

IBM SPSS Statistics Base 29

IBM

Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji [“Uwagi” na stronie 279](#).

Informacje o produkcie

Niniejsze wydanie dotyczy wersji 29, wydania 0, modyfikacji 1 produktu IBM® SPSS Statistics oraz wszystkich kolejnych wydań i modyfikacji, dopóki nie zostanie to określone inaczej w nowych wydaniach.

© Copyright International Business Machines Corporation .

Spis treści

Rozdział 1. Funkcje systemu podstawowego.....	1
Analiza mocy.....	1
Średnie.....	2
Proporcje.....	15
Korelacje.....	25
Regresja (REGRESSION).....	33
Analiza mocy: precyzja.....	36
Analiza mocy: Wartości siatki.....	37
Metaanaliza.....	37
Metaanaliza ilościowa.....	38
Metaanaliza ciągłej wielkości efektu.....	46
Metaanaliza (binarnie).....	54
Metaanaliza binarna wielkości efektu.....	63
Regresja metaanalizy.....	71
Opcje wykreślenia metaanalizy.....	75
książka kodowa.....	82
Książka kodowa – karta Raport.....	82
Książka kodowa – karta Statystyki.....	84
Częstości.....	85
Częstości: Statystyki.....	86
Wykresy częstości.....	87
Częstości: Format.....	87
statystyki opisowe.....	87
Statystyki opisowe: Opcje.....	88
Dodatkowe właściwości komendy DESCRIPTIVES.....	89
Percentyle.....	89
Percentyle-brakujące wartości.....	89
Eksploruj.....	90
Eksploracja: Statystyki.....	91
Eksploracja: Wykresy.....	91
Eksploracja: Opcje.....	92
Dodatkowe właściwości komendy EXAMINE.....	92
Tabele przestawne.....	93
Warstwy tabeli krzyżowej.....	94
Zgrupowane wykresy słupkowe tabeli krzyżowych.....	94
Tabele krzyżowe przedstawiające zmienne warstwy w warstwach tabeli.....	94
Tabele krzyżowe: Statystyki.....	94
Tabele krzyżowe: Zawartość komórek.....	96
Tabele krzyżowe: Format.....	97
Podsumowanie.....	97
Podsumowania obserwacji: Opcje.....	98
Podsumowania obserwacji: Statystyki.....	98
Średnie.....	100
Średnie: Opcje.....	100
Kostki OLAP.....	102
Kostki OLAP: Statystyki.....	103
Kostki OLAP: Różnice.....	104
Kostki OLAP: Tytuły.....	104
Proporcje.....	105
Wprowadzenie do proporcji.....	105
Proporcje dla jednej próby.....	105

Proporcje dla prób zależnych.....	108
Proporcje dla prób niezależnych.....	110
Testy t.....	113
Testy t.....	113
Test t dla prób niezależnych.....	114
Test t dla prób zależnych.....	115
Test t dla jednej próby.....	117
Dodatkowe właściwości komendy T-TEST.....	118
Jednoczynnikowa ANOVA.....	118
Jednoczynnikowa ANOVA: Kontrasty.....	119
Jednoczynnikowa ANOVA: Wielokrotne porównania post hoc.....	120
Jednoczynnikowa ANOVA: Opcje.....	121
Dodatkowe właściwości komendy ONEWAY.....	122
Analiza OML jednej zmiennej.....	122
Model OML.....	124
OML: Kontrasty.....	125
OML: Wykresy profili.....	126
OML: Porównania post hoc.....	128
OML: Zapisz.....	131
Oszacowane średnie brzegowe OML.....	132
Opcje OML.....	132
Dodatkowe właściwości komendy UNIANOVA.....	134
Korelacje parami.....	134
Korelacje parami: Opcje.....	135
Korelacje parami: przedział ufności.....	136
Dodatkowe właściwości komend CORRELATIONS (KORELACJE) oraz NONPAR CORR (KORELACJE NIEPARAMETRYCZNE).....	136
Korelacje cząstkowe.....	137
Korelacje cząstkowe: Opcje.....	137
Dodatkowe właściwości komendy KORELACJE NIEPARAMETRYCZNE.....	138
Odległości.....	138
Odległości: Miary niepodobieństwa.....	139
Odległości: Miary podobieństwa.....	139
Dodatkowe właściwości komendy PROXIMITIES.....	139
Modele liniowe.....	140
Otrzymywanie modelu liniowego.....	140
Cele.....	140
Podstawy.....	141
Wybór modelu.....	141
Zestawy.....	142
Zaawansowane.....	143
Opcje modelu.....	143
Podsumowanie modelu.....	143
Automatyczne przygotowanie danych.....	143
Ważność predyktorów.....	144
Przewidywane przez Obserwowane.....	144
Reszty.....	144
Odstające.....	144
Efekty.....	144
Współczynniki.....	145
Oszacowanie średnie.....	145
Podsumowanie tworzenia modelu.....	146
Regresja liniowa.....	146
Regresja liniowa: Metody wyboru zmiennych.....	147
Regresja liniowa: Filtrowanie.....	148
Regresja liniowa: Wykresy.....	148
Regresja liniowa: Zapisywanie zmiennych wynikowych.....	148
Regresja liniowa: Statystyki.....	150

Regresja liniowa: Opcje.....	150
Dodatkowe właściwości komendy REGRESSION.....	151
Regresja porządkowa.....	151
Opcje regresji porządkowej.....	152
Wyniki regresji porządkowej.....	153
Model położenia regresji porządkowej.....	154
Model skali regresji porządkowej.....	154
Dodatkowe właściwości komendy PLUM.....	155
Regresja liniowa sieci elastycznej.....	155
Regresja elastyczna sieci liniowej: Opcje.....	156
Regresja liniowa lasso.....	158
Regresja liniowa lasso: Opcje.....	159
Regresja liniowa grzbietowa.....	161
Regresja liniowa regresji grzbietowej: Opcje.....	162
Estymacja krzywej.....	164
Modele estymacji krzywej.....	165
Estymacja krzywej: Zapisz.....	166
Regresja metodą cząstkowych najmniejszych kwadratów.....	166
Model.....	168
Opcje.....	168
Analiza najbliższego sąsiedztwa.....	169
Sąsiedztwo.....	171
Predyktory.....	172
Podziały.....	172
Zapisywanie.....	173
Wynik.....	173
Opcje.....	174
Widok modelu.....	174
Analiza dyskryminacyjna.....	177
Analiza dyskryminacyjna: Definiuj zakres.....	178
Analiza dyskryminacyjna: Wybierz obserwacje.....	178
Analiza dyskryminacyjna: Statystyki.....	178
Analiza dyskryminacyjna: Użyj metody krokowej.....	179
Analiza dyskryminacyjna: Klasyfikacja.....	180
Analiza dyskryminacyjna: Zapisz.....	181
Dodatkowe właściwości komendy DISCRIMINANT.....	181
Analiza czynnikowa.....	181
Wybór obserwacji do analizy czynnikowej.....	182
Analiza czynnikowa: Statystyki opisowe.....	182
Analiza czynnikowa: Wyodrębnianie.....	183
Analiza czynnikowa: Rotacja.....	184
Analiza czynnikowa: Oceny czynnikowe.....	184
Analiza czynnikowa: Opcje.....	185
Dodatkowe właściwości komendy FACTOR.....	185
Wybieranie procedury analizy skupień.....	185
Dwustopniowa analiza skupień.....	186
Dwustopniowe grupowanie: Opcje.....	187
Dwustopniowe grupowanie: Wyniki.....	188
Przeglądarka skupień.....	189
Hierarchiczna analiza skupień.....	193
Metoda hierarchicznej analizy skupień.....	194
Hierarchiczna analiza skupień: Statystyki.....	195
Hierarchiczna analiza skupień: Wykresy.....	195
Hierarchiczna analiza skupień: Zapisz zmienne wynikowe.....	195
Dodatkowe właściwości składni komendy CLUSTER.....	195
Analiza skupień metodą k-średnich.....	195
Efektywność analizy skupień metodą k-średnich.....	196
Analiza skupień metodą k-średnich: Iteracja.....	197

Zapisywanie analizy skupień metodą k-średnich.....	197
Analiza skupień metodą k-średnich: Opcje.....	197
Dodatkowe właściwości komendy QUICK CLUSTER.....	198
Testy nieparametryczne.....	198
Testy nieparametryczne dla jednej próby.....	198
Testy nieparametryczne dla prób niezależnych.....	202
Testy nieparametryczne dla prób zależnych.....	204
Dodatkowe właściwości komendy NPTESTS.....	207
Wykresy tradycyjne	208
Analiza wielokrotnych odpowiedzi.....	219
Analiza wielokrotnych odpowiedzi.....	219
Definiowanie zestawów wielokrotnych odpowiedzi.....	219
Częstości wielokrotnych odpowiedzi.....	220
tabele krzyżowe wielokrotnych odpowiedzi.....	221
Przedstawianie wyników.....	223
Przedstawianie wyników.....	223
Raport – podsumowania w wierszach.....	223
Raport – podsumowania w kolumnach.....	225
Dodatkowe właściwości komendy REPORT.....	227
Analiza rzetelności.....	228
Analiza rzetelności: Statystyki.....	229
Dodatkowe właściwości komendy RELIABILITY.....	231
Ważona Kappa.....	232
Ważona Kappa: kryteria.....	233
Ważona Kappa: drukuj.....	233
Skalowanie wielowymiarowe.....	234
Skalowanie wielowymiarowe: Kształt danych.....	235
Skalowanie wielowymiarowe: Utwórz miarę na podstawie danych.....	235
Skalowanie wielowymiarowe: Model.....	235
Skalowanie wielowymiarowe: Opcje.....	236
Dodatkowe właściwości komendy ALSCAL.....	236
Statystyki ilorazowe.....	236
Statystyki ilorazowe: Statystyki.....	237
Statystyki ilorazowe.....	238
Statystyki ilorazowe: Statystyki.....	239
Wykres P-P (prawdopodobieństwo-prawdopodobieństwo).....	240
Wykres K-K (kwantyl-kwantyl).....	242
Analiza ROC.....	245
Analiza ROC: Opcje.....	246
Analiza ROC: Prezentacja.....	246
Analiza ROC: Definiuj grupy (tańcuch).....	247
Analiza ROC: Definiuj grupy (liczba).....	247
Krzywe ROC.....	248
Krzywa ROC: Opcje.....	248
Symulacja.....	249
Projektowanie symulacji opartej o plik modelu.....	249
Projektowanie symulacji opartej o niestandardowe równania.....	250
Projektowanie symulacji bez uwzględniania modelu predykcyjnego.....	251
Uruchamianie symulacji z wykorzystaniem planu symulacji.....	251
Kreator symulacji.....	252
Okno dialogowe Uruchom symulację.....	263
Praca z wynikami Symulacji w formie wykresu.....	266
Modelowanie geoprzestrzenne.....	267
Wybieranie map	267
Źródła danych	269
Reguły asocjacji geoprzestrzennych	270
Predykcja przestrzenno-czasowa	274
Zakończenie	277

Uwagi.....	279
Znaki towarowe.....	280
Indeks.....	283

Rozdział 1. Funkcje systemu podstawowego

IBM SPSS Statistics Base Edition zawiera następujące funkcje podstawowe.

Analiza mocy

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

IBM SPSS Statistics udostępnia następujące procedury analizy mocy:

Test T dla jednej próby

W analizie dla jednej próby zaobserwowane dane są gromadzone jako pojedyncza próba losowa. Zakłada się, że dane próby są niezależnie i identycznie zgodne z rozkładem normalnym ze stałą średnią i wariancją, a wnioskowanie statystyczne dotyczy parametru średniej.

Test T dla prób zależnych

W analizie dla prób zależnych obserwowane dane zawierają dwie skorelowane próby, a każda obserwacja ma dwa pomiary. Zakłada się, że dane w każdej próbie są niezależnie i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

Test T dla prób niezależnych

W analizie dla prób niezależnych dane zaobserwowane zawierają dwie niezależne próby. Zakłada się, że dane w każdej próbie są niezależnie i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

Jednoczynnikowa ANOVA

Analiza wariancji (ANOVA) to metoda statystyczna szacowania średnich kilku populacji, co do których często zakłada się, że mają rozkład normalny. Jednoczynnikowa ANOVA, często stosowany rodzaj analizy ANOVA, jest rozszerzeniem testu *t* dla dwóch prób.

Przykład. Moc testu hipotezy wykrywającego właściwą hipotezę alternatywną jest prawdopodobieństwem, że test odrzuci hipotezę testową. Ponieważ prawdopodobieństwo wystąpienia błędu typu II jest prawdopodobieństwem przyjęcia hipotezy zerowej, gdy hipoteza alternatywna jest prawdziwa, moc może być wyrażona jako (1-prawdopodobieństwo błędu typu II), czyli prawdopodobieństwo odrzucenia hipotezy zerowej, gdy hipoteza alternatywna jest prawdziwa.

Statystyki i wykresy. Test jednostronny, test dwustronny, poziom istotności, wskaźnik błędów typu I, założenia testu, odchylenie standardowe populacji, testowanie średniej populacyjnej, wartość hipotetyczna, moc dwuwymiarowa wobec wielkości próby, moc dwuwymiarowa wobec wielkości efektu, trójwymiarowy mocy wobec wielkości próby, trójwymiarowy mocy wobec wielkości efektu, kąt obrotu, pary grup, korelacja momentu iloczynowego Pearsona, średnia różnica, zakres wykresu wielkości efektu, sumaryczne odchylenie standardowe populacji, kontrasty i różnice parami, współczynniki kontrastu, test kontrastu, BONFERRONI, SIDAK, LSD, moc wobec łącznej wielkości próby, dwuwymiarowy mocy wobec sumarycznego odchylenia standardowego, trójwymiarowy mocy wobec całej próby, moc trójwymiarowa wobec łącznej wielkości próby, sumaryczne odchylenie standardowe, rozkład *t* Studenta, niecentralny rozkład *t*,

Wymagania dotyczące danych do analizy mocy

Dane

W analizie dla jednej próby zaobserwowane dane są gromadzone jako pojedyncza próba losowa.

W analizie dla prób zależnych obserwowane dane zawierają dwie skorelowane próby, a każda obserwacja ma dwa pomiary.

W analizie dla prób niezależnych dane zaobserwowane zawierają dwie niezależne próby.

Analiza wariancji (ANOVA) to metoda statystyczna szacowania średnich kilku populacji, co do których często zakłada się, że mają rozkład normalny.

Założenia

W analizie dla jednej próby zakłada się, że dane próby są niezależnie i identycznie zgodne z rozkładem normalnym ze stałą średnią i wariancją; taka analiza umożliwia wnioskowanie statystyczne dotyczące parametru średniej.

W analizie prób zależnych zakłada się, że dane w każdej próbie są niezależne i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

W analizie dla prób niezależnych zakłada się, że dane w każdej próbie są niezależnie i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

W jednoczynnikowej analizie ANOVA często zakłada się, że oszacowane średnie kilku populacji mają rozkład normalny.

Uzyskiwanie analizy mocy

1. Wybierz z menu następującą opcję:

Analiza > Analiza mocy > Porównywanie średnich > Test T dla jednej próby albo **Test T dla prób zależnych** albo **Test T dla prób niezależnych** albo **Jednoczynnikowa ANOVA**

2. Zdefiniuj wymagane założenia testu.

3. Kliknij **OK**.

Średnie

Edycja IBM SPSS Statistics Base zawiera następujące funkcje statystyczne.

Analiza mocy testu t dla jednej próby

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

W analizie dla jednej próby zaobserwowane dane są gromadzone jako pojedyncza próba losowa. Zakłada się, że dane próby są niezależnie i identycznie zgodne z rozkładem normalnym ze stałą średnią i wariancją, a wnioskowanie statystyczne dotyczy parametru średniej.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Średnie > Test T dla jednej próby

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).

3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. W przypadku wybrania opcji **Moc** jako założeń testu **Oszacowanie** należy wprowadzić odpowiednią wartość **Wielkości próby** dla wartości szacowania mocy. Wartość ta musi być liczbą całkowitą większą od 1.
5. Opcjonalnie wybierz opcję z pola **Określ**.

Wartości hipotetyczne

Ustawienie domyślne udostępnia ustawienia **Średnia populacja** i **Wartość NULL**.

Średnia populacja

Wprowadź wartość określającą średnią testowaną populację. Musi to być jedna wartość liczbowa.

Wartość hipotezy zerowej

Opcjonalnie wprowadź wartość hipotezy zerowej, która ma być testowana. Musi to być jedna wartość liczbowa.

Wielkość efektu

Statystyka F Cohena służy do szacowania wielkości efektu jako danych wejściowych dla szacowania wielkości mocy lub próbki. Zdefiniowana wielkość efektu **Wartość** jest przekazywana do kroku pośredniego w procedurze i oblicza pożądaną moc lub wielkość próby.

6. Wprowadź **Odchylenie standardowe populacji**. Wartość ta musi być jedną liczbą całkowitą większą od 0.
7. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

8. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
9. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu t dla jednej próby: Wykres” na stronie 3 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego i ustawienia wykresu trójwymiarowego).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

10. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy **Wielkość próby** została wybrana jako metoda **Oszacowania** założeń testowania, **Wartości hipotetyczna** została wybrana z listy **Określ**, a analiza **Niekierunkowa (dwustronna)** została wybrana jako **Kierunek testu**.

Analiza mocy testu t dla jednej próby: Wykres

W oknie dialogowym **Wykres** można wpływać na sposób generowania dwu- i trójwymiarowych wykresów ilustrujących moc w funkcji wielkości próby/efektu. Okno dialogowe steruje również obrotem wykresów trójwymiarowych w pionie/w poziomie.

Wykres dwuwymiarowy

Oszacowanie siły wobec wielkości próby

Gdy jest włączone, to ustawienie opcjonalne udostępnia opcje służące do sterowania dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone.

Zakres wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa od 1 i nie może być większa niż wartość pola **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wielkości efektu

Domyślnie to ustawienie opcjonalne jest wyłączone. Gdy jest włączone, wyniki zawierają wykres. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Zakres wielkości efektu

Po wybraniu tej opcji dostępna jest dolna i górna granica.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5,0.

Wykres trójwymiarowy

Oszacowanie mocy wobec

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y), obrotem w pionie/poziomie oraz zakresem wielkości próby/efektu. To ustawienie jest domyślnie wyłączone.

Wielkość efektu na osi x i wielkość próby na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Wielkość efektu na osi y i wielkość próby na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wielkości próby (oś y) i wielkości efektu (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Zakres wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa od 1 i nie może być większa niż wartość pola **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Zakres wielkości efektu

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5,0.

Analiza siły - Test T dla prób zależnych

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

W analizie dla prób zależnych obserwowane dane obejmują dwie skorelowane próby, a każda obserwacja ma dwa pomiary. Zakłada się, że dane w każdej próbie są niezależnie i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

1. Z menu wybierz:

Analizuj > Analiza mocy > Średnie > Test t dla prób zależnych

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).

3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Jeśli jako metodę testowania **Oszacowanie** wybrano **Zasilanie**, wprowadź odpowiednią wartość **Wielkości próby**. Wartość musi być dodatnią liczbą całkowitą mnie mniejszą niż 2.

5. Opcjonalnie wybierz opcję z pola **Określ**.

Wartości hipotetyczne

Ustawienie domyślne udostępnia ustawienia **Średnia** i **Odchylenie standardowe**.

Różnica średniej populacji

Jeśli jest wymagana pojedyncza średnia populacji, wprowadź wartość **Średnia różnica populacji**. Jeśli podana jest pojedyncza wartość, jest ona średnią różnicą populacji μ_d .

Uwaga: Wartość nie może być równa 0, jeśli wybrano opcję **Wielkość próby**.

Średnia populacji dla grupy 1 i grupy 2

Jeśli dla określonych par grup wymagane jest wielokrotne zapętlanie, należy wprowadzić wartości dla **Średnia populacja dla grupy 1** i **Średnia populacja dla grupy 2**. Jeśli zostanie podane wiele wartości, określają one różnice dla średniej populacji, μ_1 i μ_2 .

Uwaga: Te dwie wartości nie mogą być takie same, jeśli wybrano opcję **Wielkość próby**.

Odchylenie standardowe populacji dla średniej różnicy

Jeśli określono pojedynczą średnią populację, należy wprowadzić odchylenie standardowe populacji dla średniej wartości różnicy. Jeśli podana jest pojedyncza wartość, jest ona

odchyleniem standardowym różnic różnicy grup dla populacji, σ_d . Wartość ta musi być jedną liczbą całkowitą większą od 0.

Odchylenie standardowe populacji dla grupy 1 i grupy 2

Jeśli określono wiele średnich populacji, należy wprowadzić odchylenie standardowe populacji dla wartości grupy 1 i 2. Jeśli określono wiele wartości, są one odchyleniami standardowymi różnic grup dla populacji, σ_1 i σ_2 . Wartości muszą być pojedynczymi liczbami większymi niż 0.

Współczynnik korelacji momentu produktu Pearsona

Opcjonalnie wprowadź wartość, która określa współczynnik korelacji momentu iloczynowego Pearsona ρ . Wartość musi być wartością liczbową z zakresu od 0 do 100. nie może to być wartość 0.

Uwaga: Jeśli określona jest pojedyncza wartość **Odchylenie standardowe populacji dla średniej różnicy**, to ustawienie jest ignorowane. W przeciwnym razie wartości dla **Odchylenie standardowe populacji dla grupy 1 i grupy 2** są używane do obliczenia σ_d .

Wielkości efektów

Szacuje wielkość efektu jako dane wejściowe dla szacowania wielkości mocy lub próbki. Zdefiniowana wielkość efektu **Wartość** jest przekazywana do kroku pośredniego w procedurze i oblicza pożądaną moc lub wielkość próby.

6. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

7. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.

8. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu t dla prób zależnych: Wykres” na stronie 6 (dane wyjściowe wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

9. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy **Wielkość próby** jest wybrana jako **Oszacowanie** metody testowania i opcja **Wartości hipotetyczne** została wybrana z listy **Określ**.

Analiza mocy testu t dla prób zależnych: Wykres

W oknie dialogowym **Wykres** można wpływać na sposób generowania dwu- i trójwymiarowych wykresów ilustrujących moc w funkcji wielkości próby/efektu. Okno dialogowe steruje również obrotem wykresów trójwymiarowych w pionie/w poziomie.

Wykres dwuwymiarowy

Oszacowanie siły wobec wielkości próby

Gdy jest włączone, to ustawienie opcjonalne udostępnia opcje służące do sterowania dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone.

Zakres wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa od 1 i nie może być większa niż wartość pola **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wielkości efektu

Domyślnie to ustawienie opcjonalne jest wyłączone. Gdy jest włączone, wyniki zawierają wykres. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Zakres wielkości efektu

Po wybraniu tej opcji dostępna jest dolna i górna granica.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5,0.

Wykres trójwymiarowy

Oszacowanie mocy wobec

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y), obrotem w pionie/poziomie oraz zakresem wielkości próby/efektu. To ustawienie jest domyślnie wyłączone.

Wielkość efektu na osi x i wielkość próby na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Wielkość efektu na osi y i wielkość próby na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wielkości próby (oś y) i wielkości efektu (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Zakres wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa od 1 i nie może być większa niż wartość pola **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Zakres wielkości efektu

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5,0.

Analiza mocy testu t dla prób niezależnych

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

W analizie dla prób niezależnych zaobserwowane dane zawierają dwie niezależne próby. Zakłada się, że dane w każdej próbie są niezależnie i identycznie zgodne z rozkładem normalnym o stałej średniej i wariancji; taka analiza umożliwia wnioskowanie statystyczne dotyczące różnicy między dwiema średnimi.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Średnie > Test T dla prób niezależnych

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

Wprowadź wartość **Stosunek wielkości grup** dla oszacowania wielkości próby (wartość musi być jedną liczbą z zakresu od 0,01 do 100).

4. Jeśli jako metodę testowania **Oszacowanie** wybrano **Moc**, należy wprowadzić wartości, aby określić wielkość próby dla **Wielkość próby dla grupy 1 i grupy 2** dla porównania. Wartości muszą być liczbami całkowitymi większymi niż 1.
5. Opcjonalnie wybierz opcję z pola **Określ**.

Wartości hipotetyczne

Ustawienie domyślne udostępnia ustawienia **Średnia populacja** i **Odchylenia standardowe populacji**.

Różnica średniej populacji

Jeśli określono pojedynczą średnią populacji, wprowadź średnią wartość dla populacji. Jeśli zostanie podana pojedyncza wartość, jest ona średnią różnicą populacji σ_d . Wartość ta musi być jedną liczbą całkowitą większą od 0.

Średnia populacji dla grupy 1 i grupy 2

Jeśli określono wiele średnich populacji, należy wprowadzić średnią populację dla wartości grupy 1 i grupy 2. W przypadku podania więcej niż jednej wartości określają one średnią różnicę populacji σ_1 i σ_2 . Wartości muszą być pojedynczymi liczbami większymi niż 0.

Odchylenia standardowe populacji są

Określ, czy odchylenia standardowe populacji są **Równe dla dwóch grup**, czy **Nierówne dla dwóch grup**.

- Gdy odchylenia standardowe populacji są równe w obu grupach, wprowadź w polu **Sumaryczne odchylenie standardowe** wartość σ ; w tej sytuacji zakłada się, że wariancje obu grup są równe, czyli $\sigma_1 = \sigma_2 = \sigma$.
- Gdy odchylenia standardowe populacji nie są równe w obu grupach, wprowadź wartości **Odchylenie standardowe dla grupy 1 i grupy 2**, oznaczające σ_1 i σ_2 .

Uwaga: Gdy wartości dla **Odchylenie standardowe dla grupy 1 i grupy 2** są identyczne, są traktowane jako pojedyncza wartość.

Wielkość efektu

Szacuje wielkość efektu jako dane wejściowe dla szacowania wielkości mocy lub próbki. Zdefiniowana wielkość efektu **Wartość** jest przekazywana do kroku pośredniego w procedurze i oblicza pożądaną moc lub wielkość próby.

6. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

7. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
8. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia "Analiza mocy testu t dla prób niezależnych: Wykres" na stronie 9 (dane wyjściowe wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

9. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja "Analiza mocy: precyzja" na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy **Wielkość próby** została wybrana jako metoda **Oszacowania** założeń testowania, **Wartości hipotetyczna** została wybrana z listy **Określ**, a analiza **Niekierunkowa (dwustronna)** została wybrana jako **Kierunek testu**.

Analiza mocy testu t dla prób niezależnych: Wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami wyjściowymi w celu zilustrować dwie i trójwymiarowe potęgi według współczynnika próbkowania, wielkości efektu lub średniej liczby wykresów różnic. Okno dialogowe steruje również pionowymi i poziomymi stopniami rotacji dla wykresów trójwymiarowych.

Wykres dwuwymiarowy

Szacowanie mocy w porównaniu z przykładowym stosunkiem wielkości/mocy w porównaniu z wielkością próby

Gdy ta opcja jest włączona, to ustawienie opcjonalne udostępnia opcje służące do sterowania dwuwymiarowym wykresem mocy wobec współczynnika wielkości próby. Dodatkowe ustawienia nie są dostępne, jeśli określono wiele wartości mocy (**Szacowanie mocy w porównaniu z wielkością próby**). To ustawienie jest domyślnie wyłączone.

Przedział stosunku wielkości prób

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec współczynnika wielkości próby. Wartość musi zawierać się w przedziale od 0,01 do 100 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec współczynnika wielkości próby. Wartość musi zawierać się w zakresie od 0,01 do 100 i musi być większa niż **Dolna granica**.

Oszacowanie siły wobec wielkości efektu (lub średniej różnicy)

Domyślnie to ustawienie opcjonalne jest wyłączone. Gdy jest włączone, wyniki zawierają wykres. Jeśli w polach **Dolna granica** lub **Górna granica** nie określono żadnych wartości całkowitych, używany jest domyślny zakres wykresu wielkości efektu (lub średniej różnicy).

Zakres wielkości efektu (lub średniej różnicy)

Po wybraniu tej opcji dostępna jest dolna i górna granica.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5,0.

Wykres trójwymiarowy

Oszacowanie mocy wobec

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec współczynnika wielkości próby (oś x) i wielkości efektu (oś y), obrotem w pionie/poziomie oraz zakresem wielkości próby/efektu. To ustawienie jest domyślnie wyłączone.

Wielkość efektu (lub średnia różnica) na osi x i wielkość próby na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec współczynnika wielkości próby (oś x) i wielkości efektu (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Wielkość efektu (lub średnia różnica) na osi y i wielkość próby na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wielkości próby (oś y) i wielkości efektu (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Przedział stosunku wielkości prób

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolne ograniczenie

Steruje dolną granicą trójwymiarowego wykresu mocy wobec współczynnika wielkości próby. Wartość musi zawierać się w przedziale od 0,01 do 100 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje dolną granicą trójwymiarowego wykresu mocy wobec współczynnika wielkości próby. Wartość musi zawierać się w zakresie od 0,01 do 100 i musi być większa niż **Dolna granica**.

Zakres wielkości efektu (lub średniej różnicy)

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Dolne ograniczenie

Steruje dolną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość ta musi być większa lub równa -5,0 i nie może być większa niż wartość **Górna granica**.

Górne ograniczenie

Steruje górną granicą trójwymiarowego wykresu mocy wobec wielkości efektu. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Analiza mocy jednoczynnikowej analizy ANOVA

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Analiza wariancji (ANOVA) to metoda statystyczna szacowania średnich kilku populacji, co do których często zakłada się, że mają rozkład normalny. Jednoczynnikowa ANOVA, często stosowany rodzaj analizy ANOVA, jest rozszerzeniem testu *t* dla dwóch prób. Procedura ta umożliwia szacowanie mocy dla dwóch typów hipotez w celu porównania średnich wielu grup, całego testu i testu z określonymi kontrastami. Cały test ukierunkowany jest na hipotezę zerową, według której wszystkie średnie grupowe są równe. Test z określonymi kontrastami rozбивa ogólną hipotezę ANOVA na mniejsze średnie — łatwiejsze do opisanie i bardziej użyteczne.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Średnie > Jednoczynnikowa ANOVA

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Opcjonalnie wybierz opcję z pola **Określ**.

Wartości hipotetyczne

Ustawienie domyślne udostępnia ustawienie **Odchylenie standardowe populacji w puli**. Wprowadź pojedynczą wartość liczbową większą niż 0. Wartość domyślna to 1.

Wielkość efektu: *f* Cohena

Statystyka *F* Cohena służy do szacowania wielkości efektu jako danych wejściowych dla szacowania wielkości mocy lub próbki. Zdefiniowana wielkość efektu **Wartość** jest przekazywana do kroku pośredniego w procedurze i oblicza pożądaną moc lub wielkość próby.

Wielkość efektu: Eta-kwadrat

Eta-kwadrat (η^2) jest używane do oszacowania wielkości efektu jako danych wejściowych do oszacowania wielkości mocy lub próby. Zdefiniowana wielkość efektu **Wartość** jest przekazywana do kroku pośredniego w procedurze i oblicza pożądaną moc lub wielkość próby.

5. Po wybraniu opcji **Moc** jako założenia testowania **Oszacowanie** należy określić wartości **Wielkości grup** i **Średnie grupy**. Należy podać co najmniej dwie wartości wielkości grupy, a każda z wartości wielkości grupy musi być większa lub równa 2 (ale nie może przekraczać 2.147.483.647). Należy

również określić co najmniej dwie wartości średnie grupy (liczba określonych wartości musi być równa wielkościom grup).

Po wybraniu opcji **Wielkość próby** jako założenia testowania **Oszacowanie** należy określić wartości **Wagi grup** (opcjonalne) i **Średnie grupy**. Wagi grup przypisują wagi wielkości grupy, gdy wybrana jest opcja **Moc**.

Uwaga: Ustawienia **Wagi grup** są ignorowane, gdy określone są **Wielkości grup**.

Kliknij opcję **Dodaj**, aby uwzględnić więcej wartości **Wielkości grup**, **Wagi grup** i **Średniej grupy**.
Kliknij przycisk **Usuń**, aby usunąć istniejące wartości.

6. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
7. Opcjonalnie można kliknąć opcję **Kontrast**, aby określić ustawienia "Analiza mocy jednoczynnikowej analizy ANOVA: Kontrast" na stronie 12: Kontrast (test kontrastu, różnice parami, wielkości efektu i oszacować wielkość próby na podstawie przedziałów ufności).
8. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia "Analiza mocy jednoczynnikowej analizy ANOVA: Wykres" na stronie 13 (dane wyjściowe wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).
Uwaga: Wykres jest dostępny tylko wtedy, gdy określone są wartości **Wielkości grup**, a jako ustawienie **Oszacowania** wybrano **Moc**.
9. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja "Analiza mocy: precyzja" na stronie 36.

Uwaga: Opcja **Precyzja** jest dostępna tylko wtedy, gdy jako założenie **Oszacowania** testowania wybrano **Wielkość próby**.

Analiza mocy jednoczynnikowej analizy ANOVA: Kontrast

W oknie dialogowym **Kontrast** dostępne są opcje służące do określania kontrastu, współczynnika, wielkości efektu, różnic parami oraz ustawienia połowy szerokości przedziału ufności dla analizy mocy One-way ANOVA.

Test kontrastów

Testuj z kontrastami liniowymi

Po włączeniu tej opcji dostępne są ustawienia kontrastu i współczynników.

Kierunek testu

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

Współczynniki

W tabeli należy określić współczynniki kontrastu i zażądać testu kontrastów. Wartości w tabeli są opcjonalne. Liczba podanych wartości musi być równa liczbie wartości określonej w polach **Wielkości grup** i **Średnie grupy**. Suma dla wszystkich podanych wartości musi być równa 0. W przeciwnym razie ostatnia wartość zostanie automatycznie skorygowana tak, by warunek ten był spełniony.

Wielkość wpływu

Określ wielkość efektu testu kontrastu, który jest mierzony przez δ . Podaj pojedynczą wartość liczbową. Gdy parametr **Kierunek testu** jest ustawiony na wartość **Analiza kierunkowa (jednostronna)**, podana wartość musi mieć wartość ≥ 0 . Jeśli wartość **Moc** jest określona jako założenie testowe **Oszacowanie**, podana wartość nie może być równa 0.

Różnice parami

Oszacuj moc testu dla różnic parami

Określa, czy ma być szacowana moc testowania dla różnic parami. Domyślnie to ustawienie opcjonalne jest wyłączone, co powoduje pominięcie wyników dotyczących różnic parami.

Poprawka Bonferroniego

Korzysta z poprawki Bonferroniego przy szacowaniu mocy różnic parami. To jest ustawienie domyślne.

poprawka Sidaka

Korzysta z poprawki Sidaka przy szacowaniu mocy różnic parami.

Najmniejsza znacząca różnica

Korzysta z poprawki LSD przy szacowaniu mocy różnic parami.

Określ połowę szerokości przedziału ufności

Szacuje wielkość próby w oparciu o wartość połowy szerokości przedziału ufności. Wprowadź wartość z zakresu od 0 do 1. W przypadku testu dwumianowego o jednej próbie wartość musi być z zakresu od 0 do 0,5.

Uwaga: Zduplikowane wartości są ignorowane.

- Kliknij przycisk **Dodaj**, aby dodać określoną wartość połowy szerokości do listy.
- Podświetl istniejącą wartość połowy szerokości i kliknij przycisk **Zmień**, aby zaktualizować wartość.
- Podświetl istniejącą wartość połowy szerokości i kliknij przycisk **Usuń**, aby usunąć wartość z listy.

Analiza mocy jednoczynnikowej analizy ANOVA: Wykres

Okno dialogowe **Wykres** umożliwia sterowanie dwu- i trójwymiarowym wykresem mocy wobec wielkości próby i wielkości efektu. Okno dialogowe steruje również stopniami obrotu w pionie i poziomie dla wykresów trójwymiarowych.

Wykres dwuwymiarowy

Oszacowanie siły wobec łącznej wielkości próby

Gdy jest włączone, to ustawienie opcjonalne udostępnia opcje służące do sterowania dwuwymiarowym wykresem mocy wobec łącznej wielkości próby. To ustawienie jest domyślnie wyłączone.

Zakres całkowitej wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu łącznej wielkości próby.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa od lub równa:

- 2-krotności liczby liczb całkowitych w polu **Wielkości grup**
- 2-krotności liczby liczb całkowitych w polu **Wielkości grup** podzielonej przez najmniejszą wartość całkowitą w polu **Wielkości grup**.

Wartość nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być mniejsza niż lub równa:

- 5000/największa wartość całkowita w polu **Wielkości grup** x suma liczb całkowitych w polu **Wielkości grup**.

Wartość musi być większa niż **Dolna granica** i nie może być większa niż 2147483647.

Oszacowanie siły wobec sumarycznego odchylenia standardowego

Domyślnie to ustawienie opcjonalne jest wyłączone. Ustawienie steruje dwuwymiarowym wykresem mocy wobec sumarycznego odchylenia standardowego. Gdy jest włączone, wyniki

zawierają wykres. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu sumarycznego odchylenia standardowego.

Uwaga:

Wykres jest wyłączony, jeśli wszystkie określone wartości **Średnie grup** są takie same.

Zakres sumarycznego odchylenia standardowego

Po wybraniu tej opcji dostępna jest dolna i górna granica.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu mocy wobec sumarycznego odchylenia standardowego. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu mocy wobec sumarycznego odchylenia standardowego. Wartość musi być większa niż **Dolna granica**.

Wykres trójwymiarowy

Oszacowanie mocy wobec

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec łącznej wielkości próby (oś x) i wielkości efektu (oś y), obrotem w pionie/poziomie oraz zakresem wielkości próby/efektu. To ustawienie jest domyślnie wyłączone.

Uwaga:

Wykres jest wyłączony, jeśli wszystkie określone wartości **Średnie grup** są takie same.

Sumaryczne odchylenie standardowe na osi x i łączna wielkość próby na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec łącznej wielkości próby (oś x) i sumarycznego odchylenia standardowego (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Sumaryczne odchylenie standardowe na osi y i łączna wielkość próby na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec łącznej wielkości próby (oś y) i sumarycznego odchylenia standardowego (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Zakres całkowitej wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa niż **Dolna granica**.

Zakres sumarycznego odchylenia standardowego

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec sumarycznego odchylenia standardowego. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec sumarycznego odchylenia standardowego. Wartość musi być większa niż **Dolna granica**.

Oszacowanie siły wobec wielkości próby

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec wielkości efektu oraz zakresem wielkości próby/efektu. To ustawienie jest domyślnie wyłączone.

Uwaga:

Wykres jest wyłączony, jeśli wszystkie określone wartości **Średnie grup** są takie same.

Sumaryczne odchylenie standardowe na osi x i łączna wielkość próby na osi y

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Sumaryczne odchylenie standardowe na osi y i łączna wielkość próby na osi x

Udostępnia opcje sterowania trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i wielkości efektu (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Zakres całkowitej wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości próby.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa niż **Dolna granica**.

Zakres wielkości efektu

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie zostaną określone żadne wartości całkowite, obowiązywać będzie domyślny zakres wykresu wielkości efektu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą trójwymiarowego wykresu mocy wobec łącznej wielkości próby. Wartość musi być większa niż **Dolna granica**.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Proporcje

Edycja IBM SPSS Statistics Base zawiera następujące funkcje statystyczne.

Analiza mocy testu dwumianowego dla jednej próby

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Rozkład dwumianowy jest oparty na sekwencji prób Bernoulliego. Może być wykorzystywany do modelowania eksperymentów, w tym eksperymentów ze stałą łączną liczbą prób w założeniu uznawanych za niezależne od siebie. Każda próba prowadzi do wyniku dychotomicznego, z takim samym prawdopodobieństwem sukcesu.

Test dwumianowy dla jednej próby umożliwia wnioskowanie statystyczne dotyczące parametru proporcji poprzez porównywanie ich z wartością hipotetyczną. Metody szacowania mocy takiego testu to albo aproksymacja rozkładu normalnego, albo wyliczenie dwumianowe.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Proporcje > Test dwumianowy dla jednej próby

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Po wybraniu opcji **Szacuj moc** wprowadź odpowiednią wartość w polu **Łączna liczba prób**. Wartość musi być liczbą całkowitą nie mniejszą od 1.
5. Wprowadź wartość, która określa wartość hipotezy alternatywnej parametru proporcji w polu **Proporcja populacji**. Musi to być jedna wartość liczbowa.

Uwaga: Jeśli określona jest wartość **Moc**, wartość **Proporcja populacji** nie może być równa wartości **Wartość hipotezy zerowej**.

6. Opcjonalnie wprowadź wartość hipotezy zerowej dla parametru proporcji do przetestowania w polu **Wartość hipotezy zerowej**. Wartość musi być pojedynczą wartością liczbową z zakresu od 0 do 1. Wartością domyślną jest 0,50.
7. Wybierz metodę szacowania mocy.

Przybliżenie rozkładu normalnego

Włącza przybliżenie rozkładu normalnego. To jest ustawienie domyślne.

Zastosuj korektę ciągłości

Określa, czy w metodzie przybliżenia rozkładu normalnego ma być stosowana poprawka na ciągłość.

Wyliczenie dwumianowe

Włącza metodą wyliczenia dwumianowego. Opcjonalnie użyć pola **Limit czasu**, aby określić maksymalną liczbę minut szacowania wielkości próby. Po osiągnięciu limitu czasu analiza jest zatrzymywana i wyświetlany jest komunikat ostrzegawczy. Określona wartość musi być jedną dodatnią liczbą całkowitą wskazującą liczbę minut. Ustawieniem domyślnym jest 5 minut.

Uwaga: Wybrane założenie szacowania mocy nie ma wpływu, gdy wartość **Łącznej liczby prób** przekracza 500.

8. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

9. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu dwumianowego dla jednej próby: Wykres” na stronie 17 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko, jeśli jako założenie testu wybrano opcję **Oszacuj moc** i nie wybrano opcji **Wyliczenie dwumianowe**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy testu dwumianowego dla jednej próby: Wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami, które ilustrują dwu- i trójwymiarowe potęgi. Okno dialogowe steruje również stopniami obrotu w pionie i poziomie dla wykresów trójwymiarowych.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy wobec wartości hipotezy zerowej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy zerowej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem wartości hipotezy alternatywnej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy alternatywnej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem różnicy pomiędzy wartościami hipotetycznymi

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotetycznych. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem całkowitej liczby prób

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec łącznej liczby prób. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Wykres dla przedziału całkowitej liczby prób

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby prób. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby prób. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem całkowitej liczby prób

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i różnica między wartościami hipotetycznymi na osi y

To ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec łącznej liczby prób (oś x) i różnicy między wartościami hipotetycznymi (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i różnica między wartościami hipotetycznymi na osi x

To ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec łącznej liczby prób (oś y) i różnicy między wartościami hipotetycznymi (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Wykres dla przedziału całkowitej liczby prób

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu oszacowania mocy wobec łącznej liczby prób. Wartość musi być większa od 0 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą trójwymiarowego wykresu oszacowania mocy wobec łącznej liczby prób. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wartości hipotezy zerowej

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i wartość hipotezy alternatywnej na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś x) i alternatywnej (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i wartość hipotezy alternatywnej na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś y) i alternatywnej (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Analiza mocy testu dwumianowego dla prób zależnych

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Rozkład dwumianowy jest oparty na sekwencji prób Bernoulliego. Może być wykorzystywany do modelowania eksperymentów, w tym eksperymentów ze stałą łączną liczbą prób w założeniu

uznawanych za niezależne od siebie. Każda próba prowadzi do wyniku dychotomicznego, z takim samym prawdopodobieństwem sukcesu.

Test dwumianowy dla prób zależnych szacuje moc testu McNemara porównującego dwa parametry proporcji na podstawie dopasowanych par wybranych z dwóch powiązanych populacji dwumianowych.

1. Wybierz z menu następującą opcję:

Analiza > Analiza mocy > Proporcje > Test dwumianowy dla prób zależnych

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).

3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Jeśli jako metodę testowania **Oszacowanie** wybrano **Moc**, wprowadź wartość w polu **Łączna liczba par**.

5. Wybierz tę opcję, aby określić wartości testowane jako **Proporcje** albo **Liczebności**.

- Gdy wybrana jest opcja **Proporcje**, wprowadź wartości w polach **Proporcja 1** i **Proporcja 2**. Wartości muszą należeć do zakresu od 0 do 1.
- Gdy wybrana jest opcja **Liczebności**, wprowadź wartości w polach **Liczebność 1** i **Liczebność 2**. Wartości muszą należeć do przedziału od 0 do wartości określonej w polu **Łączna liczba par**.

Uwagi dotyczące proporcji:

- Opcja **Proporcje** jest jedyną dostępną opcją, jeśli określona jest wartość **Moc**.
- Jeśli nie jest wybrana opcja **Wartości testowe są brzegowe**: $0 < \text{Proporcja 1} + \text{Proporcja 2} \leq 1$
- Gdy wybrana jest opcja **Wartości testowe są brzegowe**:
 - $\text{Proporcja 1} * \text{Proporcja 2} > 0$
 - $\text{Proporcja 1} < 1$
 - $\text{Proporcja 2} < 1$
 - Wartości **Proporcja 1** i **Proporcja 2** nie mogą być takie same.

Uwagi dotyczące liczebności:

- Ustawienia **Liczebności** są dostępne tylko wtedy, gdy **Moc** jest wybrana jako ustawienie założeń testu **Oszacowanie**.
 - Jeśli opcja **Wartości testowe są brzegowe** nie jest wybrana: $0 < \text{Liczebność 1} + \text{Liczebność 2} \leq \text{Łączna liczba par}$
 - Gdy wybrana jest opcja **Wartości testowe są brzegowe**:
 - $\text{Liczebność 1} * \text{Liczebność 2} > 0$
 - $\text{Liczebność 1} < \text{Łączna liczba par}$
 - $\text{Liczebność 2} < \text{Łączna liczba par}$
6. Opcjonalnie można wybrać opcję **Wartości testowe są brzegowe**, aby określić, czy określone proporcje lub liczebności mają charakter brzegowy. Jeśli włączona jest opcja **Wartości testowe są brzegowe**, należy określić wartość **Korelacja między dopasowanymi parami**. Wartość musi być jedną liczbą z zakresu od -1 do 1.
7. Wybierz metodę szacowania mocy.

Przybliżenie rozkładu normalnego

Włącza przybliżenie rozkładu normalnego. To jest ustawienie domyślne.

Wyliczenie dwumianowe

Włącza metodę wyliczenia dwumianowego. Opcjonalnie użyć pola **Limit czasu**, aby określić maksymalną liczbę minut szacowania wielkości próby. Po osiągnięciu limitu czasu analiza jest

zatrzymywana i wyświetlany jest komunikat ostrzegawczy. Określona wartość musi być jedną dodatnią liczbą całkowitą wskazującą liczbę minut. Ustawieniem domyślnym jest 5 minut.

8. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

9. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.

10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu dwumianowego dla prób zależnych” na stronie 18 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko, jeśli jako założenie testu wybrano opcję **Oszacuj moc** i nie wybrano opcji **Wyliczenie dwumianowe**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy testu dwumianowego dla prób zależnych: Wykres

Okno dialogowe **Wykres** służy do sterowania dwu- i trójwymiarowymi wykresami oszacowania mocy. Okno dialogowe steruje również obrotem wykresów trójwymiarowych w pionie/w poziomie.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem całkowitej liczby par

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec łącznej liczby par. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu łącznej liczby par

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby par. Wartość musi być większa od 1 i nie może być większa niż wartość pola **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby par. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 2500.

Oszacowanie mocy względem różnicy ryzyka

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec różnicy ryzyka. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie siły wobec ilorazu ryzyka

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec ilorazu ryzyka. To ustawienie jest domyślnie wyłączone.

Wykres dla przedziału ilorazu ryzyka

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu ryzyka. Wartość nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu ryzyka. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 10.

Oszacowanie mocy wobec ilorazu szans

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec ilorazu szans. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Wykres dla przedziału ilorazu szans

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 10.

Oszacowanie siły wobec korelacji między dopasowanymi parami

Kontroluje szacowaną moc w porównaniu z korelacją między wykresem dopasowanych par. Wykres jest tworzony tylko wtedy, gdy określone są proporcje brzegowe lub liczebności (a nie proporcje niezgodne).

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem proporcji niezgodnych

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec proporcji niezgodnych. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem proporcji brzegowych

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec proporcji brzegowych. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Uwaga: To ustawienie jest dostępne tylko wtedy, gdy wybrana jest opcja **Wartości testowe są brzegowe**.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Analiza mocy testu dwumianowego dla próby niezależnych

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Rozkład dwumianowy jest oparty na sekwencji prób Bernoulliego. Może być wykorzystywany do modelowania eksperymentów, w tym eksperymentów ze stałą łączną liczbą prób w założeniu uznawanych za niezależne od siebie. Każda próba prowadzi do wyniku dychotomicznego, z takim samym prawdopodobieństwem sukcesu. Test dwumianowy dla prób niezależnych porównuje dwa niezależne parametry proporcji.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Proporcje > Test dwumianowy dla prób niezależnych

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

Opcjonalnie podaj wartość **Współczynnik wielkości grupy**. Wartość domyślna to 1.

4. Gdy wybrana jest opcja **Szacuj moc**, wprowadź wartości, aby określić **łączną liczbę prób dla grupy 1 i grupy 2**. Wartości muszą być liczbami całkowitymi większymi niż 1.
5. Określ parametry proporcji dla obu grup. Obie wartości muszą należeć do zakresu od 0 do 1.

Uwaga: Dwie wartości nie mogą być takie same, gdy określona jest wartość **Moc**.

6. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
7. Wybierz żadaną metodę testową:

Test chi-kwadrat

Szacuje moc na podstawie testu chi-kwadrat Pearsona. To jest ustawienie domyślne.

Odchylenie standardowe jest sumowane

To ustawienie opcjonalne określa, czy oszacowanie odchylenia standardowego jest sumowane, czy nie. Ustawienie jest domyślnie włączone.

Zastosuj korektę ciągłości

To ustawienie opcjonalne określa, czy stosowana jest poprawka na ciągłość. To ustawienie jest domyślnie wyłączone.

Test T

Szacuje moc na podstawie testu t Studenta.

Odchylenie standardowe jest sumowane

To ustawienie opcjonalne określa, czy oszacowanie odchylenia standardowego jest sumowane, czy nie. Ustawienie jest domyślnie włączone.

Test ilorazu wiarygodności

Szacuje moc na podstawie testu ilorazu wiarygodności.

Test dokładny Fishera

Szacuje moc na podstawie dokładnego testu Fishera.

Uwagi:

- W niektórych przypadkach obliczanie wyników dokładnego testu Fishera może się przedłużać.
- Jeśli wybrano test dokładny Fishera, wszystkie wykresy są zablokowane.

8. Wybierz metodę szacowania mocy.

Przybliżenie rozkładu normalnego

Włącza przybliżenie rozkładu normalnego. To jest ustawienie domyślne.

Wyliczenie dwumianowe

Włącza metodą wyliczenia dwumianowego. Opcjonalnie użyć pola **Limit czasu**, aby określić maksymalną liczbę minut szacowania wielkości próby. Po osiągnięciu limitu czasu analiza jest zatrzymywana i wyświetlany jest komunikat ostrzegawczy. Określona wartość musi być jedną dodatnią liczbą całkowitą wskazującą liczbę minut. Ustawieniem domyślnym jest 5 minut.

9. Wybierz, czy test jest jedno-, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu dwumianowego dla prób niezależnych: Wykres” na stronie 23 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego i ustawienia wykresu trójwymiarowego).

Uwaga: Opcja **Wykres** jest dostępna tylko, jeśli jako założenie testu wybrano opcję **Oszacuj moc** i nie wybrano opcji **Wyliczenie dwumianowe**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy testu dwumianowego dla prób niezależnych: Wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami, które ilustrują dwu- i trójwymiarowe potęgi. Okno dialogowe steruje również obrotem wykresów trójwymiarowych w pionie/w poziomie.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Szacowanie mocy w porównaniu z przykładowym stosunkiem wielkości/mocy w porównaniu z wielkością grupy

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec stosunku wielkości grup. Gdy to ustawienie jest włączone, wykres jest prezentowany. Dodatkowe ustawienia nie są dostępne, jeśli określono wiele wartości mocy (**Szacowanie mocy w porównaniu z wielkością grupy**). To ustawienie jest domyślnie wyłączone.

Przedział wykresu stosunku wielkości grup

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby par. Wartość musi być większa od 0,01 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec łącznej liczby par. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 100.

Oszacowanie mocy względem różnicy ryzyka

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec różnicy ryzyka. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie siły wobec ilorazu ryzyka

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec ilorazu ryzyka. To ustawienie jest domyślnie wyłączone.

Wykres dla przedziału ilorazu ryzyka

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu ryzyka. Wartość nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu ryzyka. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 10.

Oszacowanie mocy wobec ilorazu szans

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec ilorazu szans. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Wykres dla przedziału ilorazu szans

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 10.

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy wobec proporcji

Wybranie tego ustawienia opcjonalnego umożliwia określenie następujących opcji wykresów mocy wobec proporcji:

proporcji grupy 1 na osi x i proporcji grupy 2 na osi y

Steruje trójwymiarowym wykresem mocy wobec proporcji grupy 1 (oś x) i proporcji grupy 2 (oś y). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

proporcji grupy 1 na osi y i proporcji grupy 2 na osi x

Steruje trójwymiarowym wykresem mocy wobec proporcji grupy 2 (oś x) i proporcji grupy 1 (oś y). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy wobec wielkości grup

Wybranie tego ustawienia opcjonalnego umożliwia określenie następujących opcji wykresów mocy wobec wielkości grup:

wielkość grupy 1 na osi x i wielkość grupy 2 na osi y

Steruje trójwymiarowym wykresem mocy wobec liczby prób w grupie 1 (oś x) i liczby prób w grupie 2 (oś y). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

wielkość grupy 1 na osi y i wielkość grupy 2 na osi x

Steruje trójwymiarowym wykresem mocy wobec liczby prób w grupie 2 (oś x) i liczby prób w grupie 1 (oś y). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Zakres wykresu grupy 1

Gdy ta opcja jest wybrana, dostępne są opcje górnej i dolnej granicy przedziału wykresu grupy 1. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość ta musi być większa lub równa 2 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 2500.

Zakres wykresu grupy 2

Gdy ta opcja jest wybrana, dostępne są opcje górnej i dolnej granicy przedziału wykresu grupy 2. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość ta musi być większa lub równa 2 i nie może być większa niż **Górna granica**.

Górna granica

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec ilorazu szans. Wartość musi być większa niż **Dolna granica** i nie może być większa niż 2500.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Korelacje

Edycja IBM SPSS Statistics Base zawiera następujące funkcje statystyczne.

Analiza mocy testu korelacji Pearsona dla jednej próby

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Współczynnik korelacji liniowej Pearsona mierzy siłę zależności liniowej między dwiema ilościowymi zmiennymi losowymi, które z założenia mają dwumianowy rozkład normalny. Zgodnie z przyjętą konwencją jest to wielkość bezwymiarowa, uzyskana przez standaryzację kowariancji pomiędzy dwiema zmiennymi ciągłymi, a tym samym przyjmuje wartości od -1 do 1.

Test wykorzystuje asymptotyczną metodę Fishera do oszacowania mocy korelacji Pearsona dla jednej próby.

1. Z menu wybierz:

Analizuj > Analiza mocy > Korelacje > Współczynnik korelacji liniowej Pearsona

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).

3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Po wybraniu opcji **Szacuj moc** wprowadź odpowiednią wartość w polu **Wielkość próby w parach**. Wartość musi być jedną liczbą całkowitą większą od 3.

5. W polu **Parametr korelacji Pearsona** wprowadź wartość, która określa wartość hipotezy alternatywnej dla parametru korelacji. Wartość musi być jedną liczbą z zakresu od -1 do 1.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Parametr korelacji cząstkowej Pearsona** nie może być równa -1 lub 1 i nie może być równa wartości **Wartość hipotezy zerowej**.

6. Opcjonalnie wprowadź wartość hipotezy zerowej dla parametru korelacji do przetestowania w polu **Wartość hipotezy zerowej**. Wartość musi być jedną liczbą z zakresu od -1 do 1. Wartość domyślna wynosi 0.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Wartość hipotezy zerowej** nie może być równa -1 lub 1.

7. Opcjonalnie wybierz opcję **Użyj formuły korekty obciążenia w oszacowaniu mocy**, aby określić, czy korekta obciążenia ma być uwzględniana, czy ignorowana. To ustawienie jest domyślnie włączone, co powoduje, że składnik korekty obciążenia jest uwzględniony w szacowaniu mocy. Gdy to ustawienie nie jest wybrane, korekta obciążenia jest ignorowana.

8. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

9. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.

10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia [“Analiza mocy testu korelacji Pearsona dla jednej próby: wykres”](#) na stronie 26 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego i ustawienia wykresu trójwymiarowego).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: precyzja”](#) na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy testu korelacji Pearsona dla jednej próby: wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami, które ilustrują dwu- i trójwymiarowe potęgi. Okno dialogowe steruje również stopniami obrotu w pionie i poziomie dla wykresów trójwymiarowych.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. Ustawienia te są domyślnie wyłączone.

Oszacowanie mocy wobec wartości hipotezy zerowej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy zerowej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem wartości hipotezy alternatywnej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy alternatywnej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem różnicy pomiędzy wartościami hipotetycznymi

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotetycznych. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem wielkości próby (w parach)

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie siły wobec wielkości próby

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i różnica między wartościami hipotetycznymi na osi y

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś x) wobec różnicy między wartościami hipotetycznymi (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i różnica między wartościami hipotetycznymi na osi x

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś y) wobec różnicy między wartościami hipotetycznymi (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Przedział wykresu wielkości próby (w parach)

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicę trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wartości hipotezy zerowej

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i wartość hipotezy alternatywnej na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś x) i alternatywnej (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i wartość hipotezy alternatywnej na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś y) i alternatywnej (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Analiza siły testu korelacji Spearmana dla jednej próby

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Współczynnik korelacji rang Spearmana to nieparametryczna statystyka rangująca służąca do pomiaru monotonicznej relacji między dwiema zmiennymi, które są zwykle obcięte i nie mają rozkładu normalnego. Korelacja rang Spearmana jest równa korelacji Pearsona między wartościami rang obu zmiennych, a tym samym przyjmuje wartości od -1 do 1. Określanie mocy testu korelacji rang Spearmana jest ważnym zagadnieniem w analizie hydrologicznych danych szeregów czasowych.

Test wykorzystuje metodę asymptotyczną Fishera do oszacowania mocy dla korelacji rang Spearmana z jedną próbą.

1. Wybierz z menu następującą opcję:

Analiza > Analiza mocy > Korelacje > Spearman Rank-Order

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Po wybraniu opcji **Szacuj moc** wprowadź odpowiednią wartość w polu **Wielkość próby w parach**. Wartość musi być jedną liczbą całkowitą większą od 3.
5. W polu **Parametr korelacji Spearmana** wprowadź wartość hipotezy alternatywnej dla parametru korelacji. Wartość musi być jedną liczbą z zakresu od -1 do 1.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Parametr korelacji Spearmana** nie może być równa -1 lub 1 i nie może być równa wartości **Wartość hipotezy zerowej**.

6. Opcjonalnie wprowadź wartość hipotezy zerowej dla parametru korelacji do przetestowania w polu **Wartość hipotezy zerowej**. Wartość musi być jedną liczbą z zakresu od -1 do 1. Wartość domyślna wynosi 0.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Wartość hipotezy zerowej** nie może być równa -1 lub 1.

7. Opcjonalnie wybierz ustawienie określające sposób szacowania wariancji asymptotycznej w ramach analizy mocy.

Bonett i Wright

Szacuje wariancję metodą, którą zasugerowali Bonett i Wright. To jest ustawienie domyślne.

Fieller, Hartley i Pearson

Szacuje wariancję metodą, którą zasugerowali Fieller, Hartley i Pearson.

Caruso i Cliff

Szacuje wariancję metodą, którą zasugerowali Caruso i Cliff.

8. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

9. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.

10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia “Analiza mocy testu korelacji Spearmana dla jednej próby: wykres” na stronie 29 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja “Analiza mocy: precyzja” na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy testu korelacji Spearmana dla jednej próby: wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami, które ilustrują dwu- i trójwymiarowe potęgi. Okno dialogowe steruje również stopniami obrotu w pionie i poziomie dla wykresów trójwymiarowych.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. Ustawienia te są domyślnie wyłączone.

Oszacowanie mocy wobec wartości hipotezy zerowej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy zerowej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem wartości hipotezy alternatywnej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy alternatywnej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem różnicy pomiędzy wartościami hipotetycznymi

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotetycznych. To ustawienie jest domyślnie wyłączone.

Oszacowanie mocy względem wielkości próby (w parach)

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie siły wobec wielkości próby

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i różnica między wartościami hipotetycznymi na osi y

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś x) wobec różnicy między wartościami hipotetycznymi (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i różnica między wartościami hipotetycznymi na osi x

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś y) wobec różnicy między wartościami hipotetycznymi (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Przedział wykresu wielkości próby (w parach)

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicę trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wartości hipotezy zerowej

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i wartość hipotezy alternatywnej na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś x) i alternatywnej (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i wartość hipotezy alternatywnej na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś y) i alternatywnej (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy.

Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Analiza mocy testu korelacji cząstkowych Pearsona dla jednej próby

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Korelację cząstkową można objaśnić jako powiązanie między dwiema zmiennymi losowymi po wyeliminowaniu efektu innej zmiennej lub kilku innych zmiennych. Jest to miara użyteczna w warunkach uwikłania efektów. Podobnie jak Współczynnik korelacji Pearsona, współczynnik korelacji cząstkowej jest wielkością bezwymiarową przyjmującą wartości z przedziału od -1 do 1.

Test wykorzystuje asymptotyczną metodę Fishera do oszacowania mocy korelacji Pearsona dla jednej próby.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Korelacje > Częściowe

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Po wybraniu opcji **Moc** jako założenia **estymacji** testu, wprowadź odpowiednią wartość w polu **Wielkość próby**. Wartość musi być pojedynczą liczbą całkowitą większą niż 1.
5. Wprowadź wartość, która określa **liczbę zmiennych uznawanych za usunięte**. Wartość musi być jedną liczbą całkowitą większą lub równą 0.
6. Wprowadź wartość, która określa wartość hipotezy alternatywnej dla **Parametru korelacji cząstkowej**. Wartość musi być jedną liczbą z zakresu od -1 do 1.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Parametr korelacji cząstkowej Pearsona** nie może być równa -1 lub 1 i nie może być równa wartości **Wartość hipotezy zerowej**.

7. Opcjonalnie wprowadź wartość hipotezy zerowej dla parametru korelacji cząstkowej do przetestowania w polu **Wartość hipotezy zerowej**. Wartość musi być jedną liczbą z zakresu od -1 do 1. Wartość domyślna wynosi 0.

Uwaga: Gdy określona jest wartość **Moc**, wartość parametru **Wartość hipotezy zerowej** nie może być równa -1 lub 1.

8. Wybierz, czy test jest jedno, czy dwustronny.

Analiza niekierunkowa (dwustronna)

Gdy ta opcja jest wybrana, używany jest test dwustronny. To jest ustawienie domyślne.

Analiza kierunkowa (jednostronna)

Gdy ta opcja jest wybrana, moc obliczana jest dla testu jednostronnego.

9. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
10. Opcjonalnie można kliknąć opcję **Wykres**, aby określić ustawienia "Analiza mocy korelacji cząstkowych Pearsona: wykres" na stronie 32 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego, ustawienia wykresu trójwymiarowego i podpowiedzi).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

11. Opcjonalnie kliknij przycisk **Precyzja**, aby oszacować wielkość próby na podstawie przedziałów ufności, określając wartości przedziału ufności o połowie szerokości. Więcej informacji na ten temat zawiera sekcja "Analiza mocy: precyzja" na stronie 36.

Uwaga: **Precyzja** jest dostępna tylko wtedy, gdy wybrano opcję **Wielkość próby** jako metodę **Oszacowania** testu, a następnie wybrano analizę **Niekierunkowe (dwustronna)** jako **Kierunek testu**.

Analiza mocy korelacji cząstkowych Pearsona: wykres

W oknie dialogowym **Wykres** dostępne są opcje służące do sterowania wykresami dwu- i trójwymiarowymi, które ilustrują moc. Okno dialogowe steruje również obrotem w pionie i poziomie wykresów trójwymiarowych.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. Ustawienia te są domyślnie wyłączone.

Oszacowanie mocy wobec wartości hipotezy zerowej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy zerowej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem wartości hipotezy alternatywnej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotezy alternatywnej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem liczby zmiennych cząstkowych

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec liczby zmiennych uznawanych za usunięte. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie mocy względem różnicy pomiędzy wartościami hipotetycznymi

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wartości hipotetycznych. To ustawienie jest domyślnie wyłączone.

Oszacowanie siły wobec wielkości próby

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Wykres trójwymiarowy

Udostępnia opcje służące do sterowania trójwymiarowymi wykresami mocy. To ustawienie jest domyślnie wyłączone.

Oszacowanie siły wobec wielkości próby

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i różnica między wartościami hipotetycznymi na osi y

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś x) wobec różnicy między wartościami hipotetycznymi (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i różnica między wartościami hipotetycznymi na osi x

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś y) wobec różnicy między wartościami hipotetycznymi (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Przedział wykresu wielkości próby (w parach)

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolna granica

Steruje dolną granicą trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicę trójwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Oszacowanie mocy wobec wartości hipotezy zerowej

Wybranie tego ustawienia powoduje włączenie następujących opcji.

na osi x i wartość hipotezy alternatywnej na osi y

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś x) i alternatywnej (oś y). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

na osi y i wartość hipotezy alternatywnej na osi x

Ustawienie opcjonalne umożliwia sterowanie trójwymiarowym wykresem mocy wobec wartości hipotezy zerowej (oś y) i alternatywnej (oś x). Domyślnie wykres jest pomijany. Jeśli ta opcja będzie określona, wykres będzie widoczny.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Regresja (REGRESSION)

Edycja IBM SPSS Statistics Base zawiera następujące funkcje statystyczne.

Analiza mocy testu regresji liniowej jednej zmiennej

Ta funkcja wymaga edycji IBM SPSS Statistics Base.

Analiza mocy odgrywa kluczową rolę w planowaniu, projektowaniu i realizacji badania. Moc oblicza się zwykle przed zebraniem jakichkolwiek danych z prób, z wyjątkiem niektórych małych badań pilotażowych. Precyzyjne oszacowanie może dostarczyć badaczowi informacji o tym, jak prawdopodobne jest, że w oparciu o skończoną wielkość próby w warunkach prawdziwości hipotezy alternatywnej zostanie wykryta statystycznie istotna różnica. Jeśli moc jest zbyt mała, istnieje niewielka szansa na wykrycie znaczącej różnicy, a wyniki będą prawdopodobnie nieznaczące, nawet jeśli prawdziwe różnice faktycznie istnieją.

Regresja liniowa jednej zmiennej jest podstawową i standardową metodą statystyczną, w której badacze używają wartości kilku zmiennych w celu wyjaśnienia lub przewidzenia wartości wynikowej zmiennej ilościowej.

Test Analiza mocy testu regresji liniowej jednej zmiennej szacuje moc testu F typu III w wielu modelach regresji liniowej jednej zmiennej. Wielkość efektu jest odzwierciedlona przez wiele korelacji (częstkowych) i dostępne są strategie właściwe zarówno dla predyktorów stałych, jak i losowych. Szacowanie mocy dla predyktorów stałych jest oparte na rozkładzie niecentralnym F . W przypadku predyktorów losowych zakłada się, że zmienna przewidywana i predyktory wspólnie podążają za rozkładem normalnym wielu zmiennych. W takim przypadku szacowanie mocy opiera się na rozkładzie współczynnika wielokrotnych korelacji próby.

1. Wybierz z menu następującą opcję:

Analizuj > Analiza mocy > Regresja > Liniowa jednej zmiennej

2. Wybierz ustawienie dotyczące założeń testu (**Szacuj wielkość próby** lub **Szacuj moc**).
3. Po wybraniu opcji **Wielkość próby** należy wprowadzić wartość **Pojedyncza wartość mocy** dla wartości oszacowania wielkości próby (wartość musi być pojedynczą wartością z zakresu od 0 do 1) lub wybrać opcję **Wartości mocy siatki**, a następnie kliknąć przycisk **Siatka**, aby wyświetlić rzutowane wielkości próby dla zakresu określonych wartości mocy.

Więcej informacji na ten temat zawiera sekcja [“Analiza mocy: Wartości siatki”](#) na stronie 37.

4. Po wybraniu opcji **Szacuj moc** wprowadź odpowiednią wartość w polu **Wielkość próby** na potrzeby oszacowania mocy. Wartość musi być jedną liczbą całkowitą większą lub równą całkowitej liczbie predyktorów modelu + 2 (jeśli włączona jest opcja **Uwzględnij wyraz wolny w modelu**). W przeciwnym razie wartość musi być jedną liczbą całkowitą większą lub równą całkowitej liczbie predyktorów modelu + 1.
5. Określ wartość współczynnika korelacji częstkowej w polu **Populacja wielokrotnej korelacji częstkowej**. Wartość musi być jedną liczbą z zakresu od -1 do 1.

Uwaga: Gdy określona jest wartość **Moc**, wartość **Populacja wielokrotnej korelacji częstkowej** nie może być równa 0.

Następujące ustawienia są włączone, gdy wybrana jest opcja **Populacja wielokrotnej korelacji częstkowej**:

Łączna liczba predyktorów w modelu

Podaj liczbę wszystkich predyktorów lub predyktorów w pełnym modelu (bez uwzględniania wyrazu wolnego, jeśli ma zastosowanie). Wartość musi być jedną liczbą całkowitą większą lub równą 1.

Liczba predyktorów testowych

Podaj liczbę predyktorów testowych lub predyktorów w modelu zagnieżdżonym (bez uwzględniania wyrazu wolnego, jeśli ma zastosowanie). Wartość musi być większa lub równa 1, ale nie większa od wartości **Łączna liczba predyktorów w modelu**.

6. Określ **Wartości R-kwadrat dla** wielu współczynników korelacji dla opcji **modelu pełnego** i **modelu zagnieżdżonego**. Wartości muszą być pojedynczymi liczbami z zakresu od 0 do 1.

Uwaga: Jeśli określona jest wartość **Moc**, wartość **Model pełny** musi być większa niż wartość **Model zagnieżdżony**.

Następujące ustawienia są włączone, gdy wybrana jest opcja **Wartości R-kwadrat dla**:

Łączna liczba predyktorów - model pełny

Podaj liczbę wszystkich predyktorów lub predyktorów w pełnym modelu (bez uwzględniania wyrazu wolnego, jeśli ma zastosowanie). Wartość musi być jedną liczbą całkowitą większą lub równą 1.

Łączna liczba predyktorów - model zagnieżdżony

Podaj liczbę wszystkich predyktorów lub predyktorów w modelu zagnieżdżonym (bez uwzględniania wyrazu wolnego, jeśli ma zastosowanie). Wartość musi być większa lub równa 1, ale mniejsza od wartości **łączna liczba predyktorów w modelu**.

7. Opcjonalnie określ poziom wskaźnika błędów typu I dla testu w polu **Poziom istotności**. Wartość musi być pojedynczą wartością typu double z zakresu od 0 do 1. Wartością domyślną jest 0,05.
8. Opcjonalnie można wybrać ustawienie **Uwzględnij wyraz wolny w modelu**. Ustawienie jest domyślnie włączone. Jeśli ta opcja nie jest wybrana, wyraz wolny jest wykluczony z analizy mocy.
9. Opcjonalnie można wybrać, czy predyktory modelu mają być **Ustalone** czy **Losowe**. Ustawieniem domyślnym jest **Ustalone**.
10. Opcjonalnie można kliknąć przycisk **Wykres**, aby określić ustawienia "Analiza mocy regresji liniowej jednej zmiennej: Wykres" na stronie 35 (sposób generowania wykresu, ustawienia wykresu dwuwymiarowego i ustawienia wykresu trójwymiarowego).

Uwaga: Opcja **Wykres** jest dostępna tylko wtedy, gdy jako założenie testu wybrano opcję **Moc**.

Analiza mocy regresji liniowej jednej zmiennej: Wykres

Można wpływać na sposób generowania dwu- i trójwymiarowych wykresów ilustrujących moc w funkcji łącznej liczby prób. Można również sterować wyświetlaniem podpowiedzi i obrotem wykresów trójwymiarowych w pionie/w poziomie.

Wykres dwuwymiarowy

Udostępnia opcje służące do sterowania dwuwymiarowymi wykresami mocy. Ustawienia te są domyślnie wyłączone.

Oszacowanie mocy wobec wielokrotnej korelacji cząstkowej

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec współczynnika wielokrotnej korelacji cząstkowej. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Oszacowanie siły wobec wielkości próby

Gdy to ustawienie opcjonalne jest włączone, steruje dwuwymiarowym wykresem mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Wykres trójwymiarowy

Udostępnia opcje sterowania trójwymiarowymi wykresami oszacowania mocy, ustawieniami obrotu w pionie/w poziomie oraz określonym przez użytkownika przedziałem wielkości próby uwzględnionym na wykresie. To ustawienie jest domyślnie wyłączone.

Oszacowanie siły wobec wielkości próby

Gdy to ustawienie opcjonalne jest włączone, steruje trójwymiarowymi wykresami mocy wobec wielkości próby. To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

na osi x i wielokrotne korelacje cząstkowe na osi y

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś x) i współczynnikiem wielokrotnych korelacji cząstkowych (oś y). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

na osi y i wielokrotne korelacje cząstkowe na osi x

To ustawienie opcjonalne steruje trójwymiarowym wykresem mocy wobec wielkości próby (oś y) i współczynnikiem wielokrotnych korelacji cząstkowych (oś x). To ustawienie jest domyślnie wyłączone. Gdy to ustawienie jest włączone, wykres jest prezentowany.

Przedział wykresu wielkości próby

Po wybraniu tej opcji dostępna jest dolna i górna granica. Jeśli w polach **Dolna granica** lub **Górna granica** nie są określone żadne wartości całkowite, używany jest domyślny przedział wykresu.

Dolne ograniczenie

Steruje dolną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość ta musi być większa lub równa 4 i nie może być większa niż **Górna granica**.

Górne ograniczenie

Steruje górną granicą dwuwymiarowego wykresu oszacowania mocy wobec wielkości próby. Wartość musi być większa niż wartość w polu **Dolna granica** i nie może być większa niż 5000.

Obrót w pionie

Ustawienie opcjonalne określa kąt obrotu w pionie (zgodnie z ruchem wskazówek zegara od lewej strony) wykresu trójwymiarowego. Aby obrócić wykres w pionie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Wartość domyślna to 10.

Obrót w poziomie

Ustawienie opcjonalne określa kąt obrotu w poziomie (zgodnie z ruchem wskazówek zegara od przodu) wykresu trójwymiarowego. Aby obrócić wykres w poziomie, można użyć myszy. Ustawienie to odnosi skutek, gdy wybrany jest wykres trójwymiarowy. Wartość musi być jedną liczbą całkowitą mniejszą lub równą 359. Domyślna wartość to 325.

Analiza mocy: precyzja

Okno dialogowe **Precyzja** jest dostępne dla każdej procedury analizy mocy (z wyjątkiem regresji liniowej jednej zmiennej) i udostępnia opcje szacowania wielkości próby na podstawie określonego przedziału ufności o połowie szerokości. To okno dialogowe jest dostępne, jeśli dla oszacowania **wielkości próby** określono wartość **Pojedyncza wartość mocy** lub **Wartości mocy siatki**, a ustawienie **Kierunek testu** jest ustawione na wartość **Niekierunkowa (analiza dwustronna)**.

Uwaga: Ustawienie **Kierunek testu** nie ma zastosowania do opcji Jednoczynnikowa ANOVA,

gdy wybrane są opcje **Szacowana wielkość próby** i **Wartości mocy siatki** (kliknij element sterujący **Siatka**, aby wyświetlić okno dialogowe).

Typy przedziału ufności

Wybierz odpowiednie typy przedziałów ufności oraz czy stosowana jest korekta ciągłości.

Uwaga: **Typy przedziału ufności** są dostępne tylko dla procedur proporcji analizy mocy (proporcja dla jednej próby, test dwumianowy dla prób zależnych oraz test dwumianowy dla prób niezależnych).

Określ połowę szerokości przedziału ufności

Szacuje wielkość próby w oparciu o wartość połowy szerokości przedziału ufności. Wprowadź wartość z zakresu od 0 do 1. W przypadku testu dwumianowego o jednej próbie wartość musi być z zakresu od 0 do 0,5.

Uwaga: Zdublikowane wartości są ignorowane.

- Kliknij przycisk **Dodaj**, aby dodać określoną wartość połowy szerokości do listy.

- Podświetl istniejącą wartość połowy szerokości i kliknij przycisk **Zmień**, aby zaktualizować wartość.
- Podświetl istniejącą wartość połowy szerokości i kliknij przycisk **Usuń**, aby usunąć wartość z listy.

Analiza mocy: Wartości siatki

W oknie dialogowym **Wartości siatki** dostępne są opcje służące do określania zakresu wartości **mocy** w celu wyświetlania rzutowanych wielkości próbek w formacie siatki dla każdej określonej wartości zakresu **mocy**.

Okno dialogowe **Wartości siatki** jest dostępne dla każdej procedury analizy mocy, gdy wybrane są opcje **Szacowana wielkość próby** i **Wartości mocy siatki** (kliknij element sterujący **Siatka**, aby wyświetlić okno dialogowe).

Określ pojedynczą moc

Jeśli ta opcja jest wybrana, do wykonania analizy wymagana jest co najmniej jedna wartość. Dozwolonych jest wiele wartości, a każda wartość musi należeć do przedziału $[0, 1]$. Można określić wiele odległości rozdzielonych spacjami. Za pomocą elementów sterujących **Dodaj**, **Zmień** i **Usuń** można pracować z wartościami na liście wartości mocy.

Wszystkie wartości muszą być unikalne (zduplikowane wartości nie są dozwolone).

Określ zakres mocy

Po wybraniu tej opcji zakres wartości mocy może być określony od wartości **Początkowej** (wartość1) do wartości **Końcowej** (wartość2) z przyrostem **Co** (wartość3). Dozwolony jest tylko jeden poprawny zbiór $[wartość1 \text{ DO } wartość2 \text{ CO } wartość3]$. Musi on spełniać warunek $0 \leq wartość1 \leq wartość2 \leq 1$. Przypadki, w których $wartość1 = wartość2$, są równoważne określeniu jednej wartości1, niezależnie od wartości3.

Uwaga: Opcje **Określ pojedynczą moc** i **Określ zakres mocy** są niezależne; można wybrać jedną z tych opcji lub obie.

Metaanaliza

Metaanaliza to analiza danych uzyskanych podczas gromadzenia danych, które odpowiadają na podobne pytania badawcze. Badania te określa się mianem badań podstawowych. W metaanalizie używane są metody statystyczne na potrzeby uzyskiwania ogólnego oszacowania efektu, badania heterogeniczności między badaniami i wpływu złudzenia publikacyjnego lub, bardziej ogólnie, wpływu małych badań na rezultaty końcowe.

Produkt IBM SPSS Statistics obsługuje standardowe wielkości efektu i ogólne (wstępnie obliczone) wielkości efektu zarówno dla danych binarnych (takich jak logarytmiczny iloraz szans), jak i danych ilościowych (takich jak g Hedgesa). Informacje metaanalizyczne (na przykład wielkości efektu określonego badania i odpowiadające im błędy standardowe czy model i metoda metaanalizy) są określane na etapie deklarowania metaanalizy. Te informacje są automatycznie używane przez wszystkie następujące metaanalizy.

Obsługiwane są modele metaanalizy efektów losowych, efekt wspólnych i efektów stałych. W zależności od wybranego modelu metaanalizy dostępne są różne metody estymacji (na przykład odwrotność wariancji i Mantela-Haenszela) dla modeli efektów wspólnych i modeli efektów stałych. Dostępnych jest też kilka różnych estymatorów dla parametru wariancji między badaniami dla modelu efektów losowych.

Produkt IBM SPSS Statistics obsługuje następujące procedury metaanalizy:

- [“Metaanaliza ilościowa” na stronie 38](#)
- [“Metaanaliza ciągłej wielkości efektu” na stronie 46](#)
- [“Metaanaliza \(binarnie\)” na stronie 54](#)
- [“Metaanaliza binarna wielkości efektu” na stronie 63](#)
- [“Regresja metaanalizy” na stronie 71](#)

Metaanaliza ilościowa

Procedura Metaanaliza ilościowa przeprowadza analizę z wynikami ilościowymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Przykład

W historii przeprowadzono kilka badań w celu zbadania skuteczności pewnego leku w leczeniu cukrzycy typu II. Twierdzono, że ten lek doustny jest w stanie zmniejszyć poziom glukozy we krwi po posiłkach. Dane zostały zebrane z różnych ośrodków badawczych w latach 1979-1986.

Główny badacz chciał wyciągnąć statystyczny wniosek o skutku stosowania leku w postaci doustnej. Ze względu na fakt, że dane zostały wygenerowane z różnych badań, zaproponował sposób syntezy wyników z wielu badań, aby zrozumieć ogólny efekt i zidentyfikować te bazowe źródła zmian w wynikach.

Statistics

Poziom ufności, metoda iteracyjna, połowienie kroków, tolerancja zbieżności, średnie prób, wariancja prób, odchylenie standardowe, szacowana wielkość efektu, *d* Cohena, *g* Hedgesa, delta Glassa, różnica średnich, analiza skumulowana, metoda estymacji, metoda Trim-and-Fill, test oparty na regresji, model efektów losowych, model efektów stałych, estymator warunkowej największej wiarygodności, estymator empiryczny bayesowski, estymator Hedgesa, estymator Huntera-Schmidta, estymator DerSimoniana-Lairda, estymator Sidika-Jonkmana, korekta błędu standardowego Knappa-Hartunga, skrócona korekta błędu standardowego Knappa-Hartunga, współczynniki, test EGGERA oparty na regresji, wyraz wolny, model multiplikatywny, multiplikatywny parametr dyspersji, estymator kwadratowy, test jednorodności, miary heterogeniczności, przedział predykcji, szacowany błąd standardowy, szacowana wartość *p*, skumulowana ogólna wielkość efektu, szacowana waga badania.

Uzyskiwanie metaanalizy ilościowej

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ilościowe > Dane surowe...

2. W sekcji **Grupa poddawana leczeniu** wybierz zmienną **Rozmiar badania** oznaczającą wielkość próby dla danej grupy poddawanej leczeniu. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).
3. Wybierz zmienną **Średnia** oznaczającą średnie z prób dla danej grupy poddawanej leczeniu. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).
4. Wybierz opcję **Odchylenie standardowe**, aby określić odchylenie standardowe dla próby, albo opcję **Wariancja**, aby określić wariancję dla próby, a następnie wybierz zmienną oznaczającą odchylenie standardowe/wariancję dla danej grupy poddawanej leczeniu.
5. W sekcji **Grupa kontrolna** wybierz zmienną **Rozmiar badania** oznaczającą wielkość próby dla danej grupy kontrolnej. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).
6. Wybierz zmienną **Średnia** oznaczającą średnie z prób dla danej grupy kontrolnej. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).
7. Wybierz opcję **Odchylenie standardowe**, aby określić odchylenie standardowe dla próby, albo opcję **Wariancja**, aby określić wariancję dla próby, a następnie wybierz zmienną oznaczającą odchylenie standardowe/wariancję dla danej grupy kontrolnej.
8. Opcjonalnie wybierz zmienne **Identyfikator badania** i/lub **Etykieta badania**. Wybrana zmienna **ID badania** nie może być taka sama jak wybrana zmienna **Etykieta badania**.
9. Opcjonalnie wybierz ustawienie **Wielkość efektu**.

d Cohena

Ustawienie domyślne oszacowuje *d*Cohena. Jeżeli wybrano opcję **Skorygowany błąd standardowy**, ustawienie oszacowuje *d* Cohena i jego wariancję przy użyciu alternatywnej formuły dzielonej przez $2(N_{\text{leczzonej}} + N_{\text{kontrolnej}} - 3.94)$.

g Hedgesa

Oszacowuje g Hedgesa. Jeżeli wybrano opcję **Skorygowany błąd standardowy**, ustawienie oszacowuje g Hedgesa i jego wariancję przy użyciu alternatywnej formuły dzielonej przez $2(N_{\text{leczonej}} + N_{\text{kontrolnej}} - 3.94)$.

Delta Glassa

Oszacowuje deltę Glassa w oparciu o grupę kontrolną. Jeżeli wybrano opcję **Standaryzowany w oparciu o grupę poddawaną leczeniu**, delta Glassa jest standaryzowana w oparciu o odchylenie standardowe dla grupy poddawanej leczeniu.

Średnia różnica niestandaryzowana

Oszacowuje średnią różnicę przy założeniu, że odchylenia standardowe obu populacji są równe. Jeżeli wybrano opcję **Nierówne wariancje grup**, średnia różnica jest oszacowywana przy założeniu, że odchylenia standardowe obu populacji nie są równe.

10. Opcjonalnie wybierz ustawienie **modelu**. Jeśli ustawienia **Trim-and-Fill** są włączone, to ustawienie steruje również modelem używanym przez zestawianie w analizie obcinania i wypełniania. Jeśli ustawienia **obciążenia** są włączone, to ustawienie steruje również modelem używanym przez test oparty na regresji.

Efekty losowe

Ustawienie domyślne buduje model efektów losowych.

Efekty stałe

Buduje model efektów statycznych.

11. Opcjonalnie możesz wykonać następujące czynności:

- Kliknij przycisk **Kryteria...**, aby określić kryteria ogólne.
- Kliknij przycisk opcję **Analiza**, aby określić analizę podgrup i skumulowaną.
- Kliknij przycisk **Wnioskowanie**, aby określić metody szacowania.
- Kliknij opcję **Kontrast**, aby sterować testem kontrastu.
- Kliknij opcję **Odchylenie**, aby uzyskać dostęp do odchyleń publikacji, przeprowadzając test EGGERA oparty na regresji.
- Kliknij przycisk **Metoda Trim-and-Fill**, aby zaimplementować analizę odchyleń publikacji metodą Trim-and-Fill.
- Kliknij przycisk **Drukuj**, aby kontrolować wyniki tabeli.
- Kliknij przycisk **Zapisz**, aby zapisać szacowane statystyki w aktywnym zestawie danych.
- Kliknij przycisk **Wykres**, aby określić, które wykresy mają być uwzględnione w danych wyjściowych.

12. Kliknij **OK**.

Metaanaliza ilościowa: Kryteria

W oknie dialogowym **Kryteria** dostępne są ustawienia dotyczące określania kryteriów metaanalizy z wynikami ciągłymi dla danych surowych, które są udostępniane w aktywnym zbiorze danych na potrzeby oszacowania wielkości efektu.

Przedział ufności

To ustawienie opcjonalne określa poziom ufności. Wartość musi być wartością liczbową z zakresu od 0 do 100. Ustawieniem domyślnym jest 95.

Zasięg braków danych

Te ustawienia opcjonalne sterują sposobem, w jaki procedura obsługuje brakujące dane.

Wyłączanie obserwacji analiza po analizie

To ustawienie domyślne obejmuje wszystkie obserwacje z wystarczającą ilością danych dotyczących zmiennych używanych w poszczególnych analizach.

Wyklucz wszystkie obserwacje z brakami

Obejmuje wszystkie obserwacje z wystarczającą ilością danych na temat wszystkich zmiennych używanych we wszystkich analizach określonych w procedurze.

Braku danych zdefiniowane przez użytkownika

Ustawienia umożliwiają użytkownikowi określenie sposobu obsługi brakujących danych.

Wyklucz

W przypadku ustawienia domyślnego brakujące wartości użytkownika są traktowane jako poprawne.

Dołącz

Ignoruje nazwy brakujących wartości użytkownika i traktuje brakujące wartości użytkownika jako poprawne.

Iteracje

Maksymalna liczba iteracji

Ustawienie opcjonalne określa maksymalną liczbę iteracji w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 100. Wartość 0 oznacza, że żadne iteracje nie są wykonywane.

Maksimum kroków połowienia

Ustawienie opcjonalne określa maksymalne połowienie w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 5. Wartość 0 oznacza, że połowienie nie jest stosowane.

Zbieżność

Ustawienie opcjonalne określa tolerancję zbieżności. Wartość musi być dodatnią wartością. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartością domyślną jest $1E-6$.

Definiowanie ustawień dla procedury Metaanaliza ilościowa: Kryteria

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Dane surowe...

2. W oknie dialogowym **Metaanaliza ilościowa** kliknij opcję **Kryteria**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kryteriów.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ilościowa: Analiza

Okno dialogowe **Analiza** zawiera ustawienia określające analizę podgrup i kumulacyjną dla metaanalizy z wynikami ciągłymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Analiza podgrup

Wybierz zmienną, która wywołuje analizę podgrupy. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy skumulowanej**.

Analiza skumulowana

Wybierz zmienną, która przywołuje analizę skumulowaną i dla której przeprowadzana jest skumulowana metaanaliza. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy podgrup**. Gdy wybrana jest opcja **Rosnąco**, analiza skumulowana jest oparta na określonej zmiennej w kolejności rosnącej. Gdy wybrana jest opcja **Malejąco**, analiza skumulowana jest oparta na określonej zmiennej w porządku malejącym.

Statystyka skumulowana

Udostępnia opcje służące do zapisywania szacowanej łącznej wielkości efektu skumulowanego. Ustawienia są dostępne tylko wtedy, gdy jest wybrana zmienna **Analiza skumulowana**.

Skumulowane wielkości efektów

Zapisuje szacowaną skumulowaną łączną wielkość efektów.

Błąd standardowy

Zapisuje szacowany błąd standardowy skumulowanego ogólnego rozmiaru efektu.

Dolna granica przedziału ufności

Zapisuje dolną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Górna granica przedziału ufności

Zapisuje górną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Wartość p

Zapisuje szacowaną wartość p skumulowanej łącznej wielkości efektów.

Miejsce docelowe

Udostępnia opcje umożliwiające określenie zbioru danych lub pliku danych, w którym będzie dokonywany zapis. Po wybraniu opcji **Zbiór danych** można określić nową nazwę zbioru danych (można zachować domyślną nazwę zbioru danych). Po wybraniu opcji **Plik danych** kliknij przycisk **Przełączaj...**, aby wybrać nazwę i położenie pliku.

Definiowanie ustawień dla procedury Metaanaliza ilościowa: Analiza

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Dane surowe...

2. W oknie dialogowym **Metaanaliza ilościowa** kliknij opcję **Analiza**.

3. Wybierz i zdefiniuj odpowiednie ustawienia analizy.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa: wnioskowanie

W oknie dialogowym **Wnioskowanie** dostępne są ustawienia dotyczące określania metod estymacji dla metaanalizy ilościowej dla danych surowych, które są udostępniane w aktywnym zbiorze danych dla oszacowania wielkości efektu.

Uwaga: Okno dialogowe **Wnioskowanie** jest dostępne tylko wtedy, gdy wybrany jest model **Efekty losowe**.

Estymator

Udostępnia ustawienia dotyczące określania estymatora.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje estymatorem używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążenia, kontroluje również estymator używany przez test oparty na regresji.

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Największa wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia, które kontrolują, czy ma być stosowane dopasowanie błędu standardowego Knappa-Hartunga.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje dopasowaniem błędu standardowego używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, również steruje dopasowaniem błędu standardowego używanym przez test oparty na regresji.

Brak korekt

Ustawienie domyślne, dopasowanie nie jest stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę dopasowania Knappa-Hartunga.

Zastosuj obcięta korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas estymowania macierzy wariancji-kowariancji.

Definiowanie ustawień wnioskowania metaanalizy ilościowej

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ilościowe > Dane surowe...

2. W oknie dialogowym **Metaanaliza ciągła** kliknij opcję **Wnioskowanie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wnioskowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ilościowa: Kontrast

Okno dialogowe **Kontrast** zawiera ustawienia sterujące testem kontrastu dla metaanalizy z wynikami ciągłymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Zmienne

Lista zawiera wszystkie zmienne dostępne w zbiorze danych. Wybierz zmienne z listy i przenieś je na listę **Kontrasty**.

Kontrasty

Ta lista zawiera współczynniki, które są zapisywane jako zmienne w aktywnym zbiorze danych. Dozwolonych jest wiele zmiennych. Zmienne łańcuchowe nie są obsługiwane.

Współczynniki wprowadzane przez użytkownika

Udostępnia ustawienia służące do określania współczynników kontrastu wprowadzanych przez użytkownika. Dozwolone są tylko wartości liczbowe. Aby można było sformułować poprawny test kontrastu, liczba określonych wartości musi odpowiadać liczbie ważnych badań.

Definiowanie ustawień dla procedury Metaanaliza ilościowa: Kontrast

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ilościowe > Dane surowe...

2. W oknie dialogowym **Meta-analiza ilościowa** kliknij opcję **Kontrast**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kontrastu.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa: Odchylenie

Okno dialogowe **Odchylenie** zawiera ustawienia umożliwiające uwzględnienie odchylenia publikacji poprzez wykonywanie testów opartych na regresji metodą Eggera w ramach metaanalizy z wynikami binarnymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Test Eggera oparty na regresji

Wybranie tego ustawienia umożliwi obliczenie obciążenia publikacji poprzez test Eggera oparty na regresji.

Zmienne

Lista zawiera wszystkie dostępne zmienne zbioru danych.

Współzmiennie

Zmienne wybrane z listy **Zmienne** są traktowane jako współzmiennie. Dozwolonych jest wiele współzmiennych.

Czynniki

Zmienne wybrane z listy **Zmienne** są traktowane jako czynniki. Dozwolonych jest wiele czynników.

Uwzględnij wyraz wolny w regresji

Steruje wyrazem wolnym w teście opartym na regresji.

Uwzględnij parametr dyspersji w modelu efektów stałych

Steruje ustawieniem modelu multiplikatywnego i wprowadza parametr dyspersji multiplikatywnej do analizy. To ustawienie jest dostępne tylko wtedy, gdy wybrany jest model efektów stałych.

Szacowana statystyka oparta na rozkładzie t

Steruje dystrybucją użytą w testach opartych na regresji. Ustawienie jest domyślnie włączone, co pozwala oszacować statystyki oparte na rozkładzie T . Jeśli to ustawienie nie jest zaznaczone, statystyki są szacowane na podstawie rozkładu normalnego.

Definiowanie ustawień dla procedury Metaanaliza ilościowa: Odchylenie

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ciągłe > Dane surowe ...

2. W oknie dialogowym **Meta-analiza ilościowa** kliknij opcję **Odchylenie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia odchylenia.

4. Kliknij przycisk **Kontynuuj**.

Meta-Analiza ilościowa: Trim-and-Fill

Okno dialogowe **Metoda Trim-and-Fill** zawiera ustawienia dotyczące implementowania analizy odchylenia publikacji metodą Trim-and-Fill dla metaanalizy z wynikami ilościowymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Szacunkowa liczba brakujących badań

Kontroluje analizę metodą Trim-and-Fill dla złudzenia publikacyjnego. Wybranie tego ustawienia powoduje włączenie innych ustawień okna dialogowego.

Strona do podstawiania badań

Udostępnia opcje służące do określania strony wykresu lejkowego, na którym podstawiane są brakujące badania.

Ustalone przez zbocza próby Eggera

Ustawienie domyślne określa stronę w oparciu o szacowane nachylenie testu Eggera.

Lewo

Podstawia lewą stronę wykresu lejkowego.

Prawo

Podstawia prawą stronę wykresu lejkowego.

Metoda

Określa metodę szacowania liczby brakujących badań.

Linear

Ustawienie domyślne oblicza wartość estymatora liniowego.

Uruchom

Służy do obliczania estymatora przebiegu.

Kwadratowy

Służy do obliczania estymatora kwadratowego.

Proces iteracji

Udostępnia ustawienia służące do określania estymatora iteracji i dopasowania błędu standardowego.

Model efektów stałych

Po wybraniu tej opcji stosowany jest model efektów stałych, a estymacja iteracji i dopasowanie błędu standardowego nie będą dostępne.

Model efektów losowych

Po wybraniu tej opcji stosowany jest model efektów losowych i dostępne są następujące ustawienia.

Estymator

Udostępnia ustawienia dotyczące określania estymatora iteracji

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Maksymalna wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia służące do kontrolowania, czy do iteracji algorytmu metodą Trim-and-Fill ma zostać zastosowane dopasowanie błędu standardowego Knappa-Hartunga.

Brak korekt

Ustawienie domyślne: korekty nie są stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę korekty Knappa-Hartunga.

Zastosuj obciętą korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas szacowania macierzy wariancji-kowariancji.

Definiowanie ustawień metaanalizy ilościowej trim-and-fill

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Dane surowe...

2. W oknie dialogowym **Meta-analiza ilościowa** kliknij opcję **Trim-and-Fill**.

3. Wybierz i zdefiniuj odpowiednie ustawienia metody Trim-and-Fill.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa: wydruk

Okno dialogowe **Drukuj** zawiera ustawienia sterujące danymi wyjściowymi tabeli dla metaanalizy z wynikami ciągłymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Homogeniczność/Heterogeniczność

Udostępnia ustawienia sterujące testami homogeniczności i heterogeniczności.

Testowanie homogeniczności

Po wybraniu tej opcji w wynikach będzie się znajdować odpowiedni test homogeniczności.

Miary heterogeniczności

Jeśli ta opcja zostanie wybrana, w wynikach będą się znajdować miary heterogeniczności.

Wielkości efektów

Udostępnia następujące ustawienia wielkości efektu.

Badania indywidualne

Steruje wyświetlaniem poszczególnych badań. Po wybraniu tej opcji udostępniane są odpowiednie wyniki.

Skumulowane wielkości efektów

Steruje wyświetlaniem zbiorczej analizy. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy w oknie dialogowym **Analiza** zostanie wybrana zmienna **Analiza skumulowana**.

Przedział predykcji w modelu efektów losowych

Steruje wyświetlaniem przedziału predykcji. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy określony jest model efektów losowych.

Definiowanie ustawień dla procedury Metaanaliza ilościowa: Wydruk

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ciągłe > Dane surowe...

2. W oknie dialogowym **Meta-analiza ilościowa** kliknij opcję **Wydruk**.

3. Wybierz i zdefiniuj odpowiednie ustawienia drukowania.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa: Zapisz

W oknie dialogowym **Zapisz** dostępne są ustawienia służące do zapisywania szacowanych statystyk w aktywnym zbiorze danych w celu wykonania metaanalizy z wynikami binarnymi na surowych danych, które są udostępniane w aktywnym zbiorze danych dla oszacowania wielkości efektu.

Badania indywidualne

Udostępnia opcje służące do zapisywania szacowanej wielkości efektu.

Indywidualny rozmiar efektu

Zapisuje szacowaną wielkość efektu.

Błąd standardowy

Zapisuje szacowany błąd standardowy wielkości efektu.

Dolna granica przedziału ufności

Zapisuje szacowaną dolną granicę przedziału ufności dla wielkości efektu.

Górna granica przedziału ufności

Zapisuje szacowaną górną granicę przedziału ufności dla wielkości efektu.

Wartość p

Zapisuje szacowaną wartość p - wartość wielkości efektu.

Waga badania

Zapisuje szacowaną wagę badania.

Procent wagi badania

Zapisuje znormalizowaną wagę badania jako wartość procentową.

Definiowanie ustawień zapisywania metaanalizy ilościowej

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ilościowe > Dane surowe...

2. W oknie dialogowym **Metaanaliza ciągła** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia zapisywania szacowanych statystyk.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa: wykres

Okno dialogowe **Wykres** udostępnia ustawienia dla następujących typów wykresów:

- [“Wykres lasu” na stronie 75](#)
- [“Skumulowany wykres lasu” na stronie 77](#)
- [“Wykres bąbelkowy” na stronie 78](#)
- [“Wykres lejkowy” na stronie 79](#)
- [“Wykres Galbraitha” na stronie 80](#)

Metaanaliza ciągłej wielkości efektu

Procedura Metaanaliza wielkości efektu (ilościowo) przeprowadza metaanalizę z wynikami ilościowymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Przykład

W historii przeprowadzono kilka badań w celu zbadania skuteczności pewnego leku w leczeniu cukrzycy typu II. Twierdzono, że ten lek doustny jest w stanie zmniejszyć poziom glukozy we krwi po posiłkach. Dane zostały zebrane z różnych ośrodków badawczych w latach 1979-1986.

Główny badacz chciał wyciągnąć statystyczny wniosek o skutku stosowania leku w postaci doustnej. Ze względu na fakt, że dane zostały wygenerowane z różnych badań, zaproponował sposób syntezy wyników z wielu badań, aby zrozumieć ogólny efekt i zidentyfikować te bazowe źródła zmian w wynikach.

Statystyki

Poziom ufności, metoda iteracyjna, połowienie kroków, tolerancja zbieżności, średnie prób, wariancja prób, odchylenie standardowe, szacowana wielkość efektu, d Cohena, g Hedgesa, delta Glassa, różnica średnich, analiza skumulowana, metoda estymacji, metoda Trim-and-Fill, test oparty na regresji, model efektów losowych, model efektów stałych, estymator warunkowej największej wiarygodności, estymator empiryczny bayesowski, estymator Hedgesa, estymator Huntera-Schmidta, estymator DerSimoniana-Lairda, estymator Sidika-Jonkmana, korekta błędu standardowego Knappa-Hartunga, skrócona korekta błędu standardowego Knappa-Hartunga, współczynniki, test EGGERA oparty na regresji, wyraz wolny, model multiplikatywny, multiplikatywny parametr dyspersji, estymator kwadratowy, test jednorodności, miary heterogeniczności, przedział predykcji, szacowany błąd standardowy, szacowana wartość p , skumulowana ogólna wielkość efektu, szacowana waga badania.

Uzyskiwanie metaanalizy wielkości efektu (ilościowo)

1. Wybierz z menu następującą opcję:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. Wybierz zmienną **Wielkość efektu**, która oznacza wielkość efektu. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).

3. Wybierz opcję **Odchylenie standardowe**, aby określić odchylenie standardowe dla wielkości efektu, albo opcję **Wariancja**, aby określić wariancję dla wielkości efektu, a następnie wybierz zmienną oznaczającą odchylenie standardowe/wariancję dla wielkości efektu.

4. Opcjonalnie wybierz zmienne **Identyfikator badania** i/lub **Etykieta badania**. Wybrana zmienna **Identyfikator badania** nie może być taka sama, jak wybrana zmienna **Etykieta badania**.
5. Opcjonalnie wybierz ustawienie **modelu**. Jeśli ustawienia **Trim-and-Fill** są włączone, to ustawienie steruje również modelem używanym przez zestawianie w analizie obcinania i wypełniania. Jeśli ustawienia **obciążzeń** są włączone, to ustawienie steruje również modelem używanym przez test oparty na regresji.

Efekty losowe

Ustawienie domyślne buduje model efektów losowych.

Efekty stałe

Buduje model efektów statycznych.

6. Opcjonalnie możesz wykonać następujące czynności:

- Kliknij przycisk **Kryteria...**, aby określić kryteria ogólne.
- Kliknij przycisk opcję **Analiza**, aby określić analizę podgrup i skumulowaną.
- Kliknij przycisk **Wnioskowanie**, aby określić metody szacowania.
- Kliknij opcję **Kontrast**, aby sterować testem kontrastu.
- Kliknij opcję **Odchylenie**, aby uzyskać dostęp do odchyłeń publikacji, przeprowadzając test EGGERA oparty na regresji.
- Kliknij przycisk **Metoda Trim-and-Fill**, aby zaimplementować analizę odchyłeń publikacji metodą Trim-and-Fill.
- Kliknij przycisk **Drukuj**, aby kontrolować wyniki tabeli.
- Kliknij przycisk **Zapisz**, aby zapisać szacowane statystyki w aktywnym zestawie danych.
- Kliknij przycisk **Wykres**, aby określić, które wykresy mają być uwzględnione w danych wyjściowych.

7. Kliknij **OK**.

Metaanaliza wielkości efektu (ilościowo): Kryteria

Okno dialogowe **Kryteria** zawiera ustawienia służące do określania kryteriów metaanalizy z wynikami ilościowymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Przedział ufności

To ustawienie opcjonalnie określa poziom ufności. Wartość musi być wartością liczbową z zakresu od 0 do 100. Ustawieniem domyślnym jest 95.

Zasięg braków danych

Te ustawienia opcjonalnie sterują sposobem, w jaki procedura obsługuje brakujące dane.

Wyłączenie obserwacji analiza po analizie

To ustawienie domyślne obejmuje wszystkie obserwacje z wystarczającą ilością danych dotyczących zmiennych używanych w poszczególnych analizach.

Wyklucz wszystkie obserwacje z brakami

Obejmuje wszystkie obserwacje z wystarczającą ilością danych na temat wszystkich zmiennych używanych we wszystkich analizach określonych w procedurze.

Braku danych zdefiniowane przez użytkownika

Ustawienia umożliwiają użytkownikowi określenie sposobu obsługi brakujących danych.

Wyklucz

W przypadku ustawienia domyślnego brakujące wartości użytkownika są traktowane jako poprawne.

Dołącz

Ignoruje nazwy brakujących wartości użytkownika i traktuje brakujące wartości użytkownika jako poprawne.

Iteracje

Maksymalna liczba iteracji

Ustawienie opcjonalne określa maksymalną liczbę iteracji w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 100. Wartość 0 oznacza, że żadne iteracje nie są wykonywane.

Maksimum kroków połowienia

Ustawienie opcjonalne określa maksymalne połowienie w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 5. Wartość 0 oznacza, że połowienie nie jest stosowane.

Zbieżność

Ustawienie opcjonalne określa tolerancję zbieżności. Wartość musi być dodatnią wartością. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartością domyślną jest $1E-6$.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (ilościowo): Kryteria

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (ilościowo)** kliknij opcję **Kryteria**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kryteriów.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: analiza

W oknie dialogowym **Analiza** dostępne są ustawienia dotyczące określania podgrupy i analizy skumulowanej dla metaanalizy z ciągłymi wynikami, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Analiza podgrup

Wybierz zmienną, która wywołuje analizę podgrupy. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy skumulowanej**.

Analiza skumulowana

Wybierz zmienną, która przywołuje analizę skumulowaną i dla której przeprowadzana jest skumulowana metaanaliza. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy podgrup**. Gdy wybrana jest opcja **Rosnąco**, analiza skumulowana jest oparta na określonej zmiennej w kolejności rosnącej. Gdy wybrana jest opcja **Malejąco**, analiza skumulowana jest oparta na określonej zmiennej w porządku malejącym.

Statystyka skumulowana

Udostępnia opcje służące do zapisywania szacowanej łącznej wielkości efektu skumulowanego. Ustawienia są dostępne tylko wtedy, gdy jest wybrana zmienna **Analiza skumulowana**.

Skumulowane wielkości efektów

Zapisuje szacowaną skumulowaną łączną wielkość efektów.

Błąd standardowy

Zapisuje szacowany błąd standardowy skumulowanego ogólnego rozmiaru efektu.

Dolna granica przedziału ufności

Zapisuje dolną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Górna granica przedziału ufności

Zapisuje górną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Wartość p

Zapisuje szacowaną wartość p skumulowanej łącznej wielkości efektów.

Miejsce docelowe

Udostępnia opcje umożliwiające określenie zbioru danych lub pliku danych, w którym będzie dokonywany zapis. Po wybraniu opcji **Zbiór danych** można określić nową nazwę zbioru danych

(można zachować domyślną nazwę zbioru danych). Po wybraniu opcji **Plik danych** kliknij przycisk **Przeglądaj...**, aby wybrać nazwę i położenie pliku.

Definiowanie ustawień metaanalizy ciągłej wielkości efektu

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągłej wielkości efektu** kliknij opcję **Analiza**.

3. Wybierz i zdefiniuj odpowiednie ustawienia analizy.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: wnioskowanie

W oknie dialogowym **Wnioskowanie** dostępne są ustawienia dotyczące określania metod szacowania dla metaanalizy z ciągłymi wynikami, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Uwaga: Okno dialogowe **Wnioskowanie** jest dostępne tylko wtedy, gdy wybrany jest model **Efekty losowe**.

Estymator

Udostępnia ustawienia dotyczące określania estymatora.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje estymatorem używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, kontroluje również estymator używany przez test oparty na regresji.

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Największa wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia, które kontrolują, czy ma być stosowane dopasowanie błędu standardowego Knappa-Hartunga.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje dopasowaniem błędu standardowego używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, również steruje dopasowaniem błędu standardowego używanym przez test oparty na regresji.

Brak korekt

Ustawienie domyślne, dopasowanie nie jest stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę dopasowania Knappa-Hartunga.

Zastosuj obciążoną korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas estymowania macierzy wariancji-kowariancji.

Definiowanie ustawień wnioskowania metaanalizy ciągłej wielkości efektu

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągłej wielkości efektu** kliknij opcję **Wnioskowanie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wnioskowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: kontrast

W oknie dialogowym **Kontrast** dostępne są ustawienia sterujące testem kontrastu w celu przeprowadzenia metaanalizy z wynikami ciągłymi, gdy w aktywnym zbiorze danych zostaną wstępnie obliczone dane o wielkości efektu.

Zmienne

Lista zawiera wszystkie zmienne dostępne w zbiorze danych. Wybierz zmienne z listy i przenieś je na listę **Kontrasty**.

Kontrasty

Ta lista zawiera współczynniki, które są zapisywane jako zmienne w aktywnym zbiorze danych. Dozwolonych jest wiele zmiennych. Zmienne łańcuchowe nie są obsługiwane.

Współczynniki wprowadzane przez użytkownika

Udostępnia ustawienia służące do określania współczynników kontrastu wprowadzanych przez użytkownika. Dozwolone są tylko wartości liczbowe. Aby można było sformułować poprawny test kontrastu, liczba określonych wartości musi odpowiadać liczbie ważnych badań.

Definiowanie ustawień kontrastu metaanalizy ciągłej wielkości efektu

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągłej wielkości efektu** kliknij opcję **Kontrast**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kontrastu.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: Obciążenie

W oknie dialogowym **Obciążenie** dostępne są ustawienia pozwalające na określenie złudzenia publikacyjnego poprzez przeprowadzenie testu opartego na regresji metodą Eggera w celu przeprowadzenia metaanalizy z ciągłymi wynikami, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Test Eggera oparty na regresji

Wybranie tego ustawienia umożliwi obliczenie obciążenia publikacji poprzez test Eggera oparty na regresji.

Zmienne

Lista zawiera wszystkie dostępne zmienne zbioru danych.

Współzmiennie

Zmienne wybrane z listy **Zmienne** są traktowane jako współzmiennie. Dozwolonych jest wiele współzmiennych.

Czynniki

Zmienne wybrane z listy **Zmienne** są traktowane jako czynniki. Dozwolonych jest wiele czynników.

Uwzględnij wyraz wolny w regresji

Steruje wyrazem wolnym w teście opartym na regresji.

Uwzględnij parametr dyspersji w modelu efektów stałych

Steruje ustawieniem modelu multiplikatywnego i wprowadza parametr dyspersji multiplikatywnej do analizy. To ustawienie jest dostępne tylko wtedy, gdy wybrany jest model efektów stałych.

Szacowana statystyka oparta na rozkładzie t

Steruje dystrybucją użytą w testach opartych na regresji. Ustawienie jest domyślnie włączone, co pozwala oszacować statystyki oparte na rozkładzie T . Jeśli to ustawienie nie jest zaznaczone, statystyki są szacowane na podstawie rozkładu normalnego.

Definiowanie ustawień obciążenia metaanalizy ciągłej wielkości efektu

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza binarna wielkości efektu** kliknij opcję **Obciążenie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia obciążenia.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: Trim-and-Fill

W oknie dialogowym **Trim-and-Fill** dostępne są ustawienia implementacji analizy obcinania i wypełniania dla danych dwukierunkowe dla analizy metaanalizy z ciągłymi wynikami, gdy w aktywnym zestawie danych zostaną udostępnione dane o wielkości efektu obliczonego przed obliczeniem.

Szacunkowa liczba brakujących badań

Kontroluje analizę metodą Trim-and-Fill dla złudzenia publikacyjnego. Wybranie tego ustawienia powoduje włączenie innych ustawień okna dialogowego.

Strona do podstawiania badań

Udostępnia opcje służące do określania strony wykresu lejkowego, na którym podstawiane są brakujące badania.

Ustalone przez zbocza próby Eggera

Ustawienie domyślne określa stronę w oparciu o szacowane nachylenie testu Eggera.

Lewo

Podstawia lewą stronę wykresu lejkowego.

Prawo

Podstawia prawą stronę wykresu lejkowego.

Metoda

Określa metodę szacowania liczby brakujących badań.

Linear

Ustawienie domyślne oblicza wartość estymatora liniowego.

Uruchom

Służy do obliczania estymatora przebiegu.

Kwadratowy

Służy do obliczania estymatora kwadratowego.

Proces iteracji

Udostępnia ustawienia służące do określania estymatora iteracji i dopasowania błędu standardowego.

Model efektów stałych

Po wybraniu tej opcji stosowany jest model efektów stałych, a estymacja iteracji i dopasowanie błędu standardowego nie będą dostępne.

Model efektów losowych

Po wybraniu tej opcji stosowany jest model efektów losowych i dostępne są następujące ustawienia.

Estymator

Udostępnia ustawienia dotyczące określania estymatora iteracji

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Maksymalna wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia służące do kontrolowania, czy do iteracji algorytmu metodą Trim-and-Fill ma zostać zastosowane dopasowanie błędu standardowego Knappa-Hartunga.

Brak korekt

Ustawienie domyślne: korekty nie są stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę korekty Knappa-Hartunga.

Zastosuj obcięta korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas szacowania macierzy wariancji-kowariancji.

Definiowanie ustawień trim-and-fill metaanalizy ciągłej wielkości efektu

1. Wybierz z menu następującą opcję:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągłej wielkości efektu** kliknij opcję **Trim-and-Fill**.

3. Wybierz i zdefiniuj odpowiednie ustawienia metody Trim-and-Fill.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ilościowa efektu ciągłego: Wydruk

W oknie dialogowym **Drukuj** dostępne są ustawienia sterujące danymi wyjściowymi tabeli w celu przeprowadzenia metaanalizy z wynikami ciągłymi, gdy w aktywnym zestawie danych zostaną wstępnie obliczone dane o wielkości efektu.

Homogeniczność/Heterogeniczność

Udostępnia ustawienia sterujące testami homogeniczności i heterogeniczności.

Testowanie homogeniczności

Po wybraniu tej opcji w wynikach będzie się znajdować odpowiedni test homogeniczności.

Miary heterogeniczności

Jeśli ta opcja zostanie wybrana, w wynikach będą się znajdować miary heterogeniczności.

Wielkości efektów

Udostępnia następujące ustawienia wielkości efektu.

Badania indywidualne

Steruje wyświetlaniem poszczególnych badań. Po wybraniu tej opcji udostępniane są odpowiednie wyniki.

Skumulowane wielkości efektów

Steruje wyświetlaniem zbiorczej analizy. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy w oknie dialogowym **Analiza** zostanie wybrana zmienna **Analiza skumulowana**.

Przedział predykcji w modelu efektów losowych

Steruje wyświetlaniem przedziału predykcji. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy określony jest model efektów losowych.

Definiowanie ustawień obciążenia metaanalizy ciągłej wielkości efektu

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągła wielkości efektu** kliknij opcję **Obciążenie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia drukowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza ciągłej wielkości efektu: zapisywanie

Okno dialogowe **Zapisz** udostępnia ustawienia służące do zapisywania szacowanych statystyk w aktywnym zestawie danych w celu przeprowadzenia metaanalizy z ciągłymi wynikami, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Badania indywidualne

Udostępnia opcje służące do zapisywania szacowanej wielkości efektu.

Błąd standardowy

Zapisuje szacowany błąd standardowy wielkości efektu. Ustawienie jest dostępne tylko wtedy, gdy nie określono zmiennej **Wielkość efektu**.

Dolna granica przedziału ufności

Zapisuje szacowaną dolną granicę przedziału ufności dla wielkości efektu.

Górna granica przedziału ufności

Zapisuje szacowaną górną granicę przedziału ufności dla wielkości efektu.

Wartość p

Zapisuje szacowaną wartość p - wartość wielkości efektu.

Waga badania

Zapisuje szacowaną wagę badania.

Procent wagi badania

Zapisuje znormalizowaną wagę badania jako wartość procentową.

Definiowanie ustawień zapisywania metaanalizy ciągłej wielkości efektu

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki ciągłe > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza ciągłej wielkości efektu** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia zapisywania szacowanych statystyk.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza ciągłej wielkości efektu: Wykres

Okno dialogowe **Wykres** udostępnia ustawienia dla następujących typów wykresu:

- [“Wykres lasu” na stronie 75](#)
- [“Skumulowany wykres lasu” na stronie 77](#)
- [“Wykres bąbelkowy” na stronie 78](#)
- [“Wykres lejkowy” na stronie 79](#)
- [“Wykres Galbraitha” na stronie 80](#)

Metaanaliza (binarnie)

Procedura Metaanaliza binarna przeprowadza analizę z wynikami binarnymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Przykład

W historii przeprowadzono kilka badań w celu zbadania skuteczności pewnego leku w leczeniu cukrzycy typu II. Twierdzono, że ten lek doustny jest w stanie zmniejszyć poziom glukozy we krwi po posiłkach. Dane zostały zebrane z różnych ośrodków badawczych w latach 1979-1986.

Główny badacz chciał wyciągnąć statystyczny wniosek o skutku stosowania leku w postaci doustnej. Ze względu na fakt, że dane zostały wygenerowane z różnych badań, zaproponował sposób syntezy wyników z wielu badań, aby zrozumieć ogólny efekt i zidentyfikować te bazowe źródła zmian w wynikach.

Statystyki

Przedział ufności, iloraz logarytmu szans, iloraz logarytmu szans Peto, iloraz logarytmu ryzyka, różnica ryzyka, efekty losowe, efekty stałe, wariancja odwrotna, Mantel-Haenszel, iteracje, kroki połowienia, zbieżność, statystyki skumulowane, skumulowany efekt wielkości, warunkowa największa wiarygodność, REML, maksymalna wiarygodność, ML, empiryczny Bayesa, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, test Eggera, test Harborda, test Petersa, wyraz wolny w regresji, parametr dyspersji, homogeniczność, heterogeniczność, statystyki wykładnicze, błąd standardowy, wartość p , waga badania.

Uzyskiwanie metaanalizy binarnej

1. Wybierz z menu następującą opcję:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W sekcji **Grupa poddawana leczeniu** wybierz zmienną **Powodzenie**, która reprezentuje liczbę "powodzeń" dla danej grupy poddawanej leczeniu. Wybrana zmienna musi być zmienną numeryczną (zmiennie łańcuchowe nie są obsługiwane).
3. Wybierz zmienną **Niepowodzenie**, która reprezentuje liczbę "niepowodzeń" dla grupy poddawanej leczeniu. Wybrana zmienna musi być zmienną numeryczną (zmiennie łańcuchowe nie są obsługiwane).
4. W sekcji **Grupa kontrolna** wybierz zmienną **Powodzenie**, która reprezentuje liczbę "powodzeń" dla danej grupy kontrolnej. Wybrana zmienna musi być zmienną numeryczną (zmiennie łańcuchowe nie są obsługiwane).
5. Wybierz zmienną **Niepowodzenie** która reprezentuje liczbę "niepowodzeń" dla grupy kontrolnej. Wybrana zmienna musi być zmienną numeryczną (zmiennie łańcuchowe nie są obsługiwane).
6. Opcjonalnie wybierz zmiennę **Identyfikator badania** i/lub **Etykieta badania**. Wybrana zmienna **Identyfikator badania** nie może być taka sama, jak wybrana zmienna **Etykieta badania**.
7. Opcjonalnie wybierz ustawienie **Wielkość efektu**. Dostępne opcje to: **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto**, **Logarytmiczny iloraz ryzyka** i **Różnica ryzyka**.
8. Opcjonalnie wybierz ustawienie **modelu**. Jeśli ustawienia **Trim-and-Fill** są włączone, to ustawienie steruje również modelem używanym przez zestawianie w analizie obcinania i wypełniania. Jeśli

ustawienia **obciążeń** są włączone, to ustawienie steruje również modelem używanym przez test oparty na regresji.

Efekty losowe

Ustawienie domyślne buduje model efektów losowych.

Efekty stałe

Buduje model efektów stałych. **Odwrotna wariancja** oszacowuje wagę odwrotnej wariancji.

Mantela-Haenszel oszacowuje wagę Mantela-Haenszela.

9. Opcjonalnie możesz wykonać następujące czynności:

- Kliknij przycisk **Kryteria...**, aby określić kryteria ogólne.
- Kliknij przycisk opcję **Analiza**, aby określić analizę podgrup i skumulowaną.
- Kliknij przycisk **Wnioskowanie**, aby określić metody szacowania.
- Kliknij opcję **Kontrast**, aby sterować testem kontrastu.
- Kliknij opcję **Odchylenie**, aby uzyskać dostęp do odchyleń publikacji, przeprowadzając test EGGERA oparty na regresji.
- Kliknij przycisk **Metoda Trim-and-Fill**, aby zaimplementować analizę odchyleń publikacji metodą Trim-and-Fill.
- Kliknij przycisk **Drukuj**, aby kontrolować wyniki tabeli.
- Kliknij przycisk **Zapisz**, aby zapisać szacowane statystyki w aktywnym zestawie danych.
- Kliknij przycisk **Wykres**, aby określić, które wykresy mają być uwzględnione w danych wyjściowych.

10. Kliknij **OK**.

Metaanaliza binarna: Kryteria

W oknie dialogowym **Kryteria** dostępne są ustawienia dotyczące określania kryteriów metaanalizy z wynikami binarnymi dla danych surowych, które są udostępniane w aktywnym zbiorze danych na potrzeby oszacowania wielkości efektu.

Przedział ufności

To ustawienie opcjonalne określa poziom ufności. Wartość musi być wartością liczbową z zakresu od 0 do 100. Ustawieniem domyślnym jest 95.

Zasięg braków danych

Te ustawienia opcjonalne sterują sposobem, w jaki procedura obsługuje brakujące dane.

Wyłączanie obserwacji analiza po analizie

To ustawienie domyślne obejmuje wszystkie obserwacje z wystarczającą ilością danych dotyczących zmiennych używanych w poszczególnych analizach.

Wyklucz wszystkie obserwacje z brakami

Obejmuje wszystkie obserwacje z wystarczającą ilością danych na temat wszystkich zmiennych używanych we wszystkich analizach określonych w procedurze.

Braku danych zdefiniowane przez użytkownika

Ustawienia umożliwiają użytkownikowi określenie sposobu obsługi brakujących danych.

Wyklucz

W przypadku ustawienia domyślnego brakujące wartości użytkownika są traktowane jako poprawne.

Dołącz

Ignoruje nazwy brakujących wartości użytkownika i traktuje brakujące wartości użytkownika jako poprawne.

Iteracje

Maksymalna liczba iteracji

Ustawienie opcjonalne określa maksymalną liczbę iteracji w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 100. Wartość 0 oznacza, że żadne iteracje nie są wykonywane.

Maksimum kroków połowienia

Ustawienie opcjonalne określa maksymalne połowienie w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 5. Wartość 0 oznacza, że połowienie nie jest stosowane.

Zbieżność

Ustawienie opcjonalne określa tolerancję zbieżności. Wartość musi być dodatnią wartością. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartością domyślną jest 1E-6.

Skoryguj zerową liczebność

Ustawienia opcjonalne sterują sposobem korygowania danych o liczebności zerowej, jeśli ma to zastosowanie, przy szacowaniu wielkości efektu.

Skoryguj każdą wartość tylko tych badań, które zawierają co najmniej jedno zero

Ustawienie domyślne powoduje korygowanie tylko danych z liczebnością zerową.

Skoryguj każdą wartość wszystkich badań, tylko jeśli istnieje co najmniej jedno badanie zawierające wartość zero

Koryguje wszystkie częstości, gdy w badaniu istnieje co najmniej jedna liczba zerowa.

Koryguje każdą wartość we wszystkich badaniach

Koryguje wszystkie dane niezależnie od obecności liczebności zerowej.

Nie wprowadzaj żadnych korekt i zachowaj zera

Nie koryguje żadnych danych.

Wartość dodana

To ustawienie opcjonalne określa wartość dodawaną do danych o liczebności zerowej. Wartością domyślną jest 0,5. Wartością domyślną jest 0,5. Podana wartość musi być pojedynczą wartością liczbową większą od 0 i mniejszą lub równą 1.

Definiowanie kryteriów metaanalizy binarnej

1. Wybierz z menu następujące opcje:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kryteriów.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: analiza

Okno dialogowe **Analiza** zawiera ustawienia służące do określania analizy podgrup i skumulowanej dla metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Analiza podgrup

Wybierz zmienną, która wywołuje analizę podgrupy. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy skumulowanej**.

Analiza skumulowana

Wybierz zmienną, która przywołuje analizę skumulowaną i dla której przeprowadzana jest skumulowana metaanaliza. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy podgrup**. Gdy wybrana jest opcja **Rosnąco**, analiza skumulowana jest oparta na określonej zmiennej w kolejności rosnącej. Gdy wybrana jest opcja **Malejąco**, analiza skumulowana jest oparta na określonej zmiennej w porządku malejącym.

Statystyka skumulowana

Udostępnia opcje służące do zapisywania szacowanej łącznej wielkości efektu skumulowanego. Ustawienia są dostępne tylko wtedy, gdy jest wybrana zmienna **Analiza skumulowana**.

Skumulowane wielkości efektów

Zapisuje szacowaną skumulowaną łączną wielkość efektów.

Skumulowana wielkość efektu (postać wykładnicza)

Zapisuje szacowaną łączną wielkość efektu skumulowanego w formie wykładniczej.

Błąd standardowy

Zapisuje szacowany błąd standardowy skumulowanego ogólnego rozmiaru efektu.

Dolna granica przedziału ufności

Zapisuje dolną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Górna granica przedziału ufności

Zapisuje górną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Dolna granica przedziału ufności (postać wykładnicza)

Zapisuje dolną granicę przedziału szacowanego przedziału ufności ogólnej wielkości skumulowanych efektów w formie wykładniczej.

Górna granica przedziału ufności (forma wykładnicza)

Zapisuje górną granicę przedziału szacowanego przedziału ufności ogólnej wielkości skumulowanych efektów w formie wykładniczej.

Wartość p

Zapisuje szacowaną wartość p skumulowanej łącznej wielkości efektów.

Miejsce docelowe

Udostępnia opcje umożliwiające określenie zbioru danych lub pliku danych, w którym będzie dokonywany zapis. Po wybraniu opcji **Zbiór danych** można określić nową nazwę zbioru danych (można zachować domyślną nazwę zbioru danych). Po wybraniu opcji **Plik danych** kliknij przycisk **Przełączaj...**, aby wybrać nazwę i położenie pliku.

**Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie):
Analiza**

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Analiza**.

3. Wybierz i zdefiniuj odpowiednie ustawienia analizy.

4. Kliknij przycisk **Kontynuuj**.

Meta-analiza binarna: wnioskowanie

W oknie dialogowym **Wnioskowanie** dostępne są ustawienia dotyczące określania metod estymacji dla metaanalizy z wynikami binarnymi dla danych surowych, które są udostępniane w aktywnym zbiorze danych dla oszacowania wielkości efektu.

Uwaga: Okno dialogowe **Wnioskowanie** jest dostępne tylko wtedy, gdy wybrany jest model **Efekty losowe**.

Estymator

Udostępnia ustawienia dotyczące określania estymatora.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje estymatorem używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążenia, kontroluje również estymator używany przez test oparty na regresji.

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Największa wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia, które kontrolują, czy ma być stosowane dopasowanie błędu standardowego Knappa-Hartunga.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje dopasowaniem błędu standardowego używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, również steruje dopasowaniem błędu standardowego używanym przez test oparty na regresji.

Brak korekt

Ustawienie domyślne, dopasowanie nie jest stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę dopasowania Knappa-Hartunga.

Zastosuj obciętą korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas estymowania macierzy wariancji-kowariancji.

Definiowanie ustawień wnioskowania metaanalizy binarnej

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Wnioskowanie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wnioskowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: Kontrast

Okno dialogowe **Kontrast** zawiera ustawienia sterujące testem kontrastu dla metaanalizy z wynikami ciągłymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Zmienne

Lista zawiera wszystkie zmienne dostępne w zbiorze danych. Wybierz zmienne z listy i przenieś je na listę **Kontrasty**.

Kontrasty

Ta lista zawiera współczynniki, które są zapisywane jako zmienne w aktywnym zbiorze danych. Dozwolonych jest wiele zmiennych. Zmienne łańcuchowe nie są obsługiwane.

Współczynniki wprowadzane przez użytkownika

Udostępnia ustawienia służące do określania współczynników kontrastu wprowadzanych przez użytkownika. Dozwolone są tylko wartości liczbowe. Aby można było sformułować poprawny test kontrastu, liczba określonych wartości musi odpowiadać liczbie ważnych badań.

Wyświetl statystyki wykładnicze

Służy do sterowania uwzględnianiem statystyk w postaci wykładniczej. Gdy to ustawienie jest wybrane, wyniki zawierają statystyki w postaci wykładniczej. Dotyczy to m.in. wielkość efektów i granic przedziałów ufności. Ustawienie jest dostępne, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto** lub **Logarytmiczny iloraz ryzyka**.

Definiowanie ustawień dla procedury Metaanaliza binarna: Kontrast

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Kontrast**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kontrastu.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: Odchylenie

Okno dialogowe **Odchylenie** zawiera ustawienia umożliwiające uwzględnienie odchylenia publikacji poprzez wykonywanie testów opartych na regresji w ramach metaanalizy z wynikami binarnymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Testy oparte na regresji

Udostępnia opcje służące do określania testów opartych na regresji. Można wybrać więcej niż jeden test.

Test Eggera

Gdy ta opcja jest wybrana, przeprowadza test Eggera.

Test Harborda

Gdy ta opcja jest wybrana, przeprowadza test Harborda. Test jest dostępny, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans** lub **Logarytmiczny iloraz ryzyka**.

Test Petersa

Gdy ta opcja jest wybrana, przeprowadza test Petersa. Test jest dostępny, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans**.

Zmienne

Lista zawiera wszystkie dostępne zmienne zbioru danych.

Współzmiennie

Zmienne wybrane z listy **Zmienne** są traktowane jako współzmiennie. Dozwolonych jest wiele współzmiennych.

Czynniki

Zmienne wybrane z listy **Zmienne** są traktowane jako czynniki. Dozwolonych jest wiele czynników.

Uwzględnij wyraz wolny w regresji

Steruje wyrazem wolnym w teście opartym na regresji.

Uwzględnij parametr dyspersji w modelu efektów stałych

Steruje ustawieniem modelu multiplikatywnego i wprowadza parametr dyspersji multiplikatywnej do analizy. To ustawienie jest dostępne tylko wtedy, gdy wybrany jest model efektów stałych.

Szacowana statystyka oparta na rozkładzie t

Steruje dystrybucją użytą w testach opartych na regresji. Ustawienie jest domyślnie włączone, co pozwala oszacować statystyki oparte na rozkładzie T . Jeśli to ustawienie nie jest zaznaczone, statystyki są szacowane na podstawie rozkładu normalnego.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie): Odchylenie

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Odchylenie**.
3. Wybierz i zdefiniuj odpowiednie ustawienia obciążenia.
4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: Trim-and-Fill

W oknie dialogowym **Trim-and-Fill** dostępne są ustawienia implementacji analizy trim-and-fill (obcinania i wypełniania) dla złudzenia publikacyjnego metaanalizy z wynikami binarnymi na surowych danych, które są udostępniane w aktywnym zbiorze danych dla oszacowania wielkości efektu.

Szacunkowa liczba brakujących badań

Kontroluje analizę metodą Trim-and-Fill dla złudzenia publikacyjnego. Wybranie tego ustawienia powoduje włączenie innych ustawień okna dialogowego.

Strona do podstawiania badań

Udostępnia opcje służące do określania strony wykresu lejkowego, na którym podstawiane są brakujące badania.

Ustalone przez zbocza próby Eggera

Ustawienie domyślne określa stronę w oparciu o szacowane nachylenie testu Eggera.

Lewo

Podstawia lewą stronę wykresu lejkowego.

Prawo

Podstawia prawą stronę wykresu lejkowego.

Metoda

Określa metodę szacowania liczby brakujących badań.

Linear

Ustawienie domyślne oblicza wartość estymatora liniowego.

Uruchom

Służy do obliczania estymatora przebiegu.

Kwadratowy

Służy do obliczania estymatora kwadratowego.

Proces iteracji

Udostępnia ustawienia służące do określania estymatora iteracji i dopasowania błędu standardowego.

Model efektów stałych

Po wybraniu tej opcji stosowany jest model efektów stałych, a estymacja iteracji i dopasowanie błędu standardowego nie będą dostępne.

Model efektów losowych

Po wybraniu tej opcji stosowany jest model efektów losowych i dostępne są następujące ustawienia.

Estymator

Udostępnia ustawienia dotyczące określania estymatora iteracji

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Maksymalna wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia służące do kontrolowania, czy do iteracji algorytmu metodą Trim-and-Fill ma zostać zastosowane dopasowanie błędu standardowego Knappa-Hartunga.

Brak korekt

Ustawienie domyślne: korekty nie są stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę korekty Knappa-Hartunga.

Zastosuj obciętą korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas szacowania macierzy wariancji-kowariancji.

Definiowanie ustawień metaanalizy binarnej metodą Trim-and-Fill

1. Wybierz z menu następującą opcję:
Analizuj > Metaanaliza > Wyniki binarne > Dane surowe...
2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Trim-and-Fill**.
3. Wybierz i zdefiniuj odpowiednie ustawienia metody Trim-and-Fill.
4. Kliknij przycisk **Kontynuuj**.

Metaanaliza (binarnie): Drukuj

Okno dialogowe **Drukuj** zawiera ustawienia sterujące danymi wyjściowymi tabeli dla metaanalizy z wynikami binarnymi dla danych nieprzetworzonych, które udostępniono w aktywnym zestawie danych na potrzeby estymacji wielkości efektu.

Homogeniczność/Heterogeniczność

Udostępnia ustawienia sterujące testami homogeniczności i heterogeniczności.

Testowanie homogeniczności

Po wybraniu tej opcji w wynikach będzie się znajdować odpowiedni test homogeniczności.

Miary heterogeniczności

Jeśli ta opcja zostanie wybrana, w wynikach będą się znajdować miary heterogeniczności.

Wielkości efektów

Udostępnia następujące ustawienia wielkości efektu.

Badania indywidualne

Steruje wyświetlaniem poszczególnych badań. Po wybraniu tej opcji udostępniane są odpowiednie wyniki.

Skumulowane wielkości efektów

Steruje wyświetlaniem zbiorczej analizy. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy w oknie dialogowym **Analiza** zostanie wybrana zmienna **Analiza skumulowana**.

Przedział predykcji w modelu efektów losowych

Steruje wyświetlaniem przedziału predykcji. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy określony jest model efektów losowych.

Wyświetl statystyki wykładnicze

Steruje włączeniem statystyk wykładniczych. Po wybraniu ustawienia statystyki wykładnicze, w tym wykładnicza wielkość efektu i limity przedziału ufności, są umieszczane w wynikach. Proces ma

zastosowanie zarówno do wymaganych, jak i opcjonalnych tabel szacowania i predykcji wielkości efektu. Ustawienie jest dostępne, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto** lub **Logarytmiczny iloraz ryzyka**.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie):

Drukuj

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Drukuj**.

3. Wybierz i zdefiniuj odpowiednie ustawienia drukowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: zapisz

W oknie dialogowym **Zapisz** dostępne są ustawienia służące do zapisywania szacowanych statystyk w aktywnym zbiorze danych w celu przeprowadzenia metaanalizy z wynikami binarnymi na surowych danych, które są udostępniane w aktywnym zbiorze danych dla oszacowania wielkości efektu.

Badania indywidualne

Udostępnia opcje służące do zapisywania szacowanej wielkości efektu.

Indywidualny rozmiar efektu

Zapisuje szacowaną wielkość efektu.

Indywidualna wielkość efektu (forma wykładnicza)

Zapisuje szacowaną wielkość efektu w postaci wykładniczej.

Błąd standardowy

Zapisuje szacowany błąd standardowy wielkości efektu.

Dolna granica przedziału ufności

Zapisuje szacowaną dolną granicę przedziału ufności dla wielkości efektu.

Górna granica przedziału ufności

Zapisuje szacowaną górną granicę przedziału ufności dla wielkości efektu.

Dolna granica przedziału ufności (forma wykładnicza)

Umożliwia zapisanie dolnego limitu szacowanego przedziału ufności dla wielkości efektu w postaci wykładniczej.

Górna granica przedziału ufności (forma wykładnicza)

Umożliwia zapisanie górnego limitu szacowanego przedziału ufności dla wielkości efektu w postaci wykładniczej.

Wartość p

Zapisuje szacowaną wartość p - wartość wielkości efektu.

Waga badania

Zapisuje szacowaną wagę badania.

Procent wagi badania

Zapisuje znormalizowaną wagę badania jako wartość procentową.

Definiowanie ustawień zapisywania metaanalizy binarnej

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza binarna** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia zapisywania szacowanych statystyk.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna: Wykres

Okno dialogowe **Wykres** udostępnia ustawienia dla następujących typów wykresu:

- [“Wykres lasu” na stronie 75](#)
- [“Skumulowany wykres lasu” na stronie 77](#)
- [“Wykres bąbelkowy” na stronie 78](#)
- [“Wykres lejkowy” na stronie 79](#)
- [“Wykres Galbraitha” na stronie 80](#)
- [“Wykres L' Abba” na stronie 81](#)

Metaanaliza binarna wielkości efektu

Procedura metaanalizy binarnej wielkości efektu przeprowadza metaanalizę z wynikami binarnymi, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Przykład

W historii przeprowadzono kilka badań w celu zbadania skuteczności pewnego leku w leczeniu cukrzycy typu II. Twierdzono, że ten lek doustny jest w stanie zmniejszyć poziom glukozy we krwi po posiłkach. Dane zostały zebrane z różnych ośrodków badawczych w latach 1979-1986.

Główny badacz chciał wyciągnąć statystyczny wniosek o skutku stosowania leku w postaci doustnej. Ze względu na fakt, że dane zostały wygenerowane z różnych badań, zaproponował sposób syntezy wyników z wielu badań, aby zrozumieć ogólny efekt i zidentyfikować te bazowe źródła zmian w wynikach.

Statystyki

Przedział ufności, iloraz logarytmu szans, iloraz logarytmu szans Peto, iloraz logarytmu ryzyka, różnica ryzyka, efekty losowe, efekty stałe, wariancja odwrotna, Mantel-Haenszel, iteracje, kroki połowienia, zbieżność, statystyki skumulowane, skumulowany efekt wielkości, warunkowa największa wiarygodność, REML, maksymalna wiarygodność, ML, empiryczny Bayesa, Hedges, Hunter-Schmidt, DerSimonian-Laird, Sidik-Jonkman, Knapp-Hartung, test Eggera, test Harborda, test Petersa, wyraz wolny w regresji, parametr dyspersji, homogeniczność, heterogeniczność, statystyki wykładnicze, błąd standardowy, wartość p , waga badania.

Uzyskiwanie analizy wielkości efektu z użyciem binarnej metaanalizy

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. Wybierz zmienną **Wielkość efektu**, która oznacza wielkość efektu. Wybrana zmienna musi być zmienną numeryczną (zmienne łańcuchowe nie są obsługiwane).
3. Wybierz opcję **Odchylenie standardowe**, aby określić odchylenie standardowe dla wielkości efektu, albo opcję **Wariancja**, aby określić wariancję dla wielkości efektu, a następnie wybierz zmienną oznaczającą odchylenie standardowe/wariancję dla wielkości efektu.
4. Opcjonalnie wybierz zmienne **Identyfikator badania** i/lub **Etykieta badania**. Wybrana zmienna **Identyfikator badania** nie może być taka sama, jak wybrana zmienna **Etykieta badania**.
5. Opcjonalnie wybierz ustawienie **Wielkość efektu**. Dostępne opcje to: **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto**, **Logarytmiczny iloraz ryzyka** i **Różnica ryzyka**.
6. Opcjonalnie wybierz ustawienie **modelu**. Jeśli ustawienia **Trim-and-Fill** są włączone, to ustawienie steruje również modelem używanym przez zestawianie w analizie obcinania i wypełniania. Jeśli ustawienia **obciążień** są włączone, to ustawienie steruje również modelem używanym przez test oparty na regresji.

Efekty losowe

Ustawienie domyślne buduje model efektów losowych.

Efekty stałe

Buduje model efektów statycznych.

7. Opcjonalnie możesz wykonać następujące czynności:

- Kliknij przycisk **Kryteria...**, aby określić kryteria ogólne.
- Kliknij przycisk opcję **Analiza**, aby określić analizę podgrup i skumulowaną.
- Kliknij przycisk **Wnioskowanie**, aby określić metody szacowania.
- Kliknij opcję **Kontrast**, aby sterować testem kontrastu.
- Kliknij opcję **Odchylenie**, aby uzyskać dostęp do odchyłeń publikacji, przeprowadzając test EGGERA oparty na regresji.
- Kliknij przycisk **Metoda Trim-and-Fill**, aby zaimplementować analizę odchyłeń publikacji metodą Trim-and-Fill.
- Kliknij przycisk **Drukuj**, aby kontrolować wyniki tabeli.
- Kliknij przycisk **Zapisz**, aby zapisać szacowane statystyki w aktywnym zestawie danych.
- Kliknij przycisk **Wykres**, aby określić, które wykresy mają być uwzględnione w danych wyjściowych.

8. Kliknij **OK**.

Metaanaliza wielkości efektu (binarnie): Kryteria

Okno dialogowe **Kryteria** zawiera ustawienia służące do określania kryteriów metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Przedział ufności

To ustawienie opcjonalne określa poziom ufności. Wartość musi być wartością liczbową z zakresu od 0 do 100. Ustawieniem domyślnym jest 95.

Zasięg braków danych

Te ustawienia opcjonalne sterują sposobem, w jaki procedura obsługuje brakujące dane.

Wyłączanie obserwacji analiza po analizie

To ustawienie domyślne obejmuje wszystkie obserwacje z wystarczającą ilością danych dotyczących zmiennych używanych w poszczególnych analizach.

Wyklucz wszystkie obserwacje z brakami

Obejmuje wszystkie obserwacje z wystarczającą ilością danych na temat wszystkich zmiennych używanych we wszystkich analizach określonych w procedurze.

Braku danych zdefiniowane przez użytkownika

Ustawienia umożliwiają użytkownikowi określenie sposobu obsługi brakujących danych.

Wyklucz

W przypadku ustawienia domyślnego brakujące wartości użytkownika są traktowane jako poprawne.

Dołącz

Ignoruje nazwy brakujących wartości użytkownika i traktuje brakujące wartości użytkownika jako poprawne.

Iteracje

Maksymalna liczba iteracji

Ustawienie opcjonalne określa maksymalną liczbę iteracji w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 100. Wartość 0 oznacza, że żadne iteracje nie są wykonywane.

Maksimum kroków połowienia

Ustawienie opcjonalne określa maksymalne połowienie w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 5. Wartość 0 oznacza, że połowienie nie jest stosowane.

Zbieżność

Ustawienie opcjonalne określa tolerancję zbieżności. Wartość musi być dodatnią wartością. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartością domyślną jest $1E-6$.

Definiowanie kryteriów metaanalizy wielkości efektu binarnego

1. Wybierz z menu następujące opcje:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Kryteria**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kryteriów.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna wielkości efektu: analiza

W oknie dialogowym **Analiza** dostępne są ustawienia służące do określania podgrupy i analizy skumulowanej dla metaanalizy binarnej, gdy w aktywnym zestawie danych zostaną udostępnione wstępnie obliczone dane o wielkości efektu.

Analiza podgrup

Wybierz zmienną, która wywołuje analizę podgrupy. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy skumulowanej**.

Analiza skumulowana

Wybierz zmienną, która przywołuje analizę skumulowaną i dla której przeprowadzana jest skumulowana metaanaliza. Zmienna nie może być taka sama, jak zmienna zdefiniowana dla **analizy podgrup**. Gdy wybrana jest opcja **Rosnąco**, analiza skumulowana jest oparta na określonej zmiennej w kolejności rosnącej. Gdy wybrana jest opcja **Malejąco**, analiza skumulowana jest oparta na określonej zmiennej w porządku malejącym.

Statystyka skumulowana

Udostępnia opcje służące do zapisywania szacowanej łącznej wielkości efektu skumulowanego. Ustawienia są dostępne tylko wtedy, gdy jest wybrana zmienna **Analiza skumulowana**.

Skumulowane wielkości efektów

Zapisuje szacowaną skumulowaną łączną wielkość efektów.

Skumulowana wielkość efektu (postać wykładnicza)

Zapisuje szacowaną łączną wielkość efektu skumulowanego w formie wykładniczej.

Błąd standardowy

Zapisuje szacowany błąd standardowy skumulowanego ogólnego rozmiaru efektu.

Dolna granica przedziału ufności

Zapisuje dolną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Górna granica przedziału ufności

Zapisuje górną granicę szacowanego przedziału ufności skumulowanej łącznej wielkości efektów.

Dolna granica przedziału ufności (postać wykładnicza)

Zapisuje dolną granicę przedziału szacowanego przedziału ufności ogólnej wielkości skumulowanych efektów w formie wykładniczej.

Górna granica przedziału ufności (forma wykładnicza)

Zapisuje górną granicę przedziału szacowanego przedziału ufności ogólnej wielkości skumulowanych efektów w formie wykładniczej.

Wartość p

Zapisuje szacowaną wartość p skumulowanej łącznej wielkości efektów.

Miejsce docelowe

Udostępnia opcje umożliwiające określenie zbioru danych lub pliku danych, w którym będzie dokonywany zapis. Po wybraniu opcji **Zbiór danych** można określić nową nazwę zbioru danych (można zachować domyślną nazwę zbioru danych). Po wybraniu opcji **Plik danych** kliknij przycisk **Przełóżaj...**, aby wybrać nazwę i położenie pliku.

Definiowanie ustawień analizy dla metaanalizy binarnej wielkości efektu

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza binarna wielkości efektu** kliknij opcję **Analiza**.
3. Wybierz i zdefiniuj odpowiednie ustawienia analizy.
4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna wielkości efektu: Wnioskowanie

Okno dialogowe **Wnioskowanie** zawiera ustawienia dotyczące określania metod szacowania dla metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Estymator

Udostępnia ustawienia dotyczące określania estymatora.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje estymatorem używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, kontroluje również estymator używany przez test oparty na regresji.

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Największa wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia, które kontrolują, czy ma być stosowane dopasowanie błędu standardowego Knappa-Hartunga.

Uwagi:

- Jeśli określono ustawienia Trim-and-Fill, również steruje dopasowaniem błędu standardowego używanym przez zestawianie w analizie obcinania i wypełniania.
- Jeśli określone są ustawienia obciążeń, również steruje dopasowaniem błędu standardowego używanym przez test oparty na regresji.

Brak korekt

Ustawienie domyślne, dopasowanie nie jest stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę dopasowania Knappa-Hartunga.

Zastosuj obcięta korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas estymowania macierzy wariancji-kowariancji.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie): Wnioskowanie

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Wnioskowanie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wnioskowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza wielkości efektu (binarnie): Kontrast

Okno dialogowe **Kontrast** zawiera ustawienia do sterowania testem kontrastu dla metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Zmienne

Lista zawiera wszystkie zmienne dostępne w zbiorze danych. Wybierz zmienne z listy i przenieś je na listę **Kontrasty**.

Kontrasty

Ta lista zawiera współczynniki, które są zapisywane jako zmienne w aktywnym zbiorze danych. Dozwolonych jest wiele zmiennych. Zmienne łańcuchowe nie są obsługiwane.

Współczynniki wprowadzane przez użytkownika

Udostępnia ustawienia służące do określania współczynników kontrastu wprowadzanych przez użytkownika. Dozwolone są tylko wartości liczbowe. Aby można było sformułować poprawny test kontrastu, liczba określonych wartości musi odpowiadać liczbie ważnych badań.

Wyświetl statystyki wykładnicze

Służy do sterowania uwzględnianiem statystyk w postaci wykładniczej. Gdy to ustawienie jest wybrane, wyniki zawierają statystyki w postaci wykładniczej. Dotyczy to m.in. wielkość efektów i granic przedziałów ufności. Ustawienie jest dostępne, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto** lub **Logarytmiczny iloraz ryzyka**.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie): Kontrast

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Kontrast**.

3. Wybierz i zdefiniuj odpowiednie ustawienia Kontrast.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza wielkości efektu (binarnie): Odchylenie

W oknie dialogowym **Odchylenie** dostępne są ustawienia umożliwiające określenie odchylenia publikacji poprzez przeprowadzenie testów opartych na regresji dla metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Test Eggera oparty na regresji

Wybranie tego ustawienia umożliwia obliczenie obciążenia publikacji poprzez test Eggera oparty na regresji.

Zmienne

Lista zawiera wszystkie dostępne zmienne zbioru danych.

Współzmiennie

Zmienne wybrane z listy **Zmienne** są traktowane jako współzmiennie. Dozwolonych jest wiele współzmiennych.

Czynniki

Zmienne wybrane z listy **Zmienne** są traktowane jako czynniki. Dozwolonych jest wiele czynników.

Uwzględnij wyraz wolny w regresji

Steruje wyrazem wolnym w teście opartym na regresji.

Uwzględnij parametr dyspersji w modelu efektów stałych

Steruje ustawieniem modelu multiplikatywnego i wprowadza parametr dyspersji multiplikatywnej do analizy. To ustawienie jest dostępne tylko wtedy, gdy wybrany jest model efektów stałych.

Szacowana statystyka oparta na rozkładzie *t*

Steruje dystrybucją użytą w testach opartych na regresji. Ustawienie jest domyślnie włączone, co pozwala oszacować statystyki oparte na rozkładzie *T*. Jeśli to ustawienie nie jest zaznaczone, statystyki są szacowane na podstawie rozkładu normalnego.

Definiowanie ustawień odchylenia metaanalizy wielkości efektu (binarnie)

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Odchylenie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia obciążenia.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna wielkości efektu: Trim-and-Fill

W oknie dialogowym **Trim-and-Fill** dostępne są ustawienia implementacji analizy metodą Trim-and-Fill dla obciążenia publikacji w przypadku metaanalizy z wynikami binarnymi, gdy w aktywnym zbiorze danych zostaną udostępnione dane wielkości efektu w aktywnym zbiorze danych.

Szacunkowa liczba brakujących badań

Kontroluje analizę metodą Trim-and-Fill dla złudzenia publikacyjnego. Wybranie tego ustawienia powoduje włączenie innych ustawień okna dialogowego.

Strona do podstawiania badań

Udostępnia opcje służące do określania strony wykresu lejkowego, na którym podstawiane są brakujące badania.

Ustalone przez zbocza próby Eggera

Ustawienie domyślne określa stronę w oparciu o szacowane nachylenie testu Eggera.

Lewo

Podstawia lewą stronę wykresu lejkowego.

Prawo

Podstawia prawą stronę wykresu lejkowego.

Metoda

Określa metodę szacowania liczby brakujących badań.

Linear

Ustawienie domyślne oblicza wartość estymatora liniowego.

Uruchom

Służy do obliczania estymatora przebiegu.

Kwadratowy

Służy do obliczania estymatora kwadratowego.

Proces iteracji

Udostępnia ustawienia służące do określania estymatora iteracji i dopasowania błędu standardowego.

Model efektów stałych

Po wybraniu tej opcji stosowany jest model efektów stałych, a estymacja iteracji i dopasowanie błędu standardowego nie będą dostępne.

Model efektów losowych

Po wybraniu tej opcji stosowany jest model efektów losowych i dostępne są następujące ustawienia.

Estymator

Udostępnia ustawienia dotyczące określania estymatora iteracji

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Maksymalna wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Dopasowanie błędu standardowego

Udostępnia ustawienia służące do kontrolowania, czy do iteracji algorytmu metodą Trim-and-Fill ma zostać zastosowane dopasowanie błędu standardowego Knappa-Hartunga.

Brak korekt

Ustawienie domyślne: korekty nie są stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę korekty Knappa-Hartunga.

Zastosuj obciętą korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas szacowania macierzy wariancji-kowariancji.

Definiowanie ustawień drukowania metaanalizy binarnej wielkości efektu

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza binarna wielkości efektu** kliknij opcję **Drukuj**.

3. Wybierz i zdefiniuj odpowiednie ustawienia metody Trim-and-Fill.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza binarna wielkości efektu: Drukowanie

W oknie dialogowym **Drukuj** dostępne są ustawienia sterujące danymi wyjściowymi tabeli w celu przeprowadzenia metaanalizy z wynikami binarnymi, gdy w aktywnym zestawie danych zostaną wstępnie obliczone dane o wielkości efektu.

Homogeniczność/Heterogeniczność

Udostępnia ustawienia sterujące testami homogeniczności i heterogeniczności.

Testowanie homogeniczności

Po wybraniu tej opcji w wynikach będzie się znajdować odpowiedni test homogeniczności.

Miary heterogeniczności

Jeśli ta opcja zostanie wybrana, w wynikach będą się znajdować miary heterogeniczności.

Wielkości efektów

Udostępnia następujące ustawienia wielkości efektu.

Badania indywidualne

Steruje wyświetlaniem poszczególnych badań. Po wybraniu tej opcji udostępniane są odpowiednie wyniki.

Skumulowane wielkości efektów

Steruje wyświetlaniem zbiorczej analizy. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy w oknie dialogowym **Analiza** zostanie wybrana zmienna **Analiza skumulowana**.

Przedział predykcji w modelu efektów losowych

Steruje wyświetlaniem przedziału predykcji. Po wybraniu tej opcji udostępniane są odpowiednie wyniki. Ustawienie jest dostępne tylko wtedy, gdy określony jest model efektów losowych.

Wyświetl statystyki wykładnicze

Steruje włączeniem statystyk wykładniczych. Po wybraniu ustawienia statystyki wykładnicze, w tym wykładnicza wielkość efektu i limity przedziału ufności, są umieszczane w wynikach. Proces ma zastosowanie zarówno do wymaganych, jak i opcjonalnych tabel szacowania i predykcji wielkości efektu. Ustawienie jest dostępne, gdy **Wielkość efektu** jest określona jako **Logarytmiczny iloraz szans**, **Logarytmiczny iloraz szans Peto** lub **Logarytmiczny iloraz ryzyka**.

Definiowanie ustawień drukowania metaanalizy binarnej wielkości efektu

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza binarna wielkości efektu** kliknij opcję **Drukuj**.

3. Wybierz i zdefiniuj odpowiednie ustawienia drukowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza wielkości efektu (binarnie): Zapisz

Okno dialogowe **Zapisz** udostępnia ustawienia służące do zapisywania oszacowanych statystyk w aktywnym zbiorze danych dla metaanalizy z wynikami binarnymi, gdy dane o wstępnie obliczonej wielkości efektu zostaną udostępnione w aktywnym zbiorze danych.

Badania indywidualne

Udostępnia opcje zapisywania szacowanej wielkości efektu.

Indywidualna wielkość efektu (forma wykładnicza)

Zapisuje szacowaną wielkość efektu w postaci wykładniczej.

Błąd standardowy

Zapisuje szacowany błąd standardowy wielkości efektu.

Dolna granica przedziału ufności

Zapisuje szacowaną dolną granicę przedziału ufności dla wielkości efektu.

Górna granica przedziału ufności

Zapisuje szacowaną górną granicę przedziału ufności dla wielkości efektu.

Dolna granica przedziału ufności (forma wykładnicza)

Umożliwia zapisanie dolnego limitu szacowanego przedziału ufności dla wielkości efektu w postaci wykładniczej.

Górna granica przedziału ufności (forma wykładnicza)

Umożliwia zapisanie górnego limitu szacowanego przedziału ufności dla wielkości efektu w postaci wykładniczej.

Wartość p

Zapisuje szacowaną wartość p - wartość wielkości efektu.

Waga badania

Zapisuje szacowaną wagę badania.

Procent wagi badania

Zapisuje znormalizowaną wagę badania jako wartość procentową.

Definiowanie ustawień dla procedury Metaanaliza wielkości efektu (binarnie): Zapisywanie

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki binarne > Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza wielkości efektu (binarnie)** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia zapisywania szacowanych statystyk.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza wielkości efektu binarnego: Wykres

Okno dialogowe **Wykres** udostępnia ustawienia dla następujących typów wykresu:

- [“Wykres lasu” na stronie 75](#)
- [“Skumulowany wykres lasu” na stronie 77](#)
- [“Wykres bąbelkowy” na stronie 78](#)
- [“Wykres lejkowy” na stronie 79](#)
- [“Wykres Galbraitha” na stronie 80](#)
- [“Wykres L' Abba” na stronie 81](#)

Regresja metaanalizy

Procedura regresji metaanalizy przeprowadza analizę metaregresji.

Przykład

W historii przeprowadzono kilka badań w celu zbadania skuteczności pewnego leku w leczeniu cukrzycy typu II. Twierdzono, że ten lek doustny ten jest w stanie zmniejszyć poziom glukozy we krwi po posiłkach. Dane zostały zebrane z różnych ośrodków badawczych w latach 1979-1986.

Główny badacz chciał wyciągnąć statystyczny wniosek o skutek stosowania leku w postaci doustnej. Ze względu na fakt, że dane zostały wygenerowane z różnych badań, zaproponował sposób syntezy wyników z wielu badań, aby zrozumieć ogólny efekt i zidentyfikować te bazowe źródła zmian w wynikach.

Statystyki

Poziom ufności, metoda iteracyjna, stopniowanie, tolerancja zbieżności, średnie przykładowe, wariancja próby, odchylenie standardowe, szacowana wielkość efektu, metoda estymacji, test oparty na regresji, model efektów losowych, model efektów statycznych, parametr dyspersyjny, estymator z ograniczonymi maksymalnymi prawdopodobieństwem, empiryczny estymator Bayesa, estymator Hedgesa, estymator Huntera-Schmidta, estymator DerSimoniana-Lairda, estymator Sidika-Jonkmana, korekcja błędów standardowych Knappa-Hartunga, obcięta korekcja błędów standardowych Knappa-Hartunga, współczynniki, test współczynników modelu, statystyki wykładnicze.

Uzyskiwanie metaanalizy metodą regresji

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Metaregresja

2. Wybierz pojedynczą zmienną zależną **Wielkość efektu**, która oznacza wielkość efektu. Wybrana zmienna musi być zmienną numeryczną (zmiennie łańcuchowe nie są obsługiwane).

3. Wybierz jedno z następujących ustawień, następnie wybierz odpowiednią pojedynczą zmienną numeryczną:

Błąd standardowy

Wybierz zmienną, która określa błąd standardowy przekształcany w wagę. To jest ustawienie domyślne.

Wariancja

Wybierz zmienną, która określa wariancję przekształcaną w wagę.

Waga

Wybierz zmienną, która określa wagę.

4. Opcjonalnie dodaj zmienne czynnika do listy **Czynniki**. Dla każdej wybranej zmiennej czynnika można określić opcjonalną wartość dla każdej zmiennej, aby wyznaczyć niestandardowe **Ostatnie kategorie**. Aby przywrócić domyślne ustawienia **Ostatnie kategorie**, można kliknąć przycisk **Resetuj**.

Uwaga: Jeśli nie istnieją żadne obserwacje zgodne z podanymi wartościami **Ostatnie kategorie**, ostatnie występujące wartości są traktowane jako ostatnie kategorie.

5. Opcjonalnie wybierz numeryczne współzmiennie.
6. Opcjonalnie wybierz ustawienie **modelu**.

Efekty losowe

Ustawienie domyślne buduje model efektów losowych.

Efekty stałe

Buduje model efektów stałych. Opcjonalnie można wybrać ustawienie **Uwzględnij parametr dyspersji**.

7. Opcjonalnie można wykonać następujące czynności:
- Kliknij przycisk **Kryteria...**, aby określić kryteria ogólne.
 - Kliknij przycisk **Wnioskowanie**, aby określić metody szacowania.
 - Kliknij przycisk **Drukuj**, aby kontrolować wyniki tabeli.
 - Kliknij przycisk **Zapisz**, aby przewidzieć i zapisać szacowaną statystykę dla aktywnego zbioru danych.
 - Kliknij przycisk **Wykres**, aby określić, które wykresy mają być uwzględnione w danych wyjściowych.
8. Kliknij **OK**.

Regresja metaanalizy: Kryteria

Okno dialogowe **Kryteria** udostępnia ustawienia dotyczące określania kryteriów dla regresji metaanalizy.

Przedział ufności

To ustawienie opcjonalne określa poziom ufności. Wartość musi być wartością liczbową z zakresu od 0 do 100. Ustawieniem domyślnym jest 95.

Braku danych zdefiniowane przez użytkownika

Ustawienia umożliwiają użytkownikowi określenie sposobu obsługi brakujących danych.

Wyklucz

W przypadku ustawienia domyślnego brakujące wartości użytkownika są traktowane jako poprawne.

Dołącz

Ignoruje nazwy brakujących wartości użytkownika i traktuje brakujące wartości użytkownika jako poprawne.

Iteracje

Maksymalna liczba iteracji

Ustawienie opcjonalne określa maksymalną liczbę iteracji w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 100. Wartość 0 oznacza, że żadne iteracje nie są wykonywane.

Maksimum kroków połowienia

Ustawienie opcjonalne określa maksymalne połowienie w metodach iteracyjnych. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 5. Wartość 0 oznacza, że połowienie nie jest stosowane.

Zbieżność

Ustawienie opcjonalne określa tolerancję zbieżności. Wartość musi być dodatnią wartością. To ustawienie jest dostępne, gdy używane są metody iteracyjne. Wartością domyślną jest $1E-6$.

Definiowanie kryteriów regresji metaanalizy

1. Z menu wybierz:

Analizuj > Metaanaliza > Metaregresja

2. W oknie dialogowym **Regresja metaanalizy** kliknij opcję **Kryteria**.

3. Wybierz i zdefiniuj odpowiednie ustawienia kryteriów.

4. Kliknij przycisk **Kontynuuj**.

Regresja metaanalizy: wnioskowanie

Okno dialogowe **Wnioskowanie** udostępnia ustawienia służące do określania metod estymacji dla regresji metaanalizy.

Uwzględnij wyraz wolny w regresji

Ustawienie jest włączone domyślnie.

Szacowana statystyka oparta na rozkładzie t

Steruje dystrybucją użytą w testach opartych na regresji. Ustawienie jest domyślnie włączone, co pozwala oszacować statystyki oparte na rozkładzie T . Jeśli to ustawienie nie jest zaznaczone, statystyki są szacowane na podstawie rozkładu normalnego.

Estymator

Udostępnia ustawienia dotyczące określania estymatora.

Warunkowa największa wiarygodność (REML)

Ustawienie domyślne stosuje metodę iteracyjną i oblicza wartość ograniczonego estymatora maksymalnej wiarygodności.

Największa wiarygodność (ML)

Stosuje metodę iteracyjną i oblicza estymator maksymalnej wiarygodności.

Empiryczny Bayesa

Stosuje metodę iteracyjną i oblicza empiryczny estymator Bayesa.

Hedgesa

Stosuje metodę nieiteracyjną i oblicza estymator Hedgesa.

Huntera-Schmidta

Stosuje metodę nieiteracyjną i oblicza estymator Huntera-Schmidta.

DerSimoniana-Lairda

Stosuje metodę nieiteracyjną i oblicza estymator DerSimoniana-Lairda.

Sidika-Jonkmana

Stosuje metodę nieiteracyjną i oblicza estymator Sidika-Jonkmana.

Standardowa regulacja błędu

Udostępnia ustawienia, które kontrolują, czy ma być zastosowana standardowa regulacja błędu Knapp-Hartung.

Brak korekt

Ustawienie domyślne, dopasowanie nie jest stosowane.

Zastosuj korektę Knappa-Hartunga

Stosuje metodę dopasowania Knappa-Hartunga.

Zastosuj obciążoną korektę Knappa-Hartunga

Stosuje metodę Knappa-Hartunga i obcina wartość mniejszą od 1 podczas estymowania macierzy wariancji-kowariancji.

Definiowanie ustawień wnioskowania regresji metaanalizy

1. Z menu wybierz:

Analizuj > Metaanaliza > Metaregresja

2. W oknie dialogowym **Regresja metaanalizy** kliknij opcję **Wnioskowanie**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wnioskowania.

4. Kliknij przycisk **Kontynuuj**.

Regresja metaanalizy: Drukowanie

W oknie dialogowym **Drukuj** dostępne są ustawienia sterujące wynikami tabel dla regresji metaanalizy.

Test współczynnika modelu

Steruje testem współczynnika modelu. Domyślnie to ustawienie nie jest określone, co powoduje, że test jest pomijalny. Po określeniu ustawienia test jest udostępniany w danych wyjściowych.

Wyświetl statystyki wykładnicze

Steruje oszacowaniami parametrów. Domyślnie to ustawienie nie jest określone, co powoduje, że oszacowania parametrów są pomijalne. Po określeniu ustawienia oszacowania parametrów są udostępniane w danych wyjściowych.

Definiowanie ustawień drukowania regresji metaanalizy

1. Wybierz z menu następującą opcję:

Analizuj > Metaanaliza > Metaregresja

2. W oknie dialogowym **Regresja metaanalizy** kliknij opcję **Drukuj**.

3. Wybierz i zdefiniuj odpowiednie ustawienia drukowania.

4. Kliknij przycisk **Kontynuuj**.

Metaanaliza: regresja: Zapisz

Okno dialogowe **Zapisz** udostępnia ustawienia służące do zapisywania oszacowanych statystyk w aktywnym zbiorze danych dla procedury Metaanaliza: regresja.

Badania indywidualne

Udostępnia opcje zapisywania szacowanej wielkości efektu.

Wartości przewidywane

Prognozuje i zapisuje dopasowane wartości.

Błąd standardowy wartości przewidywanych

Prognozuje i zapisuje błąd standardowy dopasowanych wartości.

Dolna granica przedziału ufności

Zapisuje dolną granicę szacowanego przedziału ufności dopasowanych wartości.

Górna granica przedziału ufności

Zapisuje górną granicę szacowanego przedziału ufności dopasowanych wartości.

Reszty

Prognozuje i zapisuje reszty.

Błąd standardowy średniej reszt

Umożliwia predykcje i zapisuje standardowy błąd reszt.

Wpływy

Prognozuje i zapisuje wpływy.

Statystyka skumulowana

Prognozuje i zapisuje predykcję liniową.

Uwaga: Ustawienia statystyk skumulowanych są ignorowane, gdy jako **Model** wybrano **Efekty stałe**.

Stałe predykcje liniowe

Prognozuje i zapisuje predykcję liniową.

Błąd standardowy stałych predykcji liniowych

Prognozuje i zapisuje błąd standardowy predykcji liniowej.

Najlepsze liniowe predykcje nieobciążone

Prognozuje i zapisuje najlepszą liniową nieobciążoną predykcję efektów losowych.

Błąd standardowy średniej NLPN

Przewiduje i zapisuje błąd standardowy najlepszej liniowej nieobciążonej predykcji efektów losowych.

Definiowanie ustawień dla procedury Metaanaliza: regresja: Zapisz

1. Z menu wybierz:

Analizuj > Metaanaliza > Metaregresja

2. W oknie dialogowym **Metaanaliza: regresja** kliknij opcję **Zapisz**.

3. Wybierz i zdefiniuj odpowiednie ustawienia zapisywania.

4. Kliknij przycisk **Kontynuuj**.

Regresja metaanalizy: Wykres

Okno dialogowe **Wykres** udostępnia ustawienia dla definiowania wykresów bąbelkowych. Więcej informacji na ten temat zawiera sekcja [“Wykres bąbelkowy”](#) na stronie 78.

Opcje wykreślania metaanalizy

W oknie dialogowym **Wykres** dla procedur metaanalizy dostępne są następujące opcje wykresu.

Wykres lasu

Karta **Wykres lasu** udostępnia ustawienia dotyczące sterowania wykresami lasu, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zestawie danych.

Wykres lasu

Ustawienie opcjonalne włącza i wyłącza dane wyjściowe wykresu lasu. To ustawienie jest domyślnie wyłączone.

Wyświetl kolumny

Ustawienia opcjonalne określają szacowaną statystykę, która jest umieszczana w tabeli.

Wielkości efektu

Uwzględnia szacowaną wielkość efektu.

Błąd standardowy

Uwzględnia standardowy błąd szacowanej wielkości efektu.

Limity przedziału ufności

Uwzględnia dolne i górne granice przedziału ufności dla szacowanej wielkości efektu.

Wartość p

Zawiera wartość p szacowanej wielkości efektu dla każdego pojedynczego badania.

Waga

Obejmuje wagę badań podstawowych.

Wyświetl postać wykładniczą

Uwaga: To ustawienie ma zastosowanie tylko do metaanalizy z wynikami binarnymi.

Steruje kolumną wykresu oraz wyświetlaniem wielkości efektu i przedziału ufności. Po wybraniu ustawienia wykreśla iloraz szans, iloraz szans Peto lub iloraz ryzyka, a także umieszcza w tabeli statystyki wykładnicze. Po wybraniu ustawienia wykreśla iloraz szans, iloraz szans Peto lub iloraz ryzyka, a także umieszcza w tabeli przekształcone statystyki logarytmiczne.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Kolumny dodatkowe

Opcjonalnie można wybrać zmienne, które mają być wyświetlane jako dodatkowe kolumny. Kolejność zmiennych określa kolejność wyświetlania kolumn.

Sortuj

Opcjonalnie określ zmienną, według której wykres lasu jest sortowany. Jeśli zmienna jest określona, wykres jest sortowany w porządku **rosnącym** (ustawienie domyślne). Po wybraniu opcji **Malejąco** wykres lasu jest sortowany w porządku malejącym.

Pozycja kolumny wykresu

Steruje umieszczeniem kolumny wykresu.

Prawo

Ustawienie domyślne powoduje umieszczenie kolumny wykresu po prawej stronie innych kolumn tabeli.

Lewo

Umieszcza kolumnę wykresu po lewej stronie innych kolumn tabeli.

Linie odniesienia

Steruje liniami odniesienia, które są dodawane do wykresu lasu. Opcja **Całkowita wielkość efektu** dodaje wiersz, aby wskazać szacowaną ogólną wielkość efektu. Opcja **Pusta wielkość efektu** dodaje wiersz, aby wskazać pustą wielkość efektu.

Adnotacje

Steruje wyświetlaniem adnotacji.

Jednorodność

Wyświetla statystykę testu jednorodności.

Heterogeniczność

Wyświetla statystykę testu heterogeniczności.

Testuj

Umożliwia wyświetlenie testu całkowitej wielkości efektu oraz testu jednorodności między podgrupami w obecności analizy podgrupy.

Zakres przycięcia

Określa zakres przycięcia. Po wybraniu ustawienia należy określić dwie wartości liczbowe i spełnić następujące warunki: Górna granica > Dolna granica.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** jest ustawione na wartość **Różnica ryzyka**, musi być również spełniony warunek Górna granica ≤ 1 i Dolna granica ≥ -1 . Jeśli wybrano ustawienie **Wyświetl postać wykładniczą**, musi być również spełniony warunek Dolna granica ≥ 0 .

Definiowanie ustawień wykresu lasu

1. Z menu wybierz:

Analizuj > Metaanaliza > Wyniki ciągłe lub wyniki binarne > Dane surowe... lub Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykres**, następnie kliknij kartę **Wykres lasu**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu lasu.

4. Kliknij przycisk **Kontynuuj**.

Skumulowany wykres lasu

Karta **Wykres lasu** udostępnia ustawienia dotyczące sterowania wykresami lasu, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zestawie danych.

Skumulowany wykres lasu

Ustawienie opcjonalne włącza i wyłącza dane wyjściowe wykresu lasu. To ustawienie jest domyślnie wyłączone.

Wyświetl kolumny

Ustawienia opcjonalne określają szacowaną statystykę, która jest umieszczana w tabeli.

Skumulowane wielkości efektów

Obejmuje skumulowany ogólny rozmiar efektu.

Błąd standardowy

Obejmuje błąd standardowy skumulowanej ogólnej wielkości efektu.

Limity przedziału ufności

Uwzględnia dolne i górne granice przedziału ufności dla szacowanej wielkości efektu.

Wartość p

Zawiera wartość p szacowanej wielkości efektu dla każdego pojedynczego badania.

Wyświetl postać wykładniczą

Uwaga: To ustawienie ma zastosowanie tylko do metaanalizy z wynikami binarnymi.

Steruje kolumną wykresu oraz wyświetlaniem wielkości efektu i przedziału ufności. Po wybraniu ustawienia rysuje iloraz szans, iloraz szans Peto lub iloraz ryzyka, a także umieszcza w tabeli statystyki w postaci wykładniczej. Po wybraniu ustawienia rysuje iloraz szans, iloraz szans Peto lub iloraz ryzyka, a także umieszcza w tabeli przekształcone statystyki logarytmiczne.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Kolumny dodatkowe

Opcjonalnie można wybrać zmienne, które mają być wyświetlane jako dodatkowe kolumny. Kolejność zmiennych określa kolejność wyświetlania kolumn.

Pozycja kolumny wykresu

Steruje umieszczeniem kolumny wykresu.

Prawo

Ustawienie domyślne powoduje umieszczenie kolumny wykresu po prawej stronie innych kolumn tabeli.

Lewo

Umieszcza kolumnę wykresu po lewej stronie innych kolumn tabeli.

Zakres przycięcia

Określa zakres przycięcia. Po wybraniu ustawienia należy określić dwie wartości liczbowe i spełniać następujące warunki: Górna granica > Dolna granica.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** jest ustawione na wartość **Różnica ryzyka**, musi być również spełniony warunek Górna granica ≤ 1 i Dolna granica ≥ -1 . Jeśli wybrano ustawienie **Wyświetl postać wykładniczą**, musi być również spełniony warunek Dolna granica ≥ 0 .

Definiowanie skumulowanych ustawień wykresu lasu

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe lub wyniki binarne > Dane surowe... lub Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykres**, następnie kliknij kartę **Wykres lasu**.
3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu lasu.
4. Kliknij przycisk **Kontynuuj**.

Wykres bąbelkowy

Karta **Wykres bąbelkowy** udostępnia ustawienia dotyczące sterowania wykresami bąbelkowymi, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zbiorze danych.

Uwaga: Wykres bąbelkowy jest jedynym dostępnym typem wykresu dla analizy regresji metaanalizy.

Wykres bąbelkowy

Ustawienie opcjonalne włącza i wyłącza dane wyjściowe wykresu bąbelkowego. To ustawienie jest domyślnie wyłączone.

Uwaga: To ustawienie nie jest dostępne dla analizy regresji metaanalizy.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Predyktory

Określ zmienne, które są traktowane jako predyktory ciągłe na osi x. Przy tworzeniu każdego osobnego wykresu bąbelkowego dozwolone jest użycie wielu zmiennych.

Uwaga: W procedurze Metaanaliza: regresja to ustawienie ma etykietę **Predyktory wykresu bąbelkowego** i umożliwia wybór współzmiennych, dla których mają zostać utworzone wykresy bąbelkowe.

Predyktory centrowane średnią

Ustawienie opcjonalne umożliwiające sterowanie predyktorem ciągłym na osi x. To ustawienie jest domyślnie wyłączone, zatem predyktor nie jest centrowany. Gdy to ustawienie jest włączone, predyktor jest centrowany średnią.

Etykieta

Opcjonalnie można określić zmienną, według której będą tworzone etykiety na wykresie bąbelkowym. Lista **Rozmieszczanie** zawiera opcje służące do automatycznego wprowadzania etykiet (**Automatycznie**) lub **Do prawej**, **Do lewej**, **Górni** **Do dołu**.

Rysuj bąbelki proporcjonalne do wag

Określa sposób rysowania bąbelków w odniesieniu do wag. Ustawienie jest domyślnie włączone, co powoduje, że bąbelki są proporcjonalne do wag. Gdy to ustawienie jest wyłączone, wszystkie bąbelki mają tę samą wielkość.

Wyświetl krzywą dopasowaną

Steruje wyświetlaniem dopasowanych linii regresji. Ustawienie jest włączone domyślnie, co powoduje dodanie dopasowanego wiersza regresji do wykresu bąbelkowego. Gdy ustawienie jest wyłączone, dopasowana linia regresji nie jest dodawana do wykresu bąbelkowego.

Wyświetl granice przedziału ufności

Steruje wyświetlaniem granic przedziału ufności. Ustawienie jest domyślnie włączone, co powoduje dodanie zakresu przedziału ufności do wykresu bąbelkowego. Gdy ustawienie jest wyłączone, granice przedziału ufności nie są dodawane do wykresu bąbelkowego.

Zakres osi x

Określa zakres wykresu osi x. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: Górna granica > Dolna granica. Należy pamiętać, że podane wartości mają zastosowanie do wszystkich wykresów bąbelkowych.

Zakres osi Y

Określa zakres wykresu osi Y. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: Górna granica > Dolna granica. Należy pamiętać, że podane wartości mają zastosowanie do wszystkich wykresów bąbelkowych.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** ma wartość **Różnica ryzyka**, musi być również spełniony warunek Górna granica ≤ 1 i Dolna granica ≥ -1 .

Definiowanie ustawień wykresu bąbelkowego

1. Wybierz z menu następującą opcję:

Analiza > Metaanaliza > Wyniki ciągłe lub wyniki binarne > Dane surowe... lub **Wstępnie obliczona wielkość efektu...**

lub

Analizuj > Metaanaliza > Metaregresja

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykres**, a następnie kartę **Wykres bąbelkowy**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu bąbelkowego.

4. Kliknij przycisk **Kontynuuj**.

Wykres lejkowy

Karta **Wykres lejkowy** udostępnia ustawienia dotyczące sterowania wykresami lejkowymi, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zbiorze danych.

Wykres lejkowy

Ustawienie opcjonalne włącza i wyłącza dane wyjściowe wykresu lejkowego. To ustawienie jest domyślnie wyłączone.

Wartości osi Y

Kontroluje wartości osi Y.

Błąd standardowy

Wykreśla błąd standardowy.

Odwrotny błąd standardowy

Wykreśla odwrotny błąd standardowy.

Wariancja

Wykreśla wariancję.

Wariancja odwrotna

Wykreśla wariancję odwrotną.

Zakres osi x

Określa zakres wykresu osi x. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: Górna granica $>$ Dolna granica. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów lejkowych.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** ma wartość **Różnica ryzyka**, musi być również spełniony warunek Górna granica ≤ 1 i Dolna granica ≥ -1 .

Zakres osi Y

Określa zakres wykresu osi Y. Po wybraniu tego ustawienia muszą zostać określone dwie wartości liczbowe spełniające następujące warunki: Górna granica $>$ Dolna granica. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów lejkowych.

Uwzględnij badania podstawiane z metodą trim-and-fill

Określa, czy badania podstawiane są uwzględniane na wykresie lejkowym, gdy ustawienie **Szacunkowa liczba brakujących badań** jest włączone w oknie dialogowym **Trim-and-Fill**. Ustawienia są domyślnie wyłączone, co powoduje wykluczenie badań podstawianych. Jeśli wybrana jest opcja **Wyświetl ogólną wielkość efektów obserwowanych badań** jest wybrana, zostaje dodana pionowa linia odniesienia oznaczająca szacowaną ogólną wielkość efektu. Ustalenie to uwzględnia zarówno badania obserwowane, jak i podstawiane.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Etykieta

Opcjonalnie określ zmienną, na podstawie której wykres lejkowy jest sortowany. Lista **Rozmieszczanie** zawiera opcje służące do automatycznego wprowadzania etykiet (**Automatycznie**) lub **Do prawej, Do lewej, Górna i Do dołu**.

Definiowanie ustawień wykresu lejkowego

1. Z menu wybierz:

Analiza > Metaanaliza > Wyniki ciągłe lub wyniki binarne > Dane surowe... lub Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykres**, a następnie kliknij kartę **Wykres lasu**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu lejkowego.

4. Kliknij przycisk **Kontynuuj**.

Wykres Galbraitha

Karta **Wykres Galbraitha** udostępnia ustawienia dotyczące sterowania wykresami Galbraitha, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zestawie danych.

Wykres Galbraitha

Ustawienie opcjonalne włącza i wyłącza dane wyjściowe wykresu bąbelkowego. To ustawienie jest domyślnie wyłączone.

Wyświetl granice przedziału ufności

Steruje wyświetlaniem granic przedziału ufności. Ustawienie jest domyślnie włączone, co powoduje dodanie zakresu przedziału ufności do wykresu bąbelkowego. Gdy ustawienie jest wyłączone, granice przedziału ufności nie są dodawane do wykresu bąbelkowego.

Zakres osi x

Określa zakres wykresu osi x. Po wybraniu tego ustawienia należy określić dwie wartości liczbowe spełniające następujące warunki: **Górna granica > Dolna granica**. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów Galbraitha.

Zakres osi Y

Określa zakres wykresu osi Y. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: **Górna granica > Dolna granica**. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów Galbraitha.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Etykieta

Opcjonalnie można określić zmienną, według której będą tworzone etykiety na wykresie bąbelkowym. Lista **Rozmieszczanie** zawiera opcje służące do automatycznego wprowadzania etykiet (**Automatycznie**) lub **Do prawej, Do lewej, Górna i Do dołu**.

Definiowanie ustawień wykresu Galbraitha

1. Wybierz z menu następujące opcje:

Analiza > Metaanaliza > Wyniki ciągłe lub wyniki binarne > Dane surowe... lub Wstępnie obliczona wielkość efektu...

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykres**, a następnie kartę **Wykres Galbraitha**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu Galbraitha.

4. Kliknij przycisk **Kontynuuj**.

Wykres L' Abba

Karta **Wykres L'Abba** udostępnia ustawienia dotyczące sterowania wykresami L'Abba, które są wyświetlane w danych wyjściowych metaanalizy z wynikami ciągłymi i binarnymi, gdy dane surowe i wstępnie obliczone dane dotyczące wielkości efektu są udostępniane w aktywnym zbiorze danych.

Wykres L' Abb'a

Ustawienie opcjonalne włącza i wyłącza wyjście wykresu L' Abba. To ustawienie jest domyślnie wyłączone.

Rysuje punkty proporcjonalne do wag

Określa sposób rysowania kropek w odniesieniu do wag. Ustawienie jest domyślnie włączone, co powoduje, że kropki są proporcjonalne do wag. Gdy ustawienie jest wyłączone, wszystkie kropki będą rysowane w tej samej wielkości.

Linie referencyjne

Całkowita wielkość efektu

Steruje wyświetlaniem linii odniesienia, która reprezentuje szacowaną ogólną wielkość efektu. Ustawienie jest domyślnie wyłączone, co powoduje, że linia odniesienia jest pomijana. Gdy ustawienie jest włączone, wyświetlana jest linia odniesienia.

Wielkość efektu zerowego

Steruje wyświetlaniem linii referencyjnej reprezentującej brak efektu. Gdy to ustawienie jest wyłączone, linia referencyjna nie jest wyświetlana. Jeśli to ustawienie nie jest włączone, linia referencyjna jest pomijana.

Zakres osi x

Określa zakres wykresu osi x. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: Górna granica > Dolna granica. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów Galbraitha.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** ma wartość **Różnica ryzyka**, musi być również spełniony warunek $\text{Górna granica} \leq 1$ i $\text{Dolna granica} \geq 0$. Gdy w polu **Wielkość efektu** wybrana jest wartość **Logarytmiczny iloraz ryzyka**, musi być spełniony warunek $\text{Górna granica} \leq 0$.

Zakres osi Y

Określa zakres wykresu osi Y. Po wybraniu ustawienia muszą zostać ustawione dwie wartości liczbowe spełniające następujące warunki: Górna granica > Dolna granica. Należy zauważyć, że podane wartości mają zastosowanie do wszystkich wykresów Galbraitha.

W przypadku metaanalizy z wynikami binarnymi, gdy ustawienie **Wielkość efektu** ma wartość **Różnica ryzyka**, musi być również spełniony warunek $\text{Górna granica} \leq 1$ i $\text{Dolna granica} \geq 0$. Gdy w polu **Wielkość efektu** wybrana jest wartość **Logarytmiczny iloraz ryzyka**, musi być spełniony warunek $\text{Górna granica} \leq 0$.

Zmienne

Przedstawia dostępne zmienne ze zbioru danych.

Etykieta

Opcjonalnie można określić zmienną, według której będą tworzone etykiety na wykresie bąbelkowym. Lista **Rozmieszczanie** zawiera opcje służące do automatycznego wprowadzania etykiet (**Automatycznie**) lub **Do prawej**, **Do lewej**, **Górna** i **Do dołu**.

Definiowanie ustawień wykresu L' Abbego

1. Wybierz z menu następującą opcję:

Analiza > Metaanaliza > Wyniki ilościowe lub Wyniki binarne > Dane surowe...

2. W oknie dialogowym **Metaanaliza** kliknij opcję **Wykresy**, a następnie kartę **Wykresy L' Abbego**.

3. Wybierz i zdefiniuj odpowiednie ustawienia wykresu L' Abbego.

4. Kliknij przycisk **Kontynuuj**.

książka kodowa

W książce kodowej znajdują się informacje słownikowe, takie jak nazwy zmiennych, etykiety zmiennych, etykiety wartości, wartości brakujące, a także statystyki podsumowujące dla wszystkich określonych zmiennych oraz zestawy wielokrotnych podpowiedzi w aktywnym zbiorze danych. Statystyki zmiennych nominalnych i porządkowych oraz zestawów wielokrotnych podpowiedzi zawierają licznosci i obliczenia procentowe. Statystyki podsumowujące dla zmiennych ilościowych obejmują średnią, odchylenie standardowe oraz kwartyle.

Uwaga: książka kodowa ignoruje status dzielenia danych na podzbiory. Obejmuje to grupy dzielone na podzbiory, utworzone w celu wielokrotnego podstawienia brakujących wartości (dostępne za pośrednictwem opcji dodatkowej Braki danych).

Aby uzyskać dostęp do książki kodowej:

1. Z menu wybierz:

Analiza > Raporty > Książka kodowa

2. Kliknij kartę Zmienne.

3. Wybierz co najmniej jedną zmienną i/lub zestaw wielokrotnych odpowiedzi.

Opcjonalnie można wykonać następujące czynności:

- Zarządzać wyświetlanymi informacjami o zmiennych.
- Zarządzać wyświetlanymi statystykami (lub wyłączyć wszystkie statystyki podsumowujące).
- Zarządzać kolejnością wyświetlania zmiennych i zestawów wielokrotnych odpowiedzi.
- Zmieniać poziom pomiaru dla dowolnej zmiennej z listy źródłowej w celu dokonania zmiany wyświetlanej statystyki podsumowującej. Aby uzyskać dodatkowe informacje, patrz temat: "[Książka kodowa – karta Statystyki](#)" na stronie 84.

Zmianie poziomu pomiaru

Dla zmiennych można tymczasowo zmienić poziom pomiaru. (Nie można zmieniać poziomu pomiaru w przypadku zestawów wielokrotnych odpowiedzi. Są one zawsze traktowane jako nominalne.)

1. Kliknij prawym przyciskiem myszy zmienną na liście źródłowej.
2. Z menu kontekstowego wybierz poziom pomiaru.

Spowoduje to tymczasową zmianę poziomu pomiaru. W praktyce jest to przydatne wyłącznie w przypadku zmiennych numerycznych. Poziom pomiaru zmiennych łańcuchowych jest ograniczony do nominalnych lub porządkowych, które są traktowane tak samo przez procedurę Książka kodowa.

Książka kodowa – karta Raport

Na karcie Raport można zarządzać informacjami o zmiennych, dołączonymi do każdej zmiennej oraz zestawu wielokrotnych odpowiedzi, kolejnością, w jakiej są wyświetlane zestawy wielokrotnych odpowiedzi, oraz zawartością tabeli opcjonalnych informacji o pliku.

Informacje o zmiennej

W tym obszarze można zarządzać informacjami słownika wyświetlanymi dla każdej zmiennej.

Pozycja. Liczba całkowita, która oznacza położenie zmiennej w porządku pliku. Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Etykieta. Etykieta opisowa przypisana do zmiennej lub zestawu wielokrotnych odpowiedzi.

Typ. Podstawowy typ danych. Jest to typ *Numeryczny*, *Łańcuchowy* lub *Zestaw wielokrotnych odpowiedzi*.

Format. Format wyświetlania zmiennej, taki jak *A4*, *F8.2* lub *DATE11*. Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Poziom pomiaru. Dopuszczalne wartości to: *Nominalny*, *Porządkowy*, *Ilościowy* i *Nieznany*. Wyświetlana wartość to poziom pomiaru zapisany w słowniku. Nie ma na nią wpływu tymczasowe zastąpienie poziomu

pomiaru, dokonane przez zmianę poziomu pomiaru na liście zmiennych źródłowych na karcie Zmienne. Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Uwaga: poziom pomiaru zmiennych numerycznych może być „nieznany” przed pierwszym pobraniem danych, kiedy poziom pomiaru nie został jeszcze wyraźnie określony, np. podczas odczytywania danych ze źródła zewnętrznego lub w przypadku nowych danych. Aby uzyskać dodatkowe informacje, patrz temat .

Role. Niektóre okna dialogowe mają zdolność do wstępnego wybierania zmiennych do analizy w oparciu o zdefiniowane role. .

Etykiety wartości. Etykiety opisowe przypisane do określonych wartości danych. .

- Jeśli na karcie Statystyki jest wybrana opcja Liczność lub Procent, zdefiniowane etykiety wartości są dołączane do raportu, nawet jeśli pole Etykiety wartości nie zostanie tutaj zaznaczone.
- W przypadku zestawów wielokrotnych dychotomii „etykiety wartości” to etykiety zmiennych dla zmiennych elementarnych w zestawie lub etykiety wartości zliczanych, w zależności od tego, jak zdefiniowany jest zestaw. Aby uzyskać dodatkowe informacje, patrz temat .

Braki danych. Zdefiniowane braki danych. . Jeśli na karcie Statystyki jest wybrana opcja Liczność lub Procent, zdefiniowane etykiety wartości są dołączane do raportu, nawet jeśli pole Brakujące wartości nie zostanie tutaj zaznaczone. Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Atrybuty użytkownika. Atrybuty zmiennych zdefiniowane przez użytkownika. W raporcie znajdują się zarówno nazwy, jak i wartości wszystkich niestandardowych atrybutów zmiennych powiązanych z każdą ze zmiennych. Aby uzyskać dodatkowe informacje, patrz temat . Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Atrybuty zarezerwowane. Zarezerwowane atrybuty zmiennych systemowych. Można wyświetlać atrybuty systemowe, ale nie należy ich zmieniać. Nazwy atrybutów systemowych zaczynają się od znaku dolara (\$). Atrybuty inne niż wyświetlane, których nazwy rozpoczynają się od znaku @ lub \$@, nie są uwzględniane. W raporcie znajdują się zarówno nazwy, jak i wartości wszystkich atrybutów systemowych powiązanych z każdą ze zmiennych. Pole to nie jest dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Informacja o pliku

Tabela opcjonalnych informacji o pliku może zawierać dowolny z poniższych atrybutów:

Nazwa pliku. Nazwa pliku danych programu IBM SPSS Statistics. Jeśli zbiór danych nie był nigdy zapisywany w formacie programu IBM SPSS Statistics, nazwa pliku danych nie istnieje. (Jeśli nazwa pliku nie jest wyświetlana na pasku tytułu w oknie Edytora danych, oznacza to, że aktywny zbiór danych nie posiada nazwy pliku).

Lokalizacja. Położenie katalogu (folderu), w którym znajduje się plik danych programu IBM SPSS Statistics. Jeśli zbiór danych nie był nigdy zapisywany w formacie programu IBM SPSS Statistics, położenie nie występuje.

Liczba obserwacji. Liczba obserwacji w aktywnym zbiorze danych. Ogólna liczba obserwacji, łącznie ze wszystkimi obserwacjami, które mogły zostać wyłączone ze statystyk podsumowujących na podstawie ustawień filtrów.

Etykieta. Etykieta pliku (jeśli istnieje), zdefiniowana za pomocą komendy FILE LABEL.

Dokumenty. Tekst dokumentu pliku danych. .

Stan ważenia. Jeśli włączone jest ważenie, wyświetla się nazwa zmiennej ważącej. Aby uzyskać dodatkowe informacje, patrz temat .

Atrybuty użytkownika. Atrybuty pliku danych, zdefiniowane przez użytkownika. Atrybuty pliku danych, zdefiniowane za pomocą komendy DATAFILE ATTRIBUTE.

Atrybuty zarezerwowane. Zarezerwowane systemowe atrybuty pliku danych. Można wyświetlać atrybuty systemowe, ale nie należy ich zmieniać. Nazwy atrybutów systemowych zaczynają się od znaku dolara (\$). Atrybuty inne niż wyświetlane, których nazwy rozpoczynają się od znaku @ lub \$@, nie są uwzględniane. W raporcie znajdują się zarówno dane, jak i wartości wszystkich atrybutów systemowych pliku danych.

Kolejność wyświetlania zmiennych

Dostępne są następujące opcje zarządzania kolejnością wyświetlania zmiennych i zestawów wielokrotnych odpowiedzi.

Alfabetycznie. Kolejność alfabetyczna według nazw zmiennych.

Plik. Kolejność, w jakiej zmienne pojawiają się w zbiorze danych (kolejność, w jakiej są wyświetlane w Edytorze danych). W porządku rosnącym zestawy wielokrotnych odpowiedzi są wyświetlane na końcu, po wszystkich wybranych zmiennych.

Poziom pomiaru. Sortowanie według poziomu pomiaru. Powstają cztery grupy sortowania: nominalna, porządkowa, ilościowa i nieznana. Zestawy wielokrotnych odpowiedzi są traktowane jako poziom nominalny.

Uwaga: poziom pomiaru zmiennych numerycznych może być „nieznany” przed pierwszym pobraniem danych, kiedy poziom pomiaru nie został jeszcze wyraźnie określony, np. podczas odczytywania danych ze źródła zewnętrznego lub w przypadku nowych danych.

Lista zmiennych. Kolejność, w jakiej zmienne i zestawy wielokrotnych odpowiedzi pojawiają się na wybranej liście zmiennych na karcie Zmienne.

Nazwa atrybutu użytkownika. Lista opcji kolejności sortowania zawiera również nazwy wszystkich niestandardowych, zdefiniowanych przez użytkownika atrybutów. W porządku rosnącym zmienne, które nie mają atrybutu, są sortowane na górze; po nich znajdują się zmienne, które posiadają atrybut, ale bez zdefiniowanej wartości; za nimi zmienne o zdefiniowanych wartościach atrybutu w porządku alfabetycznym, według tych wartości.

Maksymalna liczba kategorii

Jeśli w raporcie znajdują się etykiety wartości, licznosci lub wartości procentowe dla każdej unikalnej wartości, można usunąć te informacje z tabeli, jeśli liczba wartości przekracza określoną wielkość. Domyślnie informacje te są usuwane, gdy liczba unikalnych wartości dla zmiennej przekracza 200.

Książka kodowa – karta Statystyki

Dzięki karcie Statystyki można zarządzać statystykami podsumowującymi, które są dołączane do raportu, lub całkowicie wyłączać wyświetlanie statystyk podsumowujących.

Liczebności i procenty

Statystyki dostępne dla zmiennych nominalnych i porządkowych, zestawów wielokrotnych odpowiedzi oraz wartości z etykietami zmiennych ilościowych to:

Liczebność. Liczebność lub liczba obserwacji posiadających daną wartość (lub zakres wartości) zmiennej.

Procent. Odsetek obserwacji posiadających daną wartość.

Tendencja centralna i dyspersja

Statystyki dostępne dla zmiennych ilościowych to:

Średnia. Miara tendencji centralnej. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.

Odchylenie standardowe. Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.

Kwartyle. Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Uwaga: na karcie Zmienne na liście zmiennych źródłowych można tymczasowo zmienić poziom pomiaru powiązany ze zmienną (w wyniku czego zmieniają się statystyki podsumowujące wyświetlane dla tej zmiennej).

Częstości

Procedura Częstości umożliwia statystyczne i graficzne przedstawianie danych, co jest użyteczne w przypadku opisywania zmiennych wielu typów. Przeglądanie danych warto rozpocząć od tej właśnie procedury.

W przypadku raportu częstości lub wykresu słupkowego można ułożyć odrębne wartości w porządku rosnącym lub malejącym, lub w porządku wynikającym z ich częstości. Raport częstości może być usunięty, jeśli zmienna posiada wiele odrębnych wartości. Etykietami wykresów mogą być częstości (domyślnie) lub procenty.

Przykład

Jaki jest rozkład klientów firmy według ich miejsca pracy? Z wyników można dowiedzieć się, że 37,5% klientów pracuje w agencjach rządowych, 24,9% w korporacjach, 28,1% w instytucjach akademickich i 9,4% w przemyśle związanym z ochroną zdrowia. W przypadku danych ilościowych, takich jak przychód ze sprzedaży, można dowiedzieć się, że średnia sprzedaż ma wartość 3576 złotych z odchyleniem standardowym 1078 złotych.

Statystyki i wykresy

Liczebności, procenty, procenty skumulowane, średnia, mediana, dominanta, odchylenie standardowe, wariancja, zakres, wartości minimum i maksimum, błąd standardowy średniej, skośność i kurtoza (obie wraz z błędami standardowymi), kwantyle, percentyle określone przez użytkownika, wykresy słupkowe, wykresy kołowe oraz histogramy.

Zagadnienia dotyczące danych

Dane

Do kodowania zmiennych kategoryjnych (nominalne lub porządkowe poziomy pomiaru) należy używać kodów numerycznych lub łańcuchów.

Założenia

Tabelaryzacje i procenty dostarczają użytecznego opisu danych z dowolnego rozkładu, szczególnie dla zmiennych o uporządkowanych lub nieuporządkowanych kategoriach. Większość opcjonalnych statystyk podsumowujących, takich jak średnia i odchylenie standardowe, jest oparta na teorii rozkładu normalnego i są one odpowiednie dla zmiennych ilościowych o rozkładach symetrycznych. Statystyki odporne, takie jak mediana, kwartyle i percentyle, są odpowiednie dla zmiennych ilościowych, niezależnie od tego, czy spełniają one założenia o normalności.

Otrzymywanie tabel częstości

1. Wybierz z menu następującą opcję:

Analiza > Statystyki opisowe > Częstości...

2. Wybierz jedną lub kilka zmiennych jakościowych lub ilościowych.

3. Opcjonalnie wybierz opcję **Utwórz tabele w stylu APA**, aby tworzyć tabele wynikowe zgodne z wytycznymi APA dotyczącymi stylu prezentacji wyników.

4. Opcjonalnie możesz wykonać następujące czynności:

- Kliknąć przycisk **Statystyki**, aby wybrać statystyki opisowe dla zmiennych ilościowych.
- Kliknąć przycisk **Wykresy**, aby wybrać wykresy słupkowe, wykresy kołowe oraz histogramy.
- Kliknąć przycisk **Format**, aby określić porządek wyświetlania wyników.
- Kliknij przycisk **Styl**, aby określić warunki automatycznego zmieniania właściwości tabel przestawnych na podstawie określonych warunków.
- Kliknij przycisk **Bootstrap**, aby uzyskać odporne oszacowania błędów standardowych i przedziały ufności dla oszacowań, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji lub współczynnik regresji. Może być także używana do opracowywania testów hipotez.

Częstości: Statystyki

Wartości percentyli. Wartości zmiennych ilościowych dzielące uporządkowane dane na grupy takie, że pewien procent znajduje się powyżej, a pewien poniżej. Kwartyłe (percentyle 25., 50. i 75.) dzielą obserwacje na cztery grupy jednakowej wielkości. Jeśli istnieje konieczność podzielenia obserwacji na liczbę równych grup różną od czterech, należy wybrać opcję **Punkty podziału dla n równych grup**. Można również określić dowolne percentyle (np. percentyl 95. jest to wartość, poniżej której znajduje się 95% obserwacji).

Tendencja centralna. Statystyki opisujące położenie rozkładu obejmują średnią, medianę, dominantę i sumę wszystkich wartości.

- *Średnia.* Miara tendencji centralnej. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.
- *Mediana.* Wartość powyżej i poniżej której połowa obserwacji spada, czyli 50th percentyla. W sytuacji parzystej liczby obserwacji mediana jest średnią dwóch środkowych obserwacji w próbie posortowanej rosnąco lub malejąco. W przeciwieństwie do średniej, na którą wpływ może mieć nawet kilka ekstremalnie dużych lub małych wartości, mediana jest miarą tendencji centralnej niewrażliwą na wartości odstające).
- *Tryb.* Najczęstsze występujące wartości. Jeśli więcej, niż jedna wartość występuje z taką samą, największą częstością, każda z nich jest dominantą (wartością modalną). Procedura Częstości podaje tylko najmniejszą z wielu dominant.
- *Suma.* Suma wartości wszystkich obserwacji nieposiadających braków danych.

Rozproszenie. Statystyki mierzące zmienność lub rozrzut danych obejmują odchylenie standardowe, wariancję, rozstęp, minimum, maksimum oraz błąd standardowy średniej.

- *Odchylenie standardowe.* Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.
- *Wariancja.* Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyleń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja jest mierzona w jednostkach będących kwadratami jednostek miary dla zmiennej, do której wariancja się odnosi.
- *Wykładnik potęgi od.* Różnica między największą a najmniejszą wartością zmiennej numerycznej; maksimum minus minimum.
- *minimum.* Najmniejsza wartość zmiennej numerycznej.
- *maximum.* Największa wartość zmiennej numerycznej.
- *Błąd standardowy średniej.* Miara o ile wartość średniej może być różna w stosunku do próbki pobranej z tej samej dystrybucji. Może być wykorzystywana do pobieżnego porównania rzeczywistej wartości średniej z wartością hipotetyczną (tj. można sądzić, że te dwie wartości są różne, jeśli iloraz różnicy i błędu standardowego jest mniejszy od -2 lub większy od +2).

Rozkład. Skośność i kurtoza są statystykami opisującymi kształt i symetrię rozkładu. Te statystyki wyświetlane są wraz ze swoimi błędami standardowymi.

- *Skośność.* Miara asymetrii rozkładu. Rozkład normalny jest symetryczny, a jego wartość skośności wynosi 0. Rozkład o dużej skośności dodatniej posiada długi kraniec z prawej strony. Gdy zaś współczynnik skośności jest ujemny, rozkład ma długi kraniec z lewej strony. Jako wytyczna, wartość skośności przekraczająca dwukrotnie swój błąd standardowy na ogół oznacza odstępstwo od symetrii rozkładu.
- *Kurtoza.* Miara zakresu, w jakim występują wartości odstających. W przypadku rozkładu normalnego wartość statystyki kurtozy wynosi zero. Kurtoza dodatnia wskazuje, że w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego. Kurtoza ujemna wskazuje, że w danych istnieje mniej skrajnych wartości odstających niż w przypadku rozkładu normalnego. Stosowana definicja kurtozy, w której wartość wynosi 0 dla rozkładu normalnego, jest czasami określana jako współczynnik ekscesu. Niektóre programy mogą raportować kurtozę w taki sposób, że wartość ta wynosi 3 dla rozkładu normalnego.

Wartości są środkami grup. Jeśli wartości danych są środkami grup (np. wiek każdej osoby, która ma więcej niż trzydzieści lat, ale mniej niż czterdzieści jest kodowany jako 35), należy wybrać tę opcję w celu oszacowania mediany i percentyli pierwotnych, nie zgrupowanych danych.

Wykresy częstości

Uwaga: Wykresy nie są uwzględniane w wynikach, gdy w oknie dialogowym Bootstrap włączona jest opcja **Realizacja metody bootstrapowej**.

Typ wykresu

Na wykresie kołowym można wyświetlić są udziały składowych w całości. Każda część wykresu kołowego odpowiada grupie zdefiniowanej przez pojedynczą zmienną grupującą. Na wykresie słupkowym wyświetlane są liczebności każdej odrębnej wartości lub kategorii w postaci osobnego słupka, co umożliwia ich wizualne porównanie. Histogram także zawiera słupki, ale są one kreślone wzdłuż skali o równych przedziałach. Wysokość każdego słupka odpowiada liczebności występowania wartości zmiennej ilościowej wewnątrz przedziału. Histogram pokazuje kształt, środek i rozrzut rozkładu. Krzywa normalna nałożona na histogram pomaga ocenić, czy rozkład danych jest normalny.

Wartości na wykresie

W przypadku wykresów słupkowych oś ilościowa może być oznaczana za pomocą procentów lub liczebności wystąpień.

Częstości: Format

Porządkuj według. Tabela częstości może być uporządkowana zgodnie z rzeczywistymi wartościami danych lub zgodnie z liczebnością (częstotliwością występowania) tych wartości w porządku rosnącym lub malejącym. Jeśli jednak niezbędny jest histogram lub percentyle, procedura Częstości traktuje zmienną jako ilościową i wyświetla jej wartości porządku rosnącym.

Wiele zmiennych. Podczas tworzenia tabel statystycznych dla wielu zmiennych można wyświetlać wszystkie zmienne w pojedynczej tabeli (**Porównaj zmienne**) lub wyświetlać osobne tabele statystyczne dla każdej zmiennej (**Przedstaw wyniki w podziale na zmienne**).

Ukryj tabele z wieloma kategoriami. Wybranie tej opcji zapobiega wyświetlaniu tabel posiadających więcej wartości niż określono.

statystyki opisowe

Procedura Statystyki opisowe umożliwia wyświetlenie w jednej tabeli statystyk podsumowujących każdą z kilku zmiennych oraz wyliczenie ocen standaryzowanych (statystyk z). Zmienne mogą być uporządkowane według wartości ich średnich (rosnąco lub malejąco), alfabetycznie lub zgodnie z porządkiem, w którym zostały wybrane (ustawienie domyślne).

Przy zapisywaniu statystyki z są dodawane do danych w Edytorze danych i można ich używać do tworzenia wykresów, wykazów danych oraz poddać je analizom. W przypadku gdy zmienne są zapisane w różnych jednostkach (np. produkt narodowy brutto na osobę oraz odsetek ludzi umiejących czytać), to przekształcenie statystyk z umieszcza zmienne na wspólnej skali, co ułatwia ich wizualne porównanie.

Przykład. Jeśli każda obserwacja zawiera całkowitą dzienną sprzedaż każdego członka personelu handlowego (np. jeden wpis dla Jana, jeden dla Anny, jeden dla Marcina itd.) i te dane są zbierane każdego dnia przez kilka miesięcy, to procedura statystyk opisowych umożliwia wyliczenie średniej dziennej sprzedaży dla każdego członka personelu handlowego i uporządkowanie wyników od największej średniej sprzedaży do najmniejszej.

Statystyki. Wielkość próby, średnia, minimum, maksimum, odchylenie standardowe, wariancja, rozstęp, suma, błąd standardowy średniej oraz kurtoza i skośność wraz z ich błędami standardowymi.

Wymagania dotyczące danych w statystykach opisowych

Dane. Należy używać zmiennych numerycznych, które po przedstawieniu ich w formie graficznej mogą służyć do zapisywania błędów, wartości skrajnych oraz anomalii rozkładowych. Procedura statystyk opisowych jest bardzo skuteczna w przypadku dużych zbiorów danych (tysiące obserwacji).

Założenia. Większość dostępnych statystyk (również statystyki z) jest opartych na teorii rozkładu normalnego i są one odpowiednie dla zmiennych ilościowych (interwałowy lub ilorazowy poziom pomiaru) o symetrycznych rozkładach (należy unikać zmiennych z nieuporządkowanymi kategoriami lub rozkładami skośnymi). Należy unikać zmiennych z nieuporządkowanymi kategoriami lub rozkładami skośnymi. Rozkład statystyk z posiada taki sam kształt jak rozkład oryginalnych danych, dlatego obliczanie statystyk z nie jest rozwiązaniem w odniesieniu do danych problemowych.

Wykonywanie statystyk opisowych

1. Wybierz z menu następujące opcje:

Analiza > Statystyki opisowe > Statystyki opisowe...

2. Wybierz co najmniej jedną zmienną.

Opcjonalnie można wykonać następujące czynności:

- Zaznaczyć opcję **Zapisz standaryzowane wartości jako zmienne**, aby zapisać statystyki z jako zmienne wynikowe.
- Kliknąć przycisk **Opcje**, aby móc skorzystać ze statystyk opcjonalnych oraz ustalić porządek wyświetlania.

Statystyki opisowe: Opcje

Średnia i suma. Domyślnie wyświetlana jest średnia lub średnia arytmetyczna.

Rozproszenie. Statystyki mierzące rozrzut lub zmienność w danych to: odchylenie standardowe, wariancja, rozstęp, minimum, maksimum oraz błąd standardowy średniej.

- *Odchylenie standardowe.* Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.
- *Wariancja.* Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyleń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja jest mierzona w jednostkach będących kwadratami jednostek miary dla zmiennej, do której wariancja się odnosi.
- *Wykładnik potęgi od.* Różnica między największą a najmniejszą wartością zmiennej numerycznej; maksimum minus minimum.
- *minimum.* Najmniejsza wartość zmiennej numerycznej.
- *maximum.* Największa wartość zmiennej numerycznej.
- *Błąd standardowy średniej.* Miara o ile wartość średniej może być różna w stosunku do próbki pobranej z tej samej dystrybucji. Może być wykorzystywana do pobieżnego porównania rzeczywistej wartości średniej z wartością hipotetyczną (tj. można sądzić, że te dwie wartości są różne, jeśli iloraz różnicy i błędu standardowego jest mniejszy od -2 lub większy od +2).

Rozkład. Kurtoza i skośność są to statystyki charakteryzujące kształt i symetrię rozkładu. Te statystyki wyświetlane są wraz ze swoimi błędami standardowymi.

- *Kurtoza.* Miara zakresu, w jakim występują wartości odstających. W przypadku rozkładu normalnego wartość statystyki kurtozy wynosi zero. Kurtoza dodatnia wskazuje, że w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego. Kurtoza ujemna wskazuje, że w danych istnieje mniej skrajnych wartości odstających niż w przypadku rozkładu normalnego. Stosowana definicja kurtozy, w której wartość wynosi 0 dla rozkładu normalnego, jest czasami określana jako współczynnik ekscesu. Niektóre programy mogą raportować kurtozę w taki sposób, że wartość ta wynosi 3 dla rozkładu normalnego.
- *Skośność.* Miara asymetrii rozkładu. Rozkład normalny jest symetryczny, a jego wartość skośności wynosi 0. Rozkład o dużej skośności dodatniej posiada długi kraniec z prawej strony. Gdy zaś współczynnik skośności jest ujemny, rozkład ma długi kraniec z lewej strony. Jako wytyczna, wartość

skośności przekraczająca dwukrotnie swój błąd standardowy na ogół oznacza odstępstwo od symetrii rozkładu.

Porządek wyświetlania. Domyślnie zmienne są wyświetlane w porządku, w którym zostały wybrane. Opcjonalnie można wyświetlać zmienne alfabetycznie oraz według średnich rosnących lub malejących.

Dodatkowe właściwości komendy DESCRIPTIVES

Język składni komend umożliwia również:

- Zapisz standaryzowane statystyki (statystyki z) dla niektórych, lecz nie dla wszystkich zmiennych (za pomocą opcji komendy VARIABLES).
- Określ nazwy nowych zmiennych, które zawierają standaryzowane statystyki (za pomocą opcji komendy VARIABLES).
- Z analizy wyklucz obserwacje z brakującymi wartościami dla jakiegokolwiek zmiennej (przy pomocy opcji komendy MISSING).
- Posortuj wyświetlone zmienne według wartości jakieś statystyki, a nie średnie (za pomocą opcji komendy SORT).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Percentyle

W procedurze Percentyles wyświetlane są tabele percentyli.

Wartości percentyli

Po wybraniu opcji **Wartości centylowe** domyślnie wyświetlane są wartości dla percentyli 5th, 10th, 25th, 50th, 75th, 90th i 95th. Wartości zmiennych ilościowych dzielące uporządkowane dane na grupy takie, że pewien procent znajduje się powyżej, a pewien poniżej.

Kwartyle

Kwartyle (percentyle 25., 50. i 75.) dzielą obserwacje na cztery grupy jednakowej wielkości.

Niestandardowe

Po wybraniu opcji **Niestandardowe** wprowadź co najmniej jedną wartość, aby uruchomić analizę. Wartości wejściowe muszą być liczbą z zakresu od 0 do 100. Użyj przycisków **Dodaj**, **Zmień** **Usuń**, aby pracować z wartościami na liście wartości centylowych.

Metoda centyla

Domyślnie metoda HAVERAGE jest wybrana do obliczania percentyli.

Uzyskiwanie tabel percentyli

1. Z menu wybierz:

Analiza > Statystyki opisowe > Percentyle ..

2. Zaznacz co najmniej jedną zmienną numeryczną i przenieś ją w pole Zmienne.

3. Wybierz opcję **Wartości centylowe**, aby określić wartości, które mają być wyliczone dla percentyla.

4. Wybierz opcję **Metoda percentyla**, aby wybrać metodę obliczania percentyla.

5. Opcjonalnie można wykonać następujące czynności:

- Kliknij opcję **Brakujące**, aby sterować traktowaniem braków danych.
- Kliknij opcję **Bootstrap**, aby uzyskać stabilne oszacowania standardowych błędów i przedziałów ufności dla oszacowań percentyli.

Percentyle-brakujące wartości

Braki danych

Kontrola sposobu postępowania z brakami danych.

Wyłączanie wszystkich obserwacji z brakami danych

Obserwacje z brakami danych dla dowolnej określonej zmiennej są wykluczane z wszystkich analiz. Jest to ustawienie domyślne.

Według zmiennej

Wykluczanie obserwacji z brakami danych w oparciu o zmienną.

Eksploruj

Procedura Eksploracja pozwala tworzyć statystyki podsumowujące oraz graficzne reprezentacje danych dla wszystkich obserwacji albo oddzielnie dla grup obserwacji. Istnieje wiele powodów, dla których warto użyć procedury Eksploracja: klasyfikowanie danych, identyfikacja wartości odstających, opis, sprawdzanie założeń oraz charakteryzowanie różnic pomiędzy podpopulacjami (grupami obserwacji). Klasyfikowanie danych może wykazać obecność wartości niezwykłych, wartości skrajnych, luk w danych lub innych osobliwości. Eksploracja danych może być pomocna w ustaleniu, czy techniki statystyczne, których użytkownik ma zamiar użyć w celu analizy danych, są odpowiednie. Eksploracja może wykazać, że należy przekształcić dane, jeśli technika wymaga rozkładu normalnego. Użytkownik może też zdecydować się na zastosowanie testów nieparametrycznych.

Przykład. Przyjrzyjmy się rozkładowi czasu nauki poruszania się po labiryncie dla szczurów według czterech różnych harmonogramów nagradzania. Dla każdej z czterech grup można sprawdzić, czy rozkład czasów jest w przybliżeniu normalny i czy cztery wariancje są równe. Można też zidentyfikować obserwacje z pięcioma największymi i pięcioma najmniejszymi czasami. Wykresy skrzynkowe oraz wykresy łydga-i-liście w sposób graficzny podsumowują rozkład czasu nauki dla każdej z grup.

Statystyki i wykresy. Średnia, mediana, średnia dla przyciętych o 5%, błąd standardowy, wariancja, odchylenie standardowe, minimum, maksimum, rozstęp, rozstęp ćwiartkowy, skośność oraz kurtozy i ich błędy standardowe, przedział ufności dla średniej (oraz określony poziom ufności), percentyle, M-estymator Hubera, estymator fali Andrewsa, M-estymator Hampela, estymator dwuwagi Tukeya, pięć największych i najmniejszych wartości, statystyka Kołmogorowa-Smirnowa z poziomem istotności Lillieforsa do testowania normalności oraz statystyka Shapiro-Wilka. Wykresy skrzynkowe, wykresy łydga-i-liście, histogramy, wykresy normalności, wykresy rozrzut-poziom z testami Levene'a oraz przekształcenia.

Wymagania dotyczące eksploracji danych

Dane. Procedura eksploracji może być wykorzystana dla zmiennych ilościowych (miar poziomów przedziałów lub proporcji). Czynniki (wykorzystywane do podziału danych na grupy obserwacji) powinny mieć rozsądną liczbę odrębnych wartości (kategorii). Wartości te mogą być krótkimi wartościami łańcuchowymi lub numerycznymi. Zmienna opisu obserwacji, wykorzystywana do oznaczania etykietami wartości skrajnych w wykresach skrzynkowych, może być krótką zmienną łańcuchową, długą zmienną łańcuchową (pierwsze 15 bajtów) lub zmienną numeryczną.

Założenia. Rozkład danych użytkownika nie musi być symetryczny czy normalny.

Eksploracja danych użytkownika

1. Wybierz z menu następującą opcję:

Analiza > Statystyki opisowe > Eksploracja...

2. Wybierz co najmniej jedną zmienną zależną.

Opcjonalnie możesz wykonać następujące czynności:

- Wybrać co najmniej jeden czynnik, którego wartość zdefiniuje grupy obserwacji.
- Wybrać zmienną identyfikacyjną do opisu obserwacji.
- Kliknąć przycisk **Statystyki**, aby uzyskać dostęp do mocnych estymatorów, wartości skrajnych, percentyli i tabel częstości.
- Kliknąć przycisk **Wykresy**, aby uzyskać dostęp do histogramów, normalnych wykresów normalności z testami oraz wykresów rozrzut-poziom z testem Levene'a.
- Kliknąć przycisk **Opcje**, aby wybrać opcje postępowania z brakami danych.

Eksploracja: Statystyki

Statystyki opisowe. Te pomiary tendencji centralnej i rozproszenia wyświetlane są domyślnie. Pomiary tendencji centralnej wskazują położenie rozkładu i zawierają średnią, medianę i średnią dla przyciętych 5%. Pomiary rozproszenia pokazują niepodobieństwo wartości i zawierają błąd standardowy, wariancję, odchylenie standardowe, minimum, maksimum, rozstęp i rozstęp ćwiartkowy. Statystyki opisowe zawierają również pomiary kształtu rozkładu; skośność i kurtoza wyświetlane są wraz z ich błędami standardowymi. Wyświetlony jest również 95% poziom przedziału ufności dla średniej. Można określić inny poziom ufności.

M-estymatory. Mocne alternatywy dla przykładowej średniej i mediany do oszacowania środka położenia. Wartości estymatorów, które są obliczane, różnią się od wag, które mają zastosowanie do obserwacji. Wyświetlane są: estymator M Hubera, estymator fali Andrewsa, estymator M Hampela oraz estymator dwuwagi Tukeya.

Wartości skrajne. Umożliwia wyświetlenie pięciu wartości największych i pięciu najmniejszych wraz z etykietami obserwacji.

Percentyle

W procedurze Percentyles wyświetlane są tabele percentyli. Po wybraniu opcji **Percentyle** domyślnie wyświetlane są wartości dla percentyli 5th, 10th, 25th, 50th, 75th, 90th i 95th. Wartości zmiennych ilościowych dzielące uporządkowane dane na grupy takie, że pewien procent znajduje się powyżej, a pewien poniżej.

Kwartyle

Kwartyle (percentyle 25., 50. i 75.) dzielą obserwacje na cztery grupy jednakowej wielkości.

Niestandardowe

Po wybraniu opcji **Niestandardowe** wprowadź co najmniej jedną wartość, aby uruchomić analizę. Wartości wejściowe muszą być liczbą z zakresu od 0 do 100. Użyj przycisków **Dodaj**, **Zmień Usui**, aby pracować z wartościami na liście wartości centylowych.

Metoda

Domyślnie metoda HAVERAGE jest wybrana do obliczania percentyli.

Eksploracja: Wykresy

Wykresy skrzynkowe. Opcje te kontrolują wyświetlanie wykresów skrzynkowych w przypadku istnienia co najmniej dwóch zmiennych zależnych. Opcja **Poziomy czynnika razem** generuje oddzielny obszar wyświetlania dla każdej zmiennej zależnej. W obrębie obszaru wykresy skrzynkowe prezentowane są dla każdej z grup zdefiniowanej przez czynnik. Opcja **Zmienne zależne razem** generuje osobny obszar wyświetlania dla każdej z grup zdefiniowanej przez czynnik. W obrębie obszaru wykresy skrzynkowe przedstawione są jeden obok drugiego dla każdej zmiennej zależnej. Taki sposób wyświetlania jest szczególnie użyteczny, kiedy różne zmienne reprezentują jedną charakterystykę podlegającą pomiarom w różnych czasach.

Opisowe. Grupa Opis pozwala na wybór wykresów łydyga-i-liście oraz histogramów.

Wykresy normalności z testami. Umożliwia wyświetlanie normalnych wykresów prawdopodobieństwa i normalnych wykresów prawdopodobieństwa bez trendu. Wyświetlana jest statystyka Kołmogorowa-Smirnowa z poziomem istotności Lillieforsa do testowania normalności. Jeśli określono wagi inne niż całkowite statystyka Shapiro-Wilka jest obliczana, gdy ważona wielkość próby mieści się w przedziale od 3 do 50. W przypadku braku wag lub wag całkowitych statystyka jest obliczana, gdy ważona wielkość próby mieści się w zakresie od 3 do 5000.

Rozrzut-poziom z testem Levene'a. Kontroluje przekształcanie danych dla wykresów rozrzut-poziom. Dla wszystkich wykresów rozrzut-poziom wyświetlone jest nachylenie linii regresji i mocne testy Levene'a na jednorodność wariancji. Jeśli wybrano przekształcenie, testy Levene'a oparte są na przekształconych danych. Jeśli nie wybrano żadnego czynnika, wykresy rozrzut-poziom nie zostaną utworzone. Opcja **Oszacowanie potęgi** umożliwia sporządzanie wykresów logarytmów naturalnych rozstępów ćwiartkowych względem logarytmów naturalnych median dla wszystkich komórek i oszacowanie potęgowych przekształceń prowadzące do osiągnięcia równych wariancji w komórkach. Wykresy rozrzut-poziom pomagają w ustaleniu potęgi dla przekształcenia w celu stabilizacji (większego wyrównania)

wariancji w poprzek grup. Opcja **Przekształcone dane** pozwala na wybór jednego ze sposobów potęgowania, między innymi postępowania zgodnie z rekomendacjami z szacowania potęg, i sporządza wykresy przekształconych danych. Wykreślany jest rozstęp ćwiartkowy i mediana przekształconych danych. Opcja **Nieprzekształcone** umożliwia sporządzenie wykresów danych nieobrobionych. Jest to tożsamy z przekształceniem z potęgą 1.

Eksploracja: Transformacje potęgi

Są to transformacje potęgi dla wykresów rozrzut-poziom. Aby przekształcić dane, należy wybrać potęgę dla transformacji. Można wybrać jedną z poniższych alternatyw:

- **Logarytm naturalny.** Transformacja z użyciem logarytmu naturalnego. Jest to wartość domyślna.
- **1/pierwiastek kwadratowy.** Dla każdej wartości danych wyliczana jest odwrotność pierwiastka kwadratowego.
- **Odwrotność.** Wyliczana jest odwrotność każdej wartości danych.
- **1/pierwiastek kwadratowy.** Wyliczany jest pierwiastek kwadratowy każdej wartości danych.
- **Kwadrat.** Każda wartość danych jest podnoszona do kwadratu.
- **Trzecia potęga.** Każda wartość danych jest podnoszona do sześcianu.

Eksploracja: Opcje

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej zależnej lub czynnika są wyłączone ze wszystkich analiz. Jest to wartość domyślna.
- **Wyłączanie obserwacji parami.** Obserwacje bez braków danych dla zmiennych w grupie (komórce) są włączane do analizy tej grupy. Obserwacja może mieć braki danych dla zmiennych wykorzystywanych w innych grupach.
- **Raportuj wartości.** Braki danych dla zmiennych czynników są traktowane jako osobna kategoria. Tworzony jest pełny raport wynikowy dla tej dodatkowej kategorii. Tabele częstości zawierają kategorie dla braków danych. Braki danych dla czynnika są uwzględniane, lecz oznaczane jako braki.

Dodatkowe właściwości komendy EXAMINE

Procedura hierarchiczna skupień stosuje składnię komend EXAMINE. Język składni komend umożliwia również:

- Żądanie całkowitych wyników i wykresów poza wynikami i wykresami dla grup określonych przez zmienne czynniki (za pomocą opcji komendy TOTAL).
- Określanie wspólnej skali dla grup wykresów skrzynkowych (za pomocą opcji komendy SCALE).
- Określanie interakcji zmiennych czynników (za pomocą opcji komendy VARIABLES).
- Określanie percentyli innych niż domyślne (za pomocą opcji komendy PERCENTILES).
- Obliczanie percentyli według jednej z pięciu metod (za pomocą opcji komendy PERCENTILES).
- Określanie potęgowego przekształcenia dla wykresów rozrzut-poziom (za pomocą opcji komendy PLOT).
- Określanie liczby wartości skrajnych, które mają być wyświetlone (za pomocą opcji komendy STATISTICS).
- Określanie parametrów M-estymatorów, mocnych estymatorów położenia (za pomocą opcji komendy MESTIMATORS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Tabele przestawne

Procedura Tabele krzyżowe pozwala tworzyć tabele drugiego rzędu i tabele wielu rzędów, a także udostępnia wiele testów i miar powiązania dla tabel drugiego rzędu. Struktura tabeli oraz fakt, czy kategorie są uporządkowane, decyduje o tym, którego testu lub miary należy użyć.

Z wyjątkiem cząstkowych współczynników gamma statystyki tabel krzyżowych i miary siły powiązania są obliczane osobno dla każdej tabeli dwuwymiarowej. Po zdefiniowaniu zawartości wierszy, kolumn i warstw (zmienna sterująca) procedura tabeli krzyżowej tworzy odrębny zestaw powiązanych statystyk i miar dla każdego czynnika definiującego warstwę (lub dla każdej kombinacji wartości dwóch lub większej liczby zmiennych sterujących). Na przykład jeśli *pleć* jest czynnikiem definiującym warstwę w tabeli przedstawiającej zależność zmiennej *zamężna/żonaty* (tak, nie) od zmiennej *jakość życia* (bardzo ciekawe, spokojne, nudne), to wyniki dla tabeli drugiego rzędu dla kobiet są obliczane oddzielnie od tych dla mężczyzn i przedstawione jako panele umieszczone jeden po drugim.

Przykład. Czy klienci pochodzący z małych firm mogą przynosić więcej zysków przy sprzedaży usług (takich jak szkolenia i doradztwo) niż klienci z dużych firm? Z tabeli krzyżowej można się przekonać, że większość małych firm (poniżej 500 zatrudnionych) daje wysoką zyskowność sprzedaży usług, natomiast większość dużych firm (więcej niż 2 500 zatrudnionych) daje niską zyskowność.

Statystyki i miary siły powiązania. Test chi-kwadrat Pearsona, iloraz wiarygodności chi-kwadrat, test powiązania liniowego, test dokładny Fishera, chi-kwadrat z poprawką Yatesa, współczynnik r Pearsona, współczynnik rho Spearmana, współczynnik kontyngencji, phi, V Craméra, współczynniki lambda symetryczne i asymetryczne, współczynniki tau Goodmana i Kruskala, współczynnik niepewności, gamma, współczynnik d Somersa, tau- b Kendalla, tau- c Kendalla, eta, kappa Cohena, oszacowanie ryzyka względnego, iloraz szans, test McNemara, statystyki Cochra i Mantela-Haenszela oraz statystyki dla proporcji kolumnowych.

Wymagania dotyczące danych w tabelach krzyżowych

Dane. Do definiowania kategorii poszczególnych zmiennych w tabeli należy się posłużyć wartością numeryczną lub łańcuchową (osiem lub mniej bajtów). Na przykład dla zmiennej *pleć* można zakodować dane jako 1 i 2 lub jako *mężczyzna* i *kobieta*.

Założenia. Niektóre statystyki i miary zakładają, że kategorie są uporządkowane (dane porządkowe) lub że wartości są wartościami ilościowymi (dane przedziałowe lub ilorazowe), jak opisano w podrozdziale poświęconym statystyce. Inne jednak dadzą prawidłowe wyniki, gdy kategorie zmiennych w tabeli nie są uporządkowane (dane nominalne). Dla statystyk opartych na chi-kwadrat (phi, V Cramér i współczynnik kontyngencji) dane powinny być losową próbą o rozkładzie wielomianowym.

Uwaga: zmienne porządkowe mogą zawierać kody liczbowe, reprezentujące kategorie (na przykład 1 = *niski*, 2 = *średni*, 3 = *wysoki*) lub wartości łańcuchowe. Jednak zakłada się, że porządek alfabetyczny wartości łańcuchowych odzwierciedla rzeczywiste uporządkowanie kategorii. Na przykład zmiennej łańcuchowej o wartościach *mało*, *średnio* i *dużo* domyślnie przypisywany jest błędny porządek: *dużo*, *mało* i *średnio*. Zasadniczo lepiej jest stosować kody numeryczne do reprezentacji danych porządkowych.

Otrzymywanie tabeli krzyżowej

1. Wybierz z menu następującą opcję:

Analizuj > Statystyki opisowe > Tabele przestawne...

2. Wybierz jedną lub więcej zmiennych w wierszach i jedną lub więcej zmiennych w kolumnach.

Opcjonalnie możesz wykonać następujące czynności:

- Wybierz jedną lub więcej zmiennych sterujących.
- Kliknij przycisk **Statystyki**, aby określić testy i pomiary sił powiązania dla tabel drugiego rzędu lub podtabel.
- Kliknij przycisk **Komórki**, aby określić zestawienie wartości obserwowanych lub oczekiwanych, udziałów procentowych i reszt.
- Kliknij przycisk **Format**, aby określić porządek kategorii.

Warstwy tabeli krzyżowej

Wybranie jednej lub kilku zmiennych warstwy powoduje utworzenie odrębnej tabeli krzyżowej dla każdej kategorii każdej zmiennej warstwy (zmiennej sterującej). Na przykład jeśli tabela zawiera jedną zmienną wiersza, jedną zmienną kolumny i jedną zmienną warstwy z dwiema kategoriami, wynikiem jest jedna tabela drugiego rzędu dla każdej kategorii zmiennej warstwy. Aby utworzyć kolejną warstwę zmiennych sterujących, należy kliknąć przycisk **Następna**. Podtabele są tworzone dla każdej kombinacji kategorii każdej zmiennej z pierwszej warstwy z każdą zmienną z drugiej warstwy itd. Ewentualne tworzone statystyki i miary sił powiązania odnoszą się jedynie do podtabel dwuczynnikowych.

Zgrupowane wykresy słupkowe tabeli krzyżowych

Pokaż zgrupowane wykresy słupkowe. Zgrupowany wykres słupkowy pozwala uzyskać podsumowanie danych dla grup obserwacji. Każdej wartości zmiennej wybranej w polu Zmienne w wierszach odpowiada osobna grupa słupków. Wysokość słupków w każdej z grup jest definiowana przez zmienną określoną w polu Zmienne w kolumnach. Każdej wartości tej zmiennej odpowiada jeden zestaw słupków wyróżnionych barwą lub wzorem. W przypadku określenia więcej niż jednej zmiennej w polu Zmienne w kolumnach lub Zmienne w wierszach następuje utworzenie osobnego zgrupowanego wykresu słupkowego dla każdej kombinacji dwóch zmiennych.

Tabele krzyżowe przedstawiające zmienne warstwy w warstwach tabeli

Pokaż zmienne warstwy w warstwach tabeli. Można wybrać wyświetlanie zmiennych warstwy (zmienne sterujące) jako warstwy tabeli w tabeli krzyżowej. Umożliwia to stworzenie widoków przedstawiających ogólne statystyki dla zmiennych w rzędach i kolumnach, a także pozwala rozwinąć kategorie zmiennych warstw.

Poniżej przedstawiono przykład wykorzystania pliku *demo.sav* (dostępny w katalogu instalacyjnym w folderze Samples) otrzymany w następujący sposób:

1. Jako zmienną w rzędzie wybierz *Kategorię przychodu w tysiącach (inccat)*, jako zmienną w kolumnach - *Posiada PDA (ownpda)*, a jako zmienną warstwy - *Wykształcenie (ed)*.
2. Wybierz **Przedstawiaj zmienne warstwy w warstwach tabeli**.
3. W podrzędnym oknie Zawartość komórek wybierz **Kolumna**.
4. Uruchom procedurę Tabele krzyżowe, kliknij dwukrotnie w tabelę krzyżową i z rozwijanej listy Wykształcenie wybierz **Wyższe**.

Wybrany widok tabeli krzyżowej pokazuje statystyki dla respondentów posiadających wyższe wykształcenie.

Tabele krzyżowe: Statystyki

Chi-kwadrat. W tabelach o dwóch wierszach i dwóch kolumnach wybranie opcji **Chi-kwadrat** pozwala obliczyć chi-kwadrat Pearsona, iloraz wiarygodności chi-kwadrat, test dokładny Fishera oraz chi-kwadrat z poprawką Yatesa w kierunku ciągłości. Dla tabel 2 x 2 niebędących wynikiem braku wierszy lub kolumn w większej tabeli test dokładny Fishera jest wykonywany, gdy tabela zawiera komórkę o oczekiwanej częstości poniżej 5. Dla wszystkich innych tabel 2 x 2 obliczana jest statystyka chi-kwadrat z poprawką Yatesa. W przypadku tabel o dowolnej liczbie wierszy i kolumn opcja **Chi-kwadrat** pozwala obliczyć chi-kwadrat Pearsona, jak również iloraz wiarygodności chi-kwadrat. Kiedy obie zmienne tabeli są ilościowe, opcja **Chi-kwadrat** powoduje obliczenie powiązania liniowego.

Korelacje. W tabelach, w których wiersze i kolumny zawierają wartości porządkowe, opcja **Korelacje** pozwala obliczyć współczynnik korelacji Spearmana, rho (tylko dla danych liczbowych). Korelacja Spearmana stanowi miarę powiązania między rangami. Kiedy obie zmienne tabeli (czynniki) są ilościowe, opcja **Korelacje** powoduje obliczenie współczynnika korelacji Pearsona r , który jest miarą liniowego powiązania między zmiennymi.

Nominalne. Dla danych nominalnych (bez zdefiniowanego porządku, np. katolik, protestant i żyd), można wybrać opcję **Współczynnik kontyngencji**, **Phi** (współczynnik) lub **V Craméra**, **Lambda** (symetryczne lub asymetryczne oraz tau Goodmana i Kruskala) oraz **Współczynnik niepewności**.

- *Współczynnik kontyngencji.* Miara powiązania oparta na teście chi-kwadrat. Współczynnik przyjmuje wartości pomiędzy 0 i 1. Wartości bliskie zero wskazują na słabe powiązanie pomiędzy zmiennymi wierszowymi i kolumnowymi, a bliskie 1 na silny związek pomiędzy tymi zmiennymi. Maksymalna, możliwa wartość współczynnika jest zależna od liczby wierszy i kolumn w tabeli.
- *Phi i V-Cramera.* Phi jest miarą powiązania chi-kwadrat, która polega na podzieleniu statystyki chi-kwadrat przez wielkość próby i biorąc pierwiastek kwadratowy z wyniku. Miara V Cramera jest również miarą siły powiązania opartą na chi-kwadrat.
- *Lambda.* Miara powiązania, która odzwierciedla proporcjonalną redukcję błędu, gdy wartości zmiennej niezależnej są używane do przewidywania wartości zmiennej zależnej. Wartość lambda wynosząca 1 oznacza, że na podstawie wartości zmiennej niezależnej można jednoznacznie przewidzieć wartość zmiennej zależnej. Wartość 0 oznacza, że zmienna niezależna nie jest pomocna w przewidywaniu wartości zmiennej zależnej.
- *Współczynnik niepewności.* Miara powiązania, która wskazuje proporcjonalną redukcję błędu, gdy wartości jednej zmiennej są używane do przewidywania wartości innej zmiennej. Na przykład wartość 0,83 oznacza, że znajomość jednej zmiennej zmniejsza błąd w przewidywaniu wartości innej zmiennej o 83%. Program oblicza zarówno symetryczną jak i asymetryczną wersję współczynnika niepewności.

Porządkowe. W przypadku tabel, gdzie zarówno wiersze, jak i kolumny zawierają wartości uporządkowane, można wybrać opcję **Gamma** (zerowy rząd dla tabel drugiego rzędu i warunkowy dla tabel od trzeciego do dziesiątego rzędu), **tau-b Kendalla** i **tau-c Kendalla**. Aby móc prognozować kategorie kolumn na podstawie kategorii wierszy, należy wybrać opcję **d Somersa**.

- *Gamma.* Symetryczna miara związku między dwiema zmiennymi porządkowymi, która waha się od -1 do 1. Wartości bliskie wartości bezwzględnej 1 wskazują na silną relację między tymi dwiema zmiennymi. Wartości bliskie zeru wskazują na brak lub słabą zależność. Dla tabel drugiego rzędu przedstawiane są wartości gamma zerowego rzędu. W przypadku tabel od trzeciego do n-tego rzędu wyświetlany jest warunkowy współczynnik gamma.
- *d Somersa.* Miara związku między dwiema zmiennymi porządkowymi, która mieści się w zakresie od -1 do 1. Wartości bliskie wartości bezwzględnej 1 wskazują na silną relację między dwiema zmiennymi, a wartości bliskie 0 wskazują na słabą relację między zmiennymi lub jej brak. Jest to asymetryczne rozwinięcie statystyki gamma, od której różni się tylko tym, że bierze pod uwagę liczbę par nie powiązanych ze względu na zmienną niezależną. Symetryczna wersja tej statystyki jest również obliczana.
- *tau-b Kendalla.* Nieparametryczna miara korelacji dla zmiennych porządkowych lub rangowanych, które biorą pod uwagę powiązania. Znak współczynnika wskazuje na kierunek zależności, a jego wartość bezwzględna wskazuje na siłę związku. Większe wartości bezwzględne wskazują na silniejsze zależności. Współczynnik przyjmuje wartości z zakresu od -1 do +1. Jednak wartości -1 lub +1 mogą zostać otrzymane jedynie dla tabel kwadratowych.
- *tau-c Kendalla.* Nieparametryczna miara powiązania zmiennych porządkowych, które ignorują powiązania. Znak współczynnika wskazuje na kierunek zależności, a jego wartość bezwzględna wskazuje na siłę związku. Większe wartości bezwzględne wskazują na silniejsze zależności. Współczynnik przyjmuje wartości z zakresu od -1 do +1. Jednak wartości -1 lub +1 mogą zostać otrzymane jedynie dla tabel kwadratowych.

Nominalne przez przedziałowe. Jeśli jedna zmienna jest jakościowa, a druga ilościowa, to należy wybrać opcję **Eta**. Zmienna kategoryjna musi być zakodowana numerycznie.

- *Eta.* Miara powiązania, która waha się od 0 do 1, przy czym 0 oznacza brak powiązania między zmiennymi wierszową i kolumnową, a wartości bliskie 1 wskazują na wysoki stopień powiązania. Eta jest odpowiednia w sytuacji, gdy zmienna zależna mierzona jest na skali przedziałowej (np. dochód), a zmienna niezależna posiada ograniczony zbiór kategorii (np. płeć). Dwie wartości eta są obliczane: jedna dla zmiennej wierszowej, traktowanej jako zmienna przedziałowa, druga — dla zmiennej kolumnowej, traktowanej w taki sam sposób.

Kappa. Cohen's kappa mierzy umowę między ocenami dwóch raterów, gdy oba te oceny oceniają ten sam obiekt. Wartość 1 oznacza doskonałą zgodność. Wartość 0 oznacza, że zgodność nie jest lepsza od przypadkowej. Kappa bazuje na tabeli kwadratowej, w której wartości wierszy i kolumn przedstawiają tę samą skalę. Każda komórka, która ma zaobserwowane wartości dla jednej zmiennej, ale nie dla drugiej, otrzymuje wartość 0. Współczynnik kappa nie jest wyliczany, jeśli typ zapisu danych (łańcuchowe lub liczbowe) nie jest taki sam dla tych dwóch zmiennych. W przypadku zmiennej łańcuchowej obie zmienne muszą mieć są samą zdefiniowaną długość.

Ryzyko. Dla 2 tabel x 2: miara siły powiązania między występowaniem czynnika a wystąpieniem zdarzenia. Jeśli przedział ufności dla tej statystyki zawiera wartość 1, nie można przyjąć, że czynnik jest powiązany ze zdarzeniem. Iloraz szans może być używany jako ocena ryzyka względnego wówczas, gdy czynnik występuje rzadko.

McNemara. Test nieparametryczny dla dwóch powiązanych zmiennych dychotomicznych. Testuje zmianę w odpowiedziach przy pomocy rozkładu chi-kwadrat. Przydatny do badania zmian w odpowiedziach spowodowanych oddziaływaniem eksperymentalnym w planach typu pretest-posttest. Dla większych tabel kwadratów zgłaszany jest test symetrii McNemara-Bowkera.

Statystyki Cochra i Mantela-Haenszela. Statystyki Cochra i Mantela-Haenszela mogą być używane do testowania niezależności między zmienną czynnika dychotomicznego a zmienną odpowiedzi dychotomicznej, warunkowo od wzorców współzmiennych zdefiniowanych przez jedną lub więcej zmiennych (kontrolnych) warstwy. Należy zauważyć, że w przeciwieństwie do innych statystyk, które są wyliczane osobno dla każdej warstwy, statystyki Cochra i Mantela-Haenszela są wyliczane raz dla wszystkich warstw.

Tabele krzyżowe: Zawartość komórek

Aby ułatwić wykrycie prawidłowości w danych, które przyczyniają się do wysokiej wartości testu chi-kwadrat, procedura Tabele krzyżowe pozwala wyświetlić oczekiwane częstości i reszty (odchylenia) trzech typów, które mierzą różnicę między częstościami oczekiwanymi a obserwowanymi. Każda komórka w tabeli może zawierać dowolną kombinację wybranych liczebności, procentów i reszt.

Counts. Liczba faktycznych obserwacji oraz liczba oczekiwanych obserwacji, jeśli zmienne w wierszach i kolumnach są od siebie niezależne. Możesz zdecydować o tym, aby ukryć liczebności, które są mniejsze od określonej liczby całkowitej. Ukryte wartości są wyświetlane jako **<N**, gdzie **N** jest podaną liczbą całkowitą. Podana liczba całkowita musi być większa lub równa 2, chociaż dozwolona jest wartość 0 i wskazuje, że żadne liczebności nie są ukryte.

Porównaj proporcje kolumn. Ta opcja wylicza porównania parami dla proporcji kolumnowych i wskazuje, które pary kolumn (dla danego rzędu) znacząco się różnią. Znaczące różnice są oznaczane w tabeli krzyżowej formatowaniem stylu APA za pomocą liter z dolnym indeksem i są obliczane na poziomie istotności wynoszącym 0,05. *Uwaga:* jeśli ta opcja zostanie określona bez wybierania zaobserwowanych liczebności lub procentów w kolumnie, wówczas te zaobserwowane liczebności są uwzględniane w tabeli krzyżowej, z literami w stylu APA z dolnym indeksem, oznaczającymi wyniki testów proporcji kolumnowych.

- **Dostosuj wartości p (metoda Bonferroniego).** Porównania parami dla proporcji kolumnowych wykorzystują korektę Bonferroniego korygującą obserwowany poziom istotności ze względu na fakt realizacji porównań wielokrotnych.

Procenty. Procenty z wszystkich wierszy lub kolumn mogą się sumować. Możliwe jest także obliczenie procentów całkowitej liczby obserwacji zawartych w tabeli (w jednej warstwie).

Uwaga: W przypadku wybrania opcji **Ukryj małe liczebności** w grupie Liczebności procenty powiązane z ukrytymi liczebnościami są także ukryte.

Reszty. Surowe, niestandardyzowane reszty przedstawiają różnicę między wartościami obserwowanymi a oczekiwanymi. Możliwe jest także obliczenie reszt standaryzowanych i skorygowanych.

- **Niestandardyzowane.** Różnica między obserwowaną wartością a oczekiwaną wartością. Wartość oczekiwana jest liczbą obserwacji, której należałoby oczekiwać w komórce, gdyby nie istniał żaden związek pomiędzy dwiema zmiennymi. Reszta dodatnia wskazuje, że w komórce jest więcej obserwacji, niż powinno być, gdyby zmienne w kolumnie i w wierszu były niezależne.

- *Standaryzowane*. Reszta podzielona przez oszacowanie jego odchylenia standardowego. Standaryzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Skorygowane standaryzowane*. Reszta dla komórki (wartość obserwowana minus oczekiwana) podzielona przez swój oszacowany błąd standardowy. Otrzymana reszta standaryzowana jest wyrażona w jednostkach odchylenia standardowego powyżej i poniżej średniej.

Utwórz tabelę w stylu APA. Powoduje, że tworzone tabele wynikowe są zgodne z wytycznymi APA dotyczącymi stylu prezentacji wyników.

Uwaga: Opcje **Obserwowane**, **Oczekiwane**, **W wierszu**, **W kolumnie** i **Łącznie** nie są dostępne, jeśli wybrano opcję **Utwórz tabelę w stylu APA**.

Wagi niecałkowite. Liczby komórek mają zazwyczaj wartości całkowite, ponieważ oznaczają liczbę obserwacji w każdej komórce. Jeśli plik danych jest jednak aktualnie ważony przez zmienną ważącą zawierającą wartości ułamkowe (na przykład 1,25), liczby komórek mogą być także wartościami ułamkowymi. Wartości można obciąć lub zaokrąglić przed lub po obliczeniu liczby komórek lub użyć ułamkowych liczb komórek do wyświetlenia w tabeli i obliczeń statystycznych.

- *Zaokrąglaj liczby komórek.* Wagi obserwacji są używane bez zmian, ale skumulowane wagi w komórkach są zaokrąglane przed obliczaniem jakichkolwiek statystyk.
- *Obetnij liczby komórek.* Wagi obserwacji są używane bez zmian, ale skumulowane wagi w komórkach są zaokrąglane przed obliczaniem jakichkolwiek statystyk.
- *Zaokrąglaj wagi obserwacji.* Wagi obserwacji są zaokrąglane przed użyciem.
- *Obetnij wagi obserwacji.* Wagi obserwacji są przycinane przed użyciem.
- *Brak korekt.* Wagi przypadku są używane w miarę, jak są używane i ułamkowe liczby komórek są używane. Gdy jednak wymagane są dokładne statystyki (dostępne tylko z opcją Losowanie i testowanie), skumulowane wagi w komórkach są obcinane lub zaokrąglane przed wyliczeniem dokładnych statystyk testowych.

Tabele krzyżowe: Format

Wiersze można uporządkować w kolejności rosnącej lub malejącej względem wartości zmiennej wiersza.

Podsumowanie

W procedurze Podsumowania obserwacji obliczane są statystyki podgrup dla zmiennych w obrębie kategorii co najmniej jednej zmiennej grupującej. Wszystkie poziomy zmiennej grupującej są umieszczone w tabeli krzyżowej. Można określić kolejność wyświetlania statystyk. Wyświetlane są również statystyki podsumowujące dla każdej zmiennej we wszystkich kategoriach. Wartości danych w każdej kategorii mogą być widoczne na ekranie lub ukryte. W przypadku dużych zbiorów danych można wybrać opcję wyświetlania tylko pierwszych n obserwacji.

Przykład. Jaka jest średnia wielkość sprzedaży danego produktu w poszczególnych regionach i dla dużych grup zawodowych klientów? Może okazać się, że średnia wielkość sprzedaży jest większa na zachodzie niż w pozostałych częściach, a największą grupę wśród kupujących stanowią zatrudnieni w dużych przedsiębiorstwach.

Statystyki. Suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość pierwszej kategorii zmiennej grupującej, wartość ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent z sumy całkowitej, procent całkowitej liczebności N , procent z sumy wewnątrz, procent liczebności N wewnątrz, średnia geometryczna i średnia harmoniczna.

Wymagania dotyczące danych w podsumowaniu obserwacji

Dane. Zmienne grupujące to zmienne kategoryjne, które mogą przyjmować wartości liczbowe lub łańcuchowe. Liczba kategorii powinna być względnie mała. Powinna istnieć możliwość nadania rang pozostałym zmiennym.

Założenia. Niektóre opcjonalne statystyki stosowane w przypadku podgrup, np. średnia oraz odchylenie standardowe, opierają się na teorii rozkładu normalnego i odpowiednie są dla zmiennych ilościowych o symetrycznym rozkładzie. Odporne estymatory, takie jak mediana i rozstęp, odpowiednie są dla zmiennych ilościowych, które mogą, ale nie muszą spełniać założenia o symetrii rozkładu.

Uzyskiwanie podsumowań obserwacji

1. Wybierz z menu następującą opcję:

Analiza > Raporty > Podsumowania obserwacji...

2. Wybierz co najmniej jedną zmienną.

Opcjonalnie można wykonać następujące czynności:

- Wybrać jedną lub więcej zmiennych grupujących w celu podziału danych na podgrupy.
- Kliknąć przycisk **Opcje**, aby zmienić tytuł tabel z wynikami, dodać nagłówek (podpis) pod nimi lub wykluczyć obserwacje z brakami danych.
- Kliknąć przycisk **Statystyki**, aby wybrać statystyki opcjonalne.
- Zaznaczyć opcję **Pokaż obserwacje**, aby wyświetlić listę obserwacji w każdej podgrupie. Domyślnie lista ta zawiera tylko 100 pierwszych obserwacji z pliku. Liczbę tę można zwiększyć lub zmniejszyć, wpisując odpowiednią wartość w polu **Ogranicz obserwacje do pierwszych n**. Można też usunąć znacznik z tego pola, co spowoduje wyświetlenie wszystkich obserwacji.

Podsumowania obserwacji: Opcje

W raporcie podsumowań można zmienić tytuł tabel z wynikami lub dodać nagłówek, który wyświetlany będzie pod tabelami. Tytuły i nagłówki można zawijać poprzez wpisanie \n w miejscach, gdzie ma nastąpić podział wiersza tekstu.

Można również wyświetlić lub ukryć podtytuły dla podsumowań oraz wykluczyć lub uwzględnić obserwacje z brakami danych dla którejkolwiek zmiennej zastosowanej w analizach. Często przydatne jest oznaczenie brakujących obserwacji znakiem kropki lub gwiazdki. Można wpisać znak, wyrażenie lub kod, którego zawartość będzie wyświetlana przy brakach danych. Jeżeli znak, wyrażenie lub kod nie zostanie wpisany, to obserwacje z brakami danych nie będą wyróżniane w wynikach.

Podsumowania obserwacji: Statystyki

Użytkownik może wybrać co najmniej jedną z poniższych statystyk podgrup w przypadku zmiennej w każdej kategorii każdej zmiennej grupującej: suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość pierwszej kategorii zmiennej grupującej, wartość ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent z sumy całkowitej, procent całkowitej liczebności N , procent z sumy wewnątrz, procent liczebności N wewnątrz, średnia geometryczna, średnia harmoniczna. Statystyki te wyświetlane są na liście Statystyki komórek w takiej kolejności, w jakiej będą wyświetlane w raporcie. Statystyki podsumowujące są również wyświetlane dla każdej zmiennej we wszystkich kategoriach.

First. Wyświetla pierwszą wartość napotkaną w pliku danych.

Średnia geometryczna. Pierwiastek n -tego stopnia z iloczynu wartości, gdzie n oznacza liczbę obserwacji.

Mediana z danych pogrupowanych. Mediana obliczona dla danych zakodowanych na grupy. Przykładowo dla danych wiekowych: jeśli każda wartość z przedziału 30–39 lat jest kodowana jako 35, wartości z przedziału 40–49 lat jako 45 itd. mediana z danych pogrupowanych jest medianą obliczoną z danych kodowanych.

Średnia harmoniczna. Służy do szacowania średniej wielkości grupy, gdy liczebności próby w grupach nie są równe. Średnia harmoniczna jest całkowitą liczbą prób podzieloną przez sumę odwrotności ich wielkości.

Kurtoza. Miara zakresu, w jakim występują wartości odstających. W przypadku rozkładu normalnego wartość statystyki kurtozy wynosi zero. Kurtoza dodatnia wskazuje, że w danych istnieje więcej dodatnich

wartości odstających niż w przypadku rozkładu normalnego. Kurtoza ujemna wskazuje, że w danych istnieje mniej skrajnych wartości odstających niż w przypadku rozkładu normalnego. Stosowana definicja kurtozy, w której wartość wynosi 0 dla rozkładu normalnego, jest czasami określana jako współczynnik ekscesu. Niektóre programy mogą raportować kurtozę w taki sposób, że wartość ta wynosi 3 dla rozkładu normalnego.

Last. Wyświetla ostatnią wartość napotkaną w pliku danych.

maximum. Największa wartość zmiennej numerycznej.

Średnia. Miara tendencji centralnej. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.

Mediana. Wartość powyżej i poniżej której połowa obserwacji spada, czyli 50th percentyla. W sytuacji parzystej liczby obserwacji mediana jest średnią dwóch środkowych obserwacji w próbie posortowanej rosnąco lub malejąco. W przeciwieństwie do średniej, na którą wpływ może mieć nawet kilka ekstremalnie dużych lub małych wartości, mediana jest miarą tendencji centralnej niewrażliwą na wartości odstające).

minimum. Najmniejsza wartość zmiennej numerycznej.

N. Liczba obserwacji (rekordów).

Procent całkowitej liczebności. Wartość procentowa łącznej liczba obserwacji w każdej z kategorii.

Procent sumy łącznie. Wartość procentowa sumy w każdej z kategorii.

Wykładnik potęgi od. Różnica między największą a najmniejszą wartością zmiennej numerycznej; maksimum minus minimum.

Skośność. Miara asymetrii rozkładu. Rozkład normalny jest symetryczny, a jego wartość skośności wynosi 0. Rozkład o dużej skośności dodatniej posiada długi kraniec z prawej strony. Gdy zaś współczynnik skośności jest ujemny, rozkład ma długi kraniec z lewej strony. Jako wytyczna, wartość skośności przekraczająca dwukrotnie swój błąd standardowy na ogół oznacza odstępstwo od symetrii rozkładu.

Odchylenie standardowe. Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.

Błąd standardowy kurtozy. Stosunek kurtozy do swojego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od + 2). Wysoka dodatnia wartość dla kurtozy wskazuje na to, iż krańce rozkładu są dłuższe niż te dla rozkładu normalnego; ujemna wartość dla kurtozy wskazuje na krótsze ogony (podobnie jak w rozkładach prostokątnych).

Błąd standardowy średniej. Miara o ile wartość średniej może być różna w stosunku do próbki pobranej z tej samej dystrybucji. Może być wykorzystywana do pobieżnego porównania rzeczywistej wartości średniej z wartością hipotetyczną (tj. można sądzić, że te dwie wartości są różne, jeśli iloraz różnicy i błędu standardowego jest mniejszy od -2 lub większy od +2).

Błąd standardowy skośności. Stosunek skośności do jego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od + 2). Wysoka dodatnia wartość dla skośności wskazuje na długi prawy kraniec; skrajnie ujemna wartość wskazuje na długi lewy kraniec.

Suma. Suma wartości wszystkich obserwacji nieposiadających braków danych.

Wariancja. Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyłeń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja jest mierzona w jednostkach będących kwadratami jednostek miary dla zmiennej, do której wariancja się odnosi.

Średnie

Procedura Średnie oblicza średnie w podgrupach i pokrewne statystyki jednej zmiennej dla zmiennych zależnych w obrębie kategorii jednej lub kilku zmiennych niezależnych. Opcjonalnie można wykonać jednoczynnikową analizę wariancji, eta i testy liniowości.

Przykład. Należy zmierzyć średnią ilość tłuszczu pochłoniętego przez trzy różne rodzaje oleju kuchennego i wykonać jednoczynnikową analizę wariancji, aby się przekonać, czy średnie się różnią od siebie.

Statystyki. Suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość pierwszej kategorii zmiennej grupującej, wartość ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent z sumy całkowitej, procent całkowitej liczebności N , procent z sumy wewnątrz, procent liczebności N wewnątrz, średnia geometryczna i średnia harmoniczna. Dostępne opcje to analiza wariancji, eta, eta kwadrat oraz testy liniowości R i R^2 .

Wymagania dotyczące danych do obliczania średnich

Dane. Zmienne zależne to zmienne ilościowe, a zmienne niezależne to zmienne jakościowe. Wartościami zmiennych kategoryalnych mogą być wartości numeryczne lub łańcuchowe.

Założenia. Niektóre opcjonalne statystyki stosowane w przypadku podgrup, np. średnia oraz odchylenie standardowe, opierają się na teorii rozkładu normalnego i odpowiednie są dla zmiennych ilościowych o symetrycznym rozkładzie. Statystyki odporne, takie jak mediana i rozstęp, są odpowiednie dla zmiennych ilościowych, które mogą, ale nie muszą, spełniać założenia o normalności. Analiza wariancji jest odporna na odstępstwa od rozkładu normalnego, ale dane w każdej komórce powinny być symetryczne. W analizie wariancji zakłada się również, że grupy pochodzą z populacji o równych wariancjach. Aby przetestować to założenie, należy użyć testu jednorodności wariancji Levene'a dostępnego w procedurze Jednoczynnikowa ANOVA.

Obliczanie średnich w podgrupach

1. Wybierz z menu następującą opcję:

Analiza > Porównywanie średnich > Średnie...

2. Wybierz co najmniej jedną zmienną zależną.

3. Użyj jednej z poniższych metod, żeby wybrać kategoryalne zmienne niezależne:

- Wybierz co najmniej jedną zmienną niezależną. Zostaną wyświetlone wyniki oddzielnie dla każdej zmiennej niezależnej.
- Wybierz co najmniej jedną warstwę zmiennych niezależnych. Każda warstwa wtórnie dzieli próbę. Jeśli w Warstwie 1 i Warstwie 2 występuje po jednej zmiennej niezależnej, to wyniki zostaną wyświetlone w jednej tabeli krzyżowej, a nie w oddzielnych tabelach dla każdej zmiennej niezależnej.

4. Opcjonalnie można kliknąć przycisk **Opcje**, aby uzyskać statystyki opcjonalne, tabelę analizy wariancji (ANOVA), eta, eta kwadrat, R i R^2 .

Średnie: Opcje

Użytkownik może wybrać co najmniej jedną z poniższych statystyk podgrup w przypadku zmiennej w każdej kategorii każdej zmiennej grupującej: suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość pierwszej kategorii zmiennej grupującej, wartość ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent z sumy całkowitej, procent całkowitej liczebności N , procent z sumy wewnątrz, procent liczebności N wewnątrz, średnia geometryczna, średnia harmoniczna. Można zmienić kolejność, w jakiej będą wyświetlane statystyki dla grup. Kolejność wyświetlania statystyk w raporcie jest zgodna z kolejnością, w jakiej występują na liście Statystyki komórek. Statystyki podsumowujące są również wyświetlane dla każdej zmiennej we wszystkich kategoriach.

First. Wyświetla pierwszą wartość napotkaną w pliku danych.

Średnia geometryczna. Pierwiastek n -tego stopnia z iloczynu wartości, gdzie n oznacza liczbę obserwacji.

Mediana z danych pogrupowanych. Mediana obliczona dla danych zakodowanych na grupy. Przykładowo dla danych wiekowych: jeśli każda wartość z przedziału 30–39 lat jest kodowana jako 35, wartości z przedziału 40–49 lat jako 45 itd. mediana z danych pogrupowanych jest medianą obliczoną z danych kodowanych.

Średnia harmoniczna. Służy do szacowania średniej wielkości grupy, gdy liczebności próby w grupach nie są równe. Średnia harmoniczna jest całkowitą liczbą prób podzieloną przez sumę odwrotności ich wielkości.

Kurtoza. Miara zakresu, w jakim występują wartości odstających. W przypadku rozkładu normalnego wartość statystyki kurtozy wynosi zero. Kurtoza dodatnia wskazuje, że w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego. Kurtoza ujemna wskazuje, że w danych istnieje mniej skrajnych wartości odstających niż w przypadku rozkładu normalnego. Stosowana definicja kurtozy, w której wartość wynosi 0 dla rozkładu normalnego, jest czasami określana jako współczynnik ekscesu. Niektóre programy mogą raportować kurtozę w taki sposób, że wartość ta wynosi 3 dla rozkładu normalnego.

Last. Wyświetla ostatnią wartość napotkaną w pliku danych.

maximum. Największa wartość zmiennej numerycznej.

Średnia. Miara tendencji centralnej. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.

Mediana. Wartość powyżej i poniżej której połowa obserwacji spada, czyli 50th percentyla. W sytuacji parzystej liczby obserwacji mediana jest średnią dwóch środkowych obserwacji w próbie posortowanej rosnąco lub malejąco. W przeciwieństwie do średniej, na którą wpływ może mieć nawet kilka ekstremalnie dużych lub małych wartości, mediana jest miarą tendencji centralnej niewrażliwą na wartości odstające).

minimum. Najmniejsza wartość zmiennej numerycznej.

N. Liczba obserwacji (rekordów).

Procent całkowitej liczebności. Wartość procentowa łącznej liczby obserwacji w każdej z kategorii.

Procent z sumy całkowitej. Wartość procentowa sumy w każdej z kategorii.

Wykładnik potęgi od. Różnica między największą a najmniejszą wartością zmiennej numerycznej; maksimum minus minimum.

Skośność. Miara asymetrii rozkładu. Rozkład normalny jest symetryczny, a jego wartość skośności wynosi 0. Rozkład o dużej skośności dodatniej posiada długi kraniec z prawej strony. Gdy zaś współczynnik skośności jest ujemny, rozkład ma długi kraniec z lewej strony. Jako wytyczna, wartość skośności przekraczająca dwukrotnie swój błąd standardowy na ogół oznacza odstępstwo od symetrii rozkładu.

Odchylenie standardowe. Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.

Błąd standardowy kurtozy. Stosunek kurtozy do swojego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od +2). Wysoka dodatnia wartość dla kurtozy wskazuje na to, iż krańce rozkładu są dłuższe niż te dla rozkładu normalnego; ujemna wartość dla kurtozy wskazuje na krótsze ogony (podobnie jak w rozkładach prostokątnych).

Błąd standardowy średniej. Miara o ile wartość średniej może być różna w stosunku do próbki pobranej z tej samej dystrybucji. Może być wykorzystywana do pobieżnego porównania rzeczywistej wartości średniej z wartością hipotetyczną (tj. można sądzić, że te dwie wartości są różne, jeśli iloraz różnicy i błędu standardowego jest mniejszy od -2 lub większy od +2).

Błąd standardowy skośności. Stosunek skośności do jego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od +

2). Wysoka dodatnia wartość dla skośności wskazuje na długi prawy kraniec; skrajnie ujemna wartość wskazuje na długi lewy kraniec.

Suma. Suma wartości wszystkich obserwacji nieposiadających braków danych.

Wariancja. Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyleń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja jest mierzona w jednostkach będących kwadratami jednostek miary dla zmiennej, do której wariancja się odnosi.

Statystyki dla pierwszej warstwy

Tabela anova i eta. Wyświetlenie jednoczynnikowej analizy wariancji oraz obliczenie wartości eta i eta kwadrat (miary powiązania) dla każdej zmiennej niezależnej w pierwszej warstwie.

Test liniowości. Oblicza sumę kwadratów, stopni swobody i średniej kwadratów powiązanych z komponentami liniowymi i nieliniowymi, a także współczynnika F, R i R-kwadrat. Liniowość nie jest wyliczana, gdy zmienna niezależna jest krótką zmienną łańcuchową.

Kostki OLAP

Procedura Kostki OLAP (ang. Online Analytical Processing – przetwarzanie analityczne w trybie online) umożliwia obliczenie sum, średnich i innych statystyk jednej zmiennej dla ilościowych zmiennych charakteryzowanych w obrębie kategorii co najmniej jednej grupującej zmiennej jakościowej. Dla każdej kategorii każdej zmiennej grupującej tworzona jest osobna warstwa w tabeli.

Przykład. Sprzedaż całkowita i przeciętna dla różnych regionów oraz asortymenty produktów w regionach.

Statystyki. Suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość zmiennej pierwszej kategorii zmiennej grupującej, wartość zmiennej ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent ogólnej liczby obserwacji, procent sumy całkowitej, procent ogólnej liczby obserwacji wewnątrz zmiennych grupujących, procent sumy całkowitej wewnątrz zmiennych grupujących, średnia geometryczna i średnia harmoniczna.

Wymagania dotyczące danych przy kostkach OLAP

Dane. Zmienne charakteryzowane należą do zmiennych ilościowych (zmienne ciągłe mierzone na skali interwałowej lub ilorazowej), a zmienne grupujące są zmiennymi jakościowymi. Wartościami zmiennych kategoryalnych mogą być wartości numeryczne lub łańcuchowe.

Założenia. Niektóre opcjonalne statystyki stosowane w przypadku podgrup, np. średnia oraz odchylenie standardowe, opierają się na teorii rozkładu normalnego i odpowiednie są dla zmiennych ilościowych o symetrycznym rozkładzie. Statystyki odporne, takie jak mediana i rozstęp, są odpowiednie dla zmiennych ilościowych, które mogą, ale nie muszą, spełniać założenia o normalności.

Otrzymywanie kostek OLAP

1. Wybierz z menu następujące opcje:

Analiza > Raporty > Kostki OLAP..

2. Wybierz co najmniej jedną ilościową zmienną charakteryzowaną.

3. Wybierz co najmniej jedną jakościową zmienną grupującą.

Opcjonalnie można wykonać następujące działania:

- Wybrać różne statystyki podsumowujące (kliknij przycisk **Statystyki**). Aby można było wybrać statystyki podsumowujące, konieczne jest zaznaczenie co najmniej jednej zmiennej grupującej.
- Wyliczyć różnice pomiędzy parami zmiennych i parami grup zdefiniowanych przez zmienną grupującą (kliknij przycisk **Różnice**).
- Utworzyć tytuły tabel użytkownika (kliknij przycisk **Tytuły**).
- Wysokie liczebności mniejsze niż podana liczba całkowita. Ukryte wartości będą wyświetlane jako **<N**, gdzie **N** jest podaną liczbą całkowitą. Określona liczba całkowita musi być większa lub równa 2.

Kostki OLAP: Statystyki

Użytkownik może wybrać co najmniej jedną z poniższych statystyk podgrup w przypadku zmiennej w każdej kategorii zmiennych podsumowania: suma, liczba obserwacji, średnia, mediana, mediana z danych pogrupowanych, błąd standardowy średniej, minimum, maksimum, rozstęp, wartość pierwszej kategorii zmiennej grupującej, wartość ostatniej kategorii zmiennej grupującej, odchylenie standardowe, wariancja, kurtoza, błąd standardowy kurtozy, skośność, błąd standardowy skośności, procent z sumy całkowitej, procent ogólnej liczby obserwacji wewnątrz zmiennych grupujących, procent sumy całkowitej wewnątrz zmiennych grupujących, średnia geometryczna i średnia harmoniczna.

Można zmienić kolejność, w jakiej będą wyświetlane statystyki dla grup. Kolejność wyświetlania statystyk w raporcie jest zgodna z kolejnością, w jakiej występują na liście Statystyki komórek. Statystyki podsumowujące są również wyświetlane dla każdej zmiennej we wszystkich kategoriach.

First. Wyświetla pierwszą wartość napotkaną w pliku danych.

Średnia geometryczna. Pierwiastek n -tego stopnia z iloczynu wartości, gdzie n oznacza liczbę obserwacji.

Mediana z danych pogrupowanych. Mediana obliczona dla danych zakodowanych na grupy. Przykładowo dla danych wiekowych: jeśli każda wartość z przedziału 30–39 lat jest kodowana jako 35, wartości z przedziału 40–49 lat jako 45 itd. mediana z danych pogrupowanych jest medianą obliczoną z danych kodowanych.

Średnia harmoniczna. Służy do szacowania średniej wielkości grupy, gdy liczebności próby w grupach nie są równe. Średnia harmoniczna jest całkowitą liczbą prób podzieloną przez sumę odwrotności ich wielkości.

Kurtoza. Miara zakresu, w jakim występują wartości odstających. W przypadku rozkładu normalnego wartość statystyki kurtozy wynosi zero. Kurtoza dodatnia wskazuje, że w danych istnieje więcej dodatnich wartości odstających niż w przypadku rozkładu normalnego. Kurtoza ujemna wskazuje, że w danych istnieje mniej skrajnych wartości odstających niż w przypadku rozkładu normalnego. Stosowana definicja kurtozy, w której wartość wynosi 0 dla rozkładu normalnego, jest czasami określana jako współczynnik ekscesu. Niektóre programy mogą raportować kurtozę w taki sposób, że wartość ta wynosi 3 dla rozkładu normalnego.

Last. Wyświetla ostatnią wartość napotkaną w pliku danych.

maximum. Największa wartość zmiennej numerycznej.

Średnia. Miara tendencji centralnej. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.

Mediana. Wartość powyżej i poniżej której połowa obserwacji spada, czyli 50th percentyla. W sytuacji parzystej liczby obserwacji mediana jest średnią dwóch środkowych obserwacji w próbie posortowanej rosnąco lub malejąco. W przeciwieństwie do średniej, na którą wpływ może mieć nawet kilka ekstremalnie dużych lub małych wartości, mediana jest miarą tendencji centralnej niewrażliwą na wartości odstające).

minimum. Najmniejsza wartość zmiennej numerycznej.

N. Liczba obserwacji (rekordów).

Procent N w. Procent obserwacji dla określonej zmiennej grupującej w ramach kategorii innych zmiennych grupujących. W przypadku tylko jednej zmiennej grupującej wartość ta jest identyczna z procentem całkowitej liczby obserwacji.

Procent sumy w. Procent sumy dla określonej zmiennej grupującej w obrębie kategorii innych zmiennych grupujących. W przypadku tylko jednej zmiennej grupującej wartość ta jest identyczna z procentem sumy całkowitej.

Procent całkowitej liczebności. Wartość procentowa łącznej liczby obserwacji w każdej z kategorii.

Procent sumy łącznie. Wartość procentowa sumy w każdej z kategorii.

Wykładnik potęgi od. Różnica między największą a najmniejszą wartością zmiennej numerycznej; maksimum minus minimum.

Skośność. Miara asymetrii rozkładu. Rozkład normalny jest symetryczny, a jego wartość skośności wynosi 0. Rozkład o dużej skośności dodatniej posiada długi kraniec z prawej strony. Gdy zaś współczynnik

skośności jest ujemny, rozkład ma długi kraniec z lewej strony. Jako wytyczna, wartość skośności przekraczająca dwukrotnie swój błąd standardowy na ogół oznacza odstępstwo od symetrii rozkładu.

Odchylenie standardowe. Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład, jeśli średnia wieku osób wynosi 45 lat, a odchylenie standardowe wynosi 10, wówczas 95% rozważanych osób znajduje się w przedziale wiekowym między 25 a 65 lat.

Błąd standardowy kurtozy. Stosunek kurtozy do swojego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od +2). Wysoka dodatnia wartość dla kurtozy wskazuje na to, iż krańce rozkładu są dłuższe niż te dla rozkładu normalnego; ujemna wartość dla kurtozy wskazuje na krótsze ogony (podobnie jak w rozkładach prostokątnych).

Błąd standardowy średniej. Miara o ile wartość średniej może być różna w stosunku do próbki pobranej z tej samej dystrybucji. Może być wykorzystywana do pobieżnego porównania rzeczywistej wartości średniej z wartością hipotetyczną (tj. można sądzić, że te dwie wartości są różne, jeśli iloraz różnicy i błędu standardowego jest mniejszy od -2 lub większy od +2).

Błąd standardowy skośności. Stosunek skośności do jego błędu standardowego może być użyty jako test normalności (czyli można odrzucić normalność, jeśli współczynnik jest mniejszy niż -2 lub większy od +2). Wysoka dodatnia wartość dla skośności wskazuje na długi prawy kraniec; skrajnie ujemna wartość wskazuje na długi lewy kraniec.

Suma. Suma wartości wszystkich obserwacji nieposiadających braków danych.

Wariancja. Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyłeń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja jest mierzona w jednostkach będących kwadratami jednostek miary dla zmiennej, do której wariancja się odnosi.

Kostki OLAP: Różnice

To okno dialogowe umożliwia wyliczenie różnic procentowych i arytmetycznych pomiędzy zmiennymi charakteryzowanymi lub pomiędzy grupami zdefiniowanymi przez zmienną grupującą. Różnice wyliczane są dla wszystkich miar wybranych w oknie dialogowym Kostki OLAP: Statystyki.

Różnice pomiędzy zmiennymi. Powoduje wyliczenie różnic między parami zmiennych. Wartości statystyk podsumowujących dla drugiej zmiennej (zmiennej odejmowanej) w każdej parze są odejmowane od wartości statystyk podsumowujących dla pierwszej zmiennej w parze. W celu obliczenia różnic procentowych jako mianownik jest używana wartość zmiennej podsumowującej dla zmiennej odejmowanej. Aby określić różnice między zmiennymi należy najpierw wybrać w głównym oknie dialogowym co najmniej dwie zmienne charakteryzowane.

Różnice między grupami obserwacji. Powoduje wyliczenie różnic między parami grup zdefiniowanych przez zmienną grupującą. Wartości statystyk podsumowujących dla drugiej kategorii (kategorii odejmowana) w każdej parze są odejmowane od wartości statystyk podsumowujących dla pierwszej kategorii w parze. W celu obliczenia różnic procentowych wynik odejmowania dzieli się przez wartość statystyki podsumowującej dla kategorii odejmowanej. Aby określić różnice między grupami należy najpierw wybrać w głównym oknie dialogowym co najmniej jedną zmienną grupującą.

Kostki OLAP: Tytuły

Możliwa jest zmiana tytułu raportu wynikowego lub dodanie nagłówka, który zostanie wyświetlony poniżej tabeli wynikowej. Można także sterować zawijaniem wierszy tytułów i nagłóweków poprzez wpisywanie \n w miejscach, gdzie w tekście ma zostać wstawiony znak podziału wiersza.

Wprowadzenie do proporcji

Procedura proporcji oblicza testy i przedziały ufności dla proporcji dwumianowych lub różnic proporcji. Dostępne są statystyki proporcji dla jednej próby (test względem konkretnej wartości), prób zależnych (różnych zmiennych) lub prób niezależnych (różnych grup obserwacji). Program oferuje różne opcje dotyczące typów statystyk testowych i przedziałów ufności. Do innych procedur, które oferują podobne funkcje, należą: **CROSSTABS**, **NPAR TESTS** i **NPTESTS**.

Proporcje dla jednej próby

Testy dla pojedynczych prób i przedziały ufności. Raport zawiera obserwowaną proporcję, oszacowanie różnicy między proporcją populacji a hipotetyczną proporcją populacji, asymptotyczne błędy standardowe w hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla proporcji.

Proporcje dla prób zależnych

Testy dla prób zależnych i przedziały ufności dla różnic między proporcjami. Raport zawiera obserwowane proporcje, oszacowanie różnic między proporcjami populacji, asymptotyczne błędy standardowe różnic populacji przy hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla różnic proporcji.

Proporcje dla prób niezależnych

Testy dla prób niezależnych i przedziały ufności. Raport zawiera obserwowane proporcje, oszacowanie różnic między proporcjami populacji, asymptotyczne błędy standardowe różnic populacji przy hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla różnic proporcji.

Proporcje dla jednej próby

Procedura Proporcje dla jednej próby zapewnia testy i przedziały ufności dla poszczególnych proporcji dwumianowych. Przyjmuje się, że dane pochodzą z prostej próby losowej, a każda hipoteza testowa lub przedział ufności jest oddzielnym testem lub odrębnym przedziałem, opartym na proporcji dwumianowej. Raport zawiera obserwowaną proporcję, oszacowanie różnicy między proporcją populacji a hipotetyczną proporcją populacji, asymptotyczne błędy standardowe w hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla proporcji.

Przykład

Statystyki

Agrestiego-Coulla, Anscombe'a, Cloppera-Pearsona (dokładny), Jeffreysa, Logit, Walda, Walda (ze skorygowaną ciągłością), ocena Wilsona (ze skorygowaną ciągłością), dokładny dwumianowy, skorygowany dwumianowy mid-p, ocena, ocena (ze skorygowaną ciągłością).

Zagadnienia dotyczące danych

Dane

Procedura wyświetla żadaną statystykę testu oraz dwustronne prawdopodobieństwa, przedziały ufności dla różnic w proporcjach, a także proporcje, błędy standardowe i liczebności dla każdej grupy lub zmiennej. Procedura jest ograniczona do maksymalnie jednej wartości testowej.

Założenia

Wykonywanie testów proporcji dla jednej próby

1. Wybierz z menu następującą opcję:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Wybierz co najmniej jedną ilościową zmienną testowaną.
3. Opcjonalnie można wykonać następujące kroki:

- Wybierz kryteria sukcesu w sekcji **Zdefiniuj sukces**:

Ostatnia wartość

Używana będzie ostatnia lub najwyższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych. To jest ustawienie domyślne.

Pierwsza wartość

Używana będzie pierwsza lub najniższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych.

Wartości

Jedna lub więcej konkretnych wartości ujętych w nawiasy. Poszczególne wartości należy rozdzielać spacjami. Dotyczy to zmiennych liczbowych i łańcuchowych. Wartości zmiennych łańcuchowych powinny być ujęte w apostrofy.

Punkt środkowy

Wartości w środku zakresu obserwowanych wartości danych lub powyżej środka. Dotyczy to tylko zmiennych liczbowych.

Punkt podziału

Wartości równe określonej podanej wartości lub od niej większe. Dotyczy to tylko zmiennych liczbowych.

- Kliknij opcję **Przedziały ufności...**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.
- Kliknij opcję **Testy...**, aby wybrać typy statystyk testowych, które mają być prezentowane, lub ukryć wszystkie testy.
- Kliknij opcję **Braku danych...**, aby sterować postępowaniem z brakami danych.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

4. Kliknij **OK**.

Proporcje dla jednej próby: Przedziały ufności

Okno dialogowe Przedziały ufności zawiera opcje służące do określania poziomu pokrycia i wybierania typów prezentowanych przedziałów ufności.

Poziom pokrycia

Określa odsetek przedziału ufności. Należy określić wartość liczbową z zakresu (0,100). 95 jest ustawieniem domyślnym.

Typ(y) przedziału:

Zawiera opcje służące do określania, które typy przedziałów ufności mają być prezentowane. Dostępne są następujące opcje:

- Agrestiego-Coulla
- Anscombe'a
- Cloppera-Pearsona (dokładny)
- Jeffreysa
- Logita
- Walda
- Walda (ze skorygowaną ciągłością)
- Ocena Wilsona
- Ocena Wilsona (ze skorygowaną ciągłością)

Określanie przedziałów ufności dla proporcji dla jednej próby

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Kliknij opcję **Przedziały ufności**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.

Proporcje dla jednej próby: Testy

Okno dialogowe Testy zawiera opcje służące do określania, które typy statystyk testowych mają być prezentowane.

Wszystko

Wyniki zawierają wszystkie statystyki testowe.

Brak

Wyniki nie zawierają żadnych statystyk testowych.

Dokładny dwumianowy

Wyświetlenie dokładnych prawdopodobieństw dwumianowych.

Skorygowany dwumianowy mid-p

Wyświetlenie skorygowanych dwumianowych prawdopodobieństw mid-p. Jest to ustawienie domyślne.

Wynik

Wyświetlenie statystyk testu oceny Z. Jest to ustawienie domyślne.

Ocena (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu oceny Z ze skorygowaną ciągłością.

Walda

Wyświetlenie statystyk testu Z Walda.

Walda (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu Z Walda ze skorygowaną ciągłością.

Wartość testowa

Określa wartość testową z zakresu od 0 do 1. Wartością domyślną jest 0,5.

Wykonywanie testów proporcji dla jednej próby

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla jednej próby kliknij opcję **Testy**.
3. Wybierz jeden lub więcej spośród dostępnych testów.

Proporcje dla jednej próby: Braki danych

Okno dialogowe Braki danych zawiera opcje dotyczące postępowania z brakami danych.

Zasięg braków danych

Wyłączanie obserwacji analiza po analizie

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących zmiennych używanych w każdej z analiz. To jest ustawienie domyślne.

Wyłączanie wszystkich obserwacji z brakami danych

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących wszystkich zmiennych używanych we wszystkich analizach.

Braku danych zdefiniowane przez użytkownika

Wyklucz powoduje, że braki danych zdefiniowane przez użytkownika będą traktowane jako brakujące wartości. **Uwzględnij** powoduje ignorowanie oznaczeń braków danych zdefiniowanych przez użytkownika i traktuje braki danych zdefiniowane przez użytkownika jako obserwacje ważne.

Definiowanie ustawień braków danych dla proporcji dla jednej próby

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla prób zależnych kliknij opcję **Braki danych**.

3. Wybierz żądane ustawienia postępowania z brakami danych.

Proporcje dla prób zależnych

Procedura Proporcje dla prób zależnych zapewnia testy i przedziały ufności dla różnicy w dwóch zależnych lub sparowanych dwumianowych proporcjach. Przyjmuje się, że dane pochodzą z prostej próby losowej, a każda hipoteza testowa lub przedział ufności jest oddzielnym testem lub odrębnym przedziałem. Raport zawiera obserwowane proporcje, oszacowanie różnic między proporcjami populacji, asymptotyczne błędy standardowe różnic populacji przy hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla różnic proporcji.

Przykład

Statystyki

Agrestiego-Mina, Bonetta-Price'a, Newcombe'a, Walda, Walda (ze skorygowaną ciągłością), Dokładny dwumianowy, Skorygowany dwumianowy mid-p, McNemara, McNemara (ze skorygowaną ciągłością).

Zagadnienia dotyczące danych

Dane

- Wymagana jest lista zmiennych zawierająca co najmniej dwie zmienne.
- Jeśli podana zostanie jedna lista zmiennych, każdy element listy zostanie połączony w pary ze wszystkimi pozostałymi elementami.

Założenia

- Jeśli dwie listy zmiennych zostaną oddzielone słowem **WITH** bez słowa kluczowego (**PAIRED**), każdy element pierwszej listy zostanie połączony w parę z każdym elementem drugiej listy.
- Jeśli dwie listy zmiennych zostaną oddzielone słowem **WITH** bez słowa kluczowego (**PAIRED**), każdy element pierwszej listy zostanie połączony w parę z każdym elementem drugiej listy. Niedopasowane zmienne zostaną zignorowane i zostanie wygenerowany komunikat ostrzegawczy.

Wykonywanie testów proporcji dla prób zależnych

1. Wybierz z menu następującą opcję:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Wybierz co najmniej jedną ilościową zmienną testowaną.

3. Opcjonalnie możesz wykonać następujące czynności:

- Wybierz kryteria sukcesu w sekcji **Zdefiniuj sukces**:

Ostatnia wartość

Używana będzie ostatnia lub najwyższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych. To jest ustawienie domyślne.

Pierwsza wartość

Używana będzie pierwsza lub najniższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych.

Wartości

Jedna lub więcej konkretnych wartości ujętych w nawiasy. Poszczególne wartości należy rozdzielać spacjami. Dotyczy to zmiennych liczbowych i łańcuchowych. Wartości zmiennych łańcuchowych powinny być ujęte w apostrofy.

Punkt środkowy

Wartości w środku zakresu obserwowanych wartości danych lub powyżej środka. Dotyczy to tylko zmiennych liczbowych.

Punkt podziału

Wartości równe określonej podanej wartości lub od niej większe. Dotyczy to tylko zmiennych liczbowych.

- Kliknij opcję **Przedziały ufności...**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.
- Kliknij opcję **Testy...**, aby wybrać typy statystyk testowych, które mają być prezentowane, lub ukryć wszystkie testy.
- Kliknij opcję **Braku danych...**, aby sterować postępowaniem z brakami danych.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

4. Kliknij **OK**.

Proporcje dla prób niezależnych: przedziały ufności

Okno dialogowe Przedziały ufności zawiera opcje służące do określania poziomu pokrycia i wybierania typów prezentowanych przedziałów ufności.

Poziom pokrycia

Określa odsetek przedziału ufności. Należy określić wartość liczbową z zakresu (0,100). 95 jest ustawieniem domyślnym.

Typ(y) przedziału:

Zawiera opcje służące do określania, które typy przedziałów ufności mają być prezentowane. Dostępne są następujące opcje:

- Agrestiego-Mina
- Bonetta-Price'a
- Newcomba
- Walda
- Walda (ze skorygowaną ciągłością)

Określanie przedziałów ufności dla proporcji dla prób zależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Kliknij opcję **Przedziały ufności**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.

Proporcje dla prób zależnych: Testy

Okno dialogowe Testy zawiera opcje służące do określania, które typy statystyk testowych mają być prezentowane.

Wszystko

Wyniki zawierają wszystkie statystyki testowe.

Brak

Wyniki nie zawierają żadnych statystyk testowych.

Dokładny dwumianowy

Wyświetlenie dokładnych prawdopodobieństw dwumianowych.

Skorygowany dwumianowy mid-p

Wyświetlenie skorygowanych dwumianowych prawdopodobieństw mid-p. Jest to ustawienie domyślne.

McNemar

Wyświetlenie statystyk testu Z McNemara. Jest to ustawienie domyślne.

McNemara (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu Z McNemara ze skorygowaną ciągłością.

Walda

Wyświetlenie statystyk testu Z Walda.

Walda (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu Z Walda ze skorygowaną ciągłością.

Wykonywanie testów proporcji dla prób zależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla prób niezależnych kliknij opcję **Testy**.

3. Wybierz jeden lub więcej spośród dostępnych testów.

Proporcje dla prób zależnych: Braki danych

Okno dialogowe Braki danych zawiera opcje dotyczące postępowania z brakami danych.

Zasięg braków danych

Wyłączanie obserwacji analiza po analizie

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących zmiennych używanych w każdej z analiz. To jest ustawienie domyślne.

Wyłączanie wszystkich obserwacji z brakami danych

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących wszystkich zmiennych używanych we wszystkich analizach.

Braku danych zdefiniowane przez użytkownika

Wyklucz powoduje, że braki danych zdefiniowane przez użytkownika będą traktowane jako brakujące wartości. **Uwzględnij** powoduje ignorowanie oznaczeń braków danych zdefiniowanych przez użytkownika i traktuje braki danych zdefiniowane przez użytkownika jako obserwacje ważne.

Definiowanie ustawień braków danych dla proporcji dla prób zależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla prób zależnych kliknij opcję **Braki danych**.

3. Wybierz żądane ustawienia postępowania z brakami danych.

Proporcje dla prób niezależnych

Procedura Proporcje dla prób niezależnych udostępnia testy i przedziały ufności dla różnicy w dwóch niezależnych dwumianowych proporcjach. Przyjmuje się, że dane pochodzą z prostej próby losowej, a każda hipoteza testowa lub przedział ufności jest oddzielnym testem lub odrębnym przedziałem. Raport zawiera obserwowane proporcje, oszacowanie różnic między proporcjami populacji, asymptotyczne błędy standardowe różnic populacji przy hipotezach zerowych i alternatywnych, określone statystyki testów z prawdopodobieństwem dwustronnym oraz określone przedziały ufności dla różnic proporcji.

Przykład

Statystyki

Agrestiego-Mina, Bonetta-Price'a, Newcombe'a, Walda, Walda (ze skorygowaną ciągłością), Dokładny dwumianowy, Skorygowany dwumianowy mid-p, McNemara, McNemara (ze skorygowaną ciągłością).

Zagadnienia dotyczące danych

Dane

- Wymagana jest co najmniej jedna zmienna zależna i jedna zmienna identyfikująca porównywane grupy.
- Zmienna grupująca może być liczbowa lub łańcuchowa.

Założenia

Wykonywanie testów proporcji dla prób niezależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Wybierz co najmniej jedną ilościową zmienną testowaną.

3. W polu **Zmienna grupująca** wybierz jedną zmienną, która identyfikuje dwie porównywane grupy.

4. Opcjonalnie określ ustawienia dla zmiennej wybranej w polu **Zmienna grupująca**.

- W przypadku wybrania opcji **Wartości** można podać dwie wartości liczbowe lub łańcuchowe z nawiasami jako wartości do porównania. Wartości łańcuchowe powinny być ujęte w apostrofy. Obserwacje z innymi wartościami są ignorowane.
- **Punkt środkowy** ma zastosowanie tylko do zmiennych liczbowych. Obserwacje w punkcie środkowym lub powyżej niego w rozkładzie zmiennej grupującej są przypisywane do drugiej grupy, a obserwacje poniżej punktu środkowego są przypisywane do pierwszej grupy.
- **Punkt podziału** ma zastosowanie tylko do zmiennych liczbowych i umożliwia określenie jednej zmiennej liczbowej z nawiasami. Obserwacje w punkcie podziału lub powyżej niego w zmiennej grupującej są przypisywane do drugiej grupy, a obserwacje poniżej punktu podziału są przypisywane do pierwszej grupy.

5. Opcjonalnie możesz wykonać następujące czynności:

- Wybierz kryteria sukcesu w sekcji **Zdefiniuj sukces**:

Ostatnia wartość

Używana będzie ostatnia lub najwyższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych. To jest ustawienie domyślne.

Pierwsza wartość

Używana będzie pierwsza lub najniższa wartość spośród posortowanych odrębnych wartości w danych. Dotyczy to zmiennych liczbowych i łańcuchowych.

Wartości

Jedna lub więcej konkretnych wartości ujętych w nawiasy. Poszczególne wartości należy rozdzielać spacjami. Dotyczy to zmiennych liczbowych i łańcuchowych. Wartości zmiennych łańcuchowych powinny być ujęte w apostrofy.

Punkt środkowy

Wartości w środku zakresu obserwowanych wartości danych lub powyżej środka. Dotyczy to tylko zmiennych liczbowych.

Punkt podziału

Wartości równe określonej podanej wartości lub od niej większe. Dotyczy to tylko zmiennych liczbowych.

- Kliknij opcję **Przedziały ufności...**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.

- Kliknij opcję **Testy...**, aby wybrać typy statystyk testowych, które mają być prezentowane, lub ukryć wszystkie testy.
- Kliknij opcję **Braku danych...**, aby sterować postępowaniem z brakami danych.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

6. Kliknij **OK**.

Proporcje dla jednej próby: przedziały ufności

Okno dialogowe Przedziały ufności zawiera opcje służące do określania poziomu pokrycia i wybierania typów prezentowanych przedziałów ufności.

Poziom pokrycia

Określa odsetek przedziału ufności. Należy określić wartość liczbową z zakresu (0,100). 95 jest ustawieniem domyślnym.

Typ(y) przedziału:

Zawiera opcje służące do określania, które typy przedziałów ufności mają być prezentowane. Dostępne są następujące opcje:

- Agrestiego-Caffo
- Browna-Li-Jeffreysa
- Haucka-Andersona
- Newcomba
- Newcomba (ze skorygowaną ciągłością)
- Walda
- Walda (ze skorygowaną ciągłością)

Określanie przedziałów ufności dla proporcji dla prób niezależnych

1. Wybierz z menu następującą opcję:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. Kliknij opcję **Przedziały ufności**, aby wybrać typy przedziałów ufności, które mają być prezentowane, lub ukryć wszystkie przedziały ufności.

Proporcje dla prób niezależnych: Testy

Okno dialogowe Testy zawiera opcje służące do określania, które typy statystyk testowych mają być prezentowane.

Wszyscy

Wyniki zawierają wszystkie statystyki testowe.

Brak

Wyniki nie zawierają żadnych statystyk testowych.

Haucka-Andersona

Wyświetlenie statystyk testu Z Haucka-Andersona.

Walda

Wyświetlenie statystyk testu Z Walda.

Walda (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu Z Walda ze skorygowaną ciągłością.

Walda H₀

Wyświetlenie statystyk testu Z Walda przy szacunkach wariancji przy H₀.

Walda H_0 (ze skorygowaną ciągłością)

Wyświetlenie statystyk testu Z Walda ze skorygowaną ciągłością przy szacunkach wariancji przy H_0 .

Wykonywanie testów proporcji dla prób niezależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla prób niezależnych kliknij opcję **Testy**.

3. Wybierz jeden lub więcej spośród dostępnych testów.

Proporcje dla prób niezależnych: Braki danych

Okno dialogowe Braki danych zawiera opcje dotyczące postępowania z brakami danych.

Zasięg braków danych

Wyłączanie obserwacji analiza po analizie

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących zmiennych używanych w każdej z analiz. To jest ustawienie domyślne.

Wyklucz wszystkie obserwacje z brakami

Oznacza uwzględnienie wszystkich obserwacji z wystarczającą ilością danych dotyczących wszystkich zmiennych używanych we wszystkich analizach.

Braku danych zdefiniowane przez użytkownika

Wyklucz powoduje, że braki danych zdefiniowane przez użytkownika będą traktowane jako brakujące wartości. **Uwzględnij** powoduje ignorowanie oznaczeń braków danych zdefiniowanych przez użytkownika i traktuje braki danych zdefiniowane przez użytkownika jako obserwacje ważne.

Definiowanie ustawień braków danych dla proporcji dla prób niezależnych

1. Wybierz z menu następującą opcję:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

2. W oknie dialogowym Proporcje dla prób niezależnych kliknij opcję **Braki danych**.

3. Wybierz żądane ustawienia postępowania z brakami danych.

Testy t

Testy t

Dostępne są trzy typy testów t:

Test t dla prób niezależnych (test t dla dwóch prób). Porównuje średnie jednej zmiennej dla dwóch grup obserwacji. Wynikami tego testu są statystyki opisowe dla każdej grupy i testy jednorodności wariancji Levene'a, jak również wartości t dla równych i nierównych wariancji oraz 95% przedział ufności dla różnicy średnich.

Test t dla prób zależnych (test t dla zmiennych zależnych). Porównuje średnie dwóch zmiennych dla jednej grupy obserwacji. Test ten jest również używany w badaniach opartych na parach dopasowanych i planach przypadek-kontrola. Jego wyniki zawierają statystyki opisowe dla zmiennych testowanych, korelację między nimi, statystyki opisowe dla różnic w próbach zależnych, test t oraz 95% przedział ufności.

Test t dla jednej próby. Porównuje średnią dla jednej zmiennej ze znaną lub hipotetyczną wartością. Statystyki opisowe dla zmiennych testowanych wyświetlane są wraz z testem t. Do wyników domyślnych należy 95% przedział ufności dla różnicy między średnią dla zmiennej testowanej a hipotetyczną wartością testowaną.

Test t dla prób niezależnych

Procedura testu t dla prób niezależnych porównuje średnie dwóch grup obserwacji i automatyzuje obliczenia wielkości efektu testu t. W idealnych warunkach obiekty powinny być losowo przypisane do dwóch grup, tak aby każda różnica ich reakcji była wynikiem oddziaływania (lub braku oddziaływania) tylko jednego czynnika. Nie jest tak w przypadku porównywania średniego dochodu mężczyzn i kobiet. Płeć badanych nie jest przypisywana losowo. W takich przypadkach należy zadbać o to, żeby różnice innych czynników nie pomniejszyły, ani nie powiększyły, znaczącej różnicy średnich. Na różnice średniego dochodu mogą mieć także wpływ takie czynniki jak wykształcenie (a nie tylko płeć).

Przykład

Pacjentów z wysokim ciśnieniem krwi przydziela się losowo do jednej z dwóch grup: grupy placebo lub grupy terapeutycznej. Badani z grupy placebo otrzymują tabletkę nieaktywną, natomiast badani z grupy terapeutycznej przyjmują nowy lek, który ma obniżać ciśnienie krwi. Po dwumiesięcznym leczeniu, za pomocą testu t dla dwóch prób, porównuje się średnie ciśnienie krwi pacjentów z grupy placebo i grupy terapeutycznej. Każdy pacjent należy do jednej grupy i jego ciśnienie jest mierzone jeden raz.

Statystyki

Dla każdej zmiennej: wielkość próby, średnia, odchylenie standardowe, błąd standardowy średniej oraz oszacowanie wielkości efektu dla testu t. Dla różnicy średnich: średnia, błąd standardowy i przedział ufności (można określić poziom ufności). Testy: test Levene'a równości wariancji oraz testy t równości średnich dla wariancji wspólnych i oddzielnych.

Zagadnienia dotyczące danych

Dane

Wartości badanej zmiennej ilościowej znajdują się w pojedynczej kolumnie w pliku danych. Procedura wykorzystuje do podziału obserwacji na dwie grupy zmienną grupującą o dwóch wartościach. Zmienna grupująca może być zmienną numeryczną (np. 1 i 2 lub 6,25 i 12,5) albo krótką zmienną łańcuchową (np. *tak* i *nie*). Można też użyć zmiennej ilościowej, takiej jak *wiek*, aby podzielić obserwacje na dwie grupy, określając punkt podziału (punkt podziału 21 dzieli obserwacje *wiek* na grupy osób poniżej i powyżej 21. roku życia).

Założenia

W przypadku testu t zakładającego równość wariancji obserwacje powinny być niezależnymi, losowymi próbkami z rozkładów normalnych z tą samą wariancją populacji. W przypadku testu t nie zakładającego równości wariancji obserwacje powinny być niezależnymi, losowymi próbkami z rozkładów normalnych. Test t dla dwóch prób jest dość odporny na odstępstwa od normalności. Podczas graficznego sprawdzania rozkładów należy upewnić się, czy są one symetryczne i czy nie zawierają obserwacji odstających.

Wykonywanie testu t dla prób niezależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób niezależnych...

- Wybierz co najmniej jedną ilościową zmienną testowaną. Dla każdej zmiennej obliczany jest oddzielny test t.
- Wybierz pojedynczą zmienną grupującą, a następnie kliknij przycisk **Definiuj grupy**, aby określić dwa kody dla grup, które mają być porównane.
- Opcjonalnie można wykonać następujące czynności:
 - Wybierz opcję **Szacuj wielkości efektów**, aby sterować szacowaniem wielkości efektu testu t.
 - Kliknij przycisk **Opcje**, aby określić przedział ufności oraz sposób postępowania z brakami danych.
 - Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

Test t dla prób niezależnych: Definiuj grupy

W przypadku liczbowych zmiennych grupujących zdefiniuj dwie grupy dla testu t , określając dwie wartości lub punkt podziału:

- **Użyj określonych wartości.** Wprowadź jedną wartość dla grupy 1 i inną wartość dla grupy 2. Obserwacje z dowolnymi innymi wartościami są wykluczane z analizy. Liczby te nie muszą być liczbami całkowitymi (dopuszczalnymi wartościami są np. 6,25 i 12,5).
- **Punkt podziału.** Można również wprowadzić liczbę dzielącą wartości zmiennej grupującej na dwie grupy. Wszystkie obserwacje, których wartości są mniejsze od punktu podziału, tworzą jedną grupę, zaś obserwacje, których wartości są większe od punktu podziału lub mu równe, tworzą drugą grupę.

W przypadku łańcuchowych zmiennych grupujących wprowadź łańcuch dla grupy 1 i inną wartość dla grupy 2, na przykład *tak* i *nie*. Obserwacje z innymi łańcuchami są wykluczane z analizy.

Test t dla prób niezależnych: Opcje

Oszacowanie przedziału ufności. Domyślnie, dla różnicy między średnimi wyświetlany jest przedział ufności 95%. Aby zmienić przedział ufności, wprowadź odpowiednią wartość z przedziału od 1 do 99.

Braki danych. Jeśli testowanych jest kilka zmiennych i dla co najmniej jednej z nich brakuje danych, można określić, które obserwacje mają być uwzględnione (lub wykluczone).

- **Wyłączanie obserwacji analiza po analizie.** Każdy test t wykorzystuje wszystkie obserwacje zawierające ważne dane dla testowanej zmiennej. Wielkości prób mogą się różnić w poszczególnych testach.
- **Wyłączanie wszystkich obserwacji z brakami.** Każdy test t wykorzystuje tylko obserwacje zawierające ważne dane dla wszystkich zmiennych użytych w żądanych testach t . Wielkość próby jest stała we wszystkich testach.

Test t dla prób zależnych

Test t dla prób zależnych porównuje średnie dwóch zmiennych z jednej grupy. Procedura oblicza różnice między wartościami dwóch zmiennych dla każdej obserwacji i testuje, czy średnie różnią się od 0. Procedura umożliwia również automatyzację obliczenia wielkości efektu testu t .

Przykład

W badaniach nad wysokim ciśnieniem krwi sprawdza się je wszystkim pacjentom na początku leczenia i po jego zakończeniu. Dlatego każdy obiekt ma dwa pomiary, często zwane pomiarami *przed* i *po*. Alternatywnym modelem, dla którego stosuje się ten test, jest badanie dopasowanych par lub kontroli obserwacji, w przypadku których każdy członek grupy eksperymentalnej i odpowiadający mu członek grupy kontrolnej należą do tej samej obserwacji w pliku danych. W przypadku badania nad wysokim ciśnieniem krwi pacjenci i obiekty kontrolne mogą być dobrani według wieku (75-letni pacjent i 75-letni członek grupy kontrolnej).

Statystyki

Dla każdej zmiennej: średnia, rozmiar próby, odchylenie standardowe i błąd standardowy średniej.
Dla każdej pary zmiennych: korelacja, przeciętna różnica średnich, test t , przedział ufności dla różnicy średnich (można określić poziom ufności) oraz oszacowanie wielkości efektu dla testu t . Odchylenie standardowe oraz błąd standardowy różnicy średnich.

Zagadnienia dotyczące danych

Dane

Dla każdego testu t dla prób zależnych należy podać dwie zmienne ilościowe (interwałowy lub ilorazowy poziom pomiaru). W przypadku badania dopasowanych par lub kontroli obserwacji odpowiedzi dla każdego członka grupy eksperymentalnej i odpowiadającego mu członka grupy kontrolnej muszą należeć do tej samej obserwacji w pliku danych.

Założenia

Obserwacje każdej pary powinny być wykonane w jednakowych warunkach. Różnice średnich powinny posiadać rozkład normalny. Wariancje każdej zmiennej mogą, ale nie muszą być równe.

Wykonywanie testu t dla prób zależnych

1. Z menu wybierz:

Analiza > Porównywanie średnich > Test t dla prób zależnych...

2. Wybierz co najmniej jedną parę zmiennych.

3. Opcjonalnie zmień/wybierz opcję **Szacuj wielkości efektów**. Ustawienia sterują sposobem obliczania standaryzatora podczas szacowania d Cohena oraz poprawki Hedgesa dla każdej pary zmiennych.

Odchylenie standardowe różnicy

Mianownik używany przy szacowaniu wielkości efektu. d Cohena wykorzystuje odchylenie standardowe średniej różnicy próby. Przy obliczaniu poprawki Hedgesa wykorzystuje się odchylenie standardowe średniej różnicy próby skorygowane przez współczynnik poprawki.

Skorygowane odchylenie standardowe różnicy

Mianownik używany przy szacowaniu wielkości efektu. Wartość d Cohena wykorzystuje odchylenie standardowe średniej różnicy próby skorygowane przez korelację między miarami. Przy obliczaniu poprawki Hedgesa wykorzystuje się odchylenie standardowe średniej różnicy próby skorygowane przez korelację między miarami powiększoną o współczynnik poprawki.

Średnia wariancji

Mianownik używany przy szacowaniu wielkości efektu. Wartość d Cohena wykorzystuje pierwiastek kwadratowy z średniej wariancji miar. Przy obliczaniu poprawki Hedgesa wykorzystuje się pierwiastek kwadratowy średniej wariancji miar powiększony o współczynnik poprawki.

4. Opcjonalnie można wykonać następujące kroki:

- Wybierz opcję **Szacuj wielkości efektów**, aby sterować szacowaniem wielkości efektu testu t . Po wybraniu ustawienia można dodatkowo wpływać na sposób obliczania standaryzatora przy szacowaniu d Cohena i poprawki Hedgesa dla każdej pary zmiennych.
- Kliknij przycisk **Opcje**, aby określić przedział ufności oraz sposób postępowania z brakami danych.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

Test t dla prób zależnych: Opcje

Oszacowanie przedziału ufności. Domyślnie, dla różnicy między średnimi wyświetlany jest przedział ufności 95%. Aby zmienić przedział ufności, wprowadź odpowiednią wartość z przedziału od 1 do 99.

Braki danych. Jeśli testowanych jest kilka zmiennych i dla co najmniej jednej z nich brakuje danych, można określić, które obserwacje mają być uwzględnione (lub wykluczone):

- **Wyłączanie obserwacji analiza po analizie.** Każdy test t wykorzystuje wszystkie obserwacje zawierające ważne dane dla testowanych par zmiennych. Wielkości prób mogą się różnić w poszczególnych testach.
- **Wyłączanie wszystkich obserwacji z brakami.** Każdy test t wykorzystuje tylko te obserwacje, które mają poprawne dane dla wszystkich testowanych par zmiennych. Wielkość próby jest stała we wszystkich testach.

Dodatkowe właściwości komendy T-TEST

Język składni komend umożliwia również:

- Tworzenie testu t dla jednej próby lub dla prób niezależnych uruchamiając jedną komendę.
- Testowanie zmiennej względem każdej zmiennej na liście za pomocą testu t dla prób zależnych (przy użyciu opcji komendy PAIRS).

- Możliwe jest sterowanie szacowaniem wielkości efektu testu t (za pomocą opcji komendy ES).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Test t dla jednej próby

Procedura testu t dla jednej próby umożliwia sprawdzenie, czy średnia jednej zmiennej różni się od określonej stałej, i automatyzuje obliczenia wielkości efektu testu t .

Przykłady

Badacz może chcieć sprawdzić, czy średnia ocena IQ dla grupy studentów różni się od 100. Albo producent płatków śniadaniowych może pobrać próbkę opakowań z linii produkcyjnej i sprawdzić, czy średnia masa próbek różni się od 1,3 funta (590 g) na poziomie ufności 95%.

Statystyki

Dla każdej zmiennej testowanej: średnia, odchylenie standardowe, błąd standardowy średniej oraz oszacowanie wielkości efektu dla testu t . Średnia różnica między każdą wartością danych i hipotetyczną wartością testową, test t weryfikujący, czy różnica ta wynosi 0 oraz przedział ufności dla tej różnicy (poziom ufności można określić samodzielnie).

Zagadnienia dotyczące danych

Dane

Aby zweryfikować wartości zmiennej ilościowej względem hipotetycznej wartości testowej, należy wybrać zmienną ilościową i wprowadzić hipotetyczną wartość testową.

Założenia

Ten test zakłada, że dane mają rozkład normalny. Jest on jednak dość odporny na odstępstwa od rozkładu normalnego.

Wykonywanie testu t dla jednej próby

1. Wybierz z menu następującą opcję:

Analizuj > Porównywanie średnich > Test T dla jednej próby...

2. Wybierz jedną lub więcej zmiennych, które mają być przetestowane względem jednej wartości hipotetycznej.
3. Wprowadź numeryczną wartość testową, z którą porównywana ma być średnia z każdej próby.
4. Opcjonalnie można wykonać następujące kroki:

- Wybierz opcję **Szacuj wielkości efektów**, aby sterować szacowaniem wielkości efektu testu t .
- Kliknij przycisk **Opcje**, aby określić przedział ufności oraz sposób postępowania z brakami danych.

Test t dla jednej próby: Opcje

Oszacowanie przedziału ufności. Domyślnie przedział ufności dla różnicy między średnią i hipotetyczną wartością testową ma wartość 95%. Aby zmienić przedział ufności, wprowadź odpowiednią wartość z przedziału od 1 do 99.

Braki danych. Jeśli testowanych jest kilka zmiennych i dla co najmniej jednej z nich brakuje danych, można określić, które obserwacje mają być uwzględnione (lub wykluczone).

- **Wyłączenie obserwacji analiza po analizie.** Każdy test t wykorzystuje wszystkie obserwacje zawierające ważne dane dla testowanej zmiennej. Wielkości prób mogą się różnić w poszczególnych testach.
- **Wyłączenie wszystkich obserwacji z brakami.** Każdy test t wykorzystuje tylko obserwacje zawierające ważne dane dla wszystkich zmiennych użytych w żądanych testach t . Wielkość próby jest stała we wszystkich testach.

Dodatkowe właściwości komendy T-TEST

Język składni komend umożliwia również:

- Tworzenie testu t dla jednej próby lub dla prób niezależnych uruchamiając jedną komendę.
- Testowanie zmiennej względem każdej zmiennej na liście za pomocą testu t dla prób zależnych (przy użyciu opcji komendy PAIRS).
- Możliwe jest sterowanie szacowaniem wielkości efektu testu t (za pomocą opcji komendy ES).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Dodatkowe właściwości komendy T-TEST

Język składni komend umożliwia również:

- Tworzenie testu t dla jednej próby lub dla prób niezależnych uruchamiając jedną komendę.
- Testowanie zmiennej względem każdej zmiennej na liście za pomocą testu t dla prób zależnych (przy użyciu opcji komendy PAIRS).
- Możliwe jest sterowanie szacowaniem wielkości efektu testu t (za pomocą opcji komendy ES).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Jednoczynnikowa ANOVA

Procedura Jednoczynnikowa ANOVA generuje jednoczynnikową analizę wariancji dla ilościowej zmiennej zależnej i pojedynczego czynnika (niezależnego) oraz szacuje wielkość efektu w jednoczynnikowej analizie ANOVA. Analiza wariancji jest wykorzystywana do testowania hipotezy o równości kilku średnich. Technika ta jest rozszerzeniem testu t dla dwóch prób.

Oprócz wykazania różnic między średnimi może zaistnieć potrzeba określenia, które średnie są różne. Istnieją dwa typy testów służące do porównywania średnich: kontrasty a priori i testy post hoc. Kontrasty są testami zdefiniowanymi *przed* przeprowadzeniem eksperymentu, natomiast testy post hoc są przeprowadzane *po* zakończeniu eksperymentu. Można także przeprowadzić test trendów w kategoriach.

Przykład

Podczas gotowania pączki wchłaniają tłuszcz w różnych ilościach. Przeprowadza się eksperyment z użyciem trzech rodzajów tłuszczu: oleju z orzeszków ziemnych, oleju kukurydzianego i smalcu. Olej orzechowy i olej kukurydziany są tłuszczami nienasyconymi, natomiast smalec jest tłuszczem nasyconym. Wraz z ustaleniem, czy ilość wchłoniętego tłuszczu zależy od rodzaju użytego tłuszczu, można zdefiniować kontrast a priori w celu ustalenia, czy ilość wchłoniętego tłuszczu jest różna dla tłuszczów nasyconych i nienasyconych.

Statystyki

Dla każdej grupy: liczba obserwacji, średnia, odchylenie standardowe, błąd standardowy średniej, minimum, maksimum, 95% przedział ufności dla średniej oraz oszacowanie wielkości efektu dla jednoczynnikowej analizy ANOVA. Test Levene'a jednorodności wariancji, tabela analizy wariancji dla każdej zmiennej zależnej, określone przez użytkownika kontrasty a priori oraz testy rozstępów post hoc i wielokrotne porównania: Bonferroniego, Sidaka, rzeczywiście znacząca różnica Tukeya, GT2 Hochberga, Gabriela, Dunnetta, test F Ryana-Einota-Gabriela-Welscha ($F R-E-G-W$), test rozstępów Ryana-Einota-Gabriela-Welscha ($Q R-E-G-W$), T2 Tamhane'a, T3 Dunnetta, Gamesa-Howella, C Dunnetta, test Duncana wielokrotnych rozstępów, Studenta-Newmana-Keulsa (S-N-K), b Tukeya, Wallera-Duncana, Scheff'a i najmniej znacząca różnica.

Zagadnienia dotyczące danych

Dane

Zmienna zależna powinna być zmienną ilościową (poziom interwału pomiaru).

Założenia

Każda grupa jest niezależną próbą losową z normalnej populacji. Analiza wariancji jest odporna na odstępstwa od rozkładu normalnego, ale dane powinny być symetryczne. Grupy powinny pochodzić

z populacji o równych wariancjach. Aby przetestować to założenie, należy użyć testu jednorodności wariancji Levene'a.

Uzyskiwanie jednoczynnikowej analizy wariancji

1. Wybierz z menu następującą opcję:

Analizuj > Porównywanie średnich > Jednoczynnikowa ANOVA...

2. Wybierz co najmniej jedną zmienną zależną.

3. Wybierz pojedynczy czynnik niezależny.

Opcjonalnie można wykonać następujące czynności:

- Wybierz opcję **Szacuj wielkość efektu dla całych testów**, aby wykonać obliczenia wielkości efektu dla całego testu. Gdy ta opcja jest zaznaczona, wyniki zawierają tabelę wielkości efektu ANOVA.
- Kliknij opcję **Kontrasty**, aby podzielić międzygrupowe sumy kwadratów na składniki trendu lub określić kontrasty a priori.
- Kliknij opcję **Post Hoc**, aby użyć testów rozstępów post hoc i wielokrotnych porównań parami w celu określenia, które średnie są różne.
- Kliknij przycisk **Opcje**, aby określić przedział ufności oraz sposób postępowania z brakami danych.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.

Jednoczynnikowa ANOVA: Kontrasty

Międzygrupowe sumy kwadratów można podzielić na składniki trendu lub określić kontrasty a priori.

Wielomianowe

Dzieli międzygrupowe sumy kwadratów na składniki trendu. Można przeprowadzić test trendu zmiennej zależnej dla uporządkowanych poziomów czynnika. Można na przykład przeprowadzić test trendu liniowego (rosnącego lub malejącego) płacy dla uporządkowanych poziomów najwyższego zdobytego wykształcenia.

- **Stopnia.** Można wybrać wielomian od pierwszego do piątego stopnia.

Współczynniki

Określone przez użytkownika kontrasty a priori, które mają być przetestowane przez test *t*. Należy wprowadzić współczynnik dla każdej grupy (kategorii) czynnika i kliknąć przycisk **Dodaj** po każdym wpisie. Każda nowa wartość jest dodawana na dole listy współczynników. Aby określić dodatkowe zestawy kontrastów, kliknij przycisk **Następny**. Do przechodzenia między zestawami kontrastów służą przyciski **Dalej** i **Wstecz**.

Szacuj wielkość efektu dla kontrastów

Steruje obliczaniem wielkości efektu dla całego testu. Gdy to ustawienie jest włączone, należy wybrać co najmniej jedną z następujących opcji, aby obliczyć wielkości efektu. To ustawienie jest włączone, gdy określono co najmniej jeden kontrast, a wyniki w tabeli efektów ANOVA są uwzględnione w wynikach.

Jako standaryzatora używaj sumarycznego odchylenia standardowego dla wszystkich grup

Powoduje wykorzystanie sumarycznego odchylenia standardowego dla wszystkich grup jako standaryzatora przy szacowaniu wielkości efektu. Jest to ustawienie domyślne – dostępne, gdy wybrana jest opcja **Szacuj wielkość efektu dla kontrastów**.

Jako standaryzatora używaj sumarycznego odchylenia standardowego dla grup uczestniczących w kontraście

Powoduje wykorzystanie sumarycznego odchylenia standardowego dla grup uczestniczących w kontraście jako standaryzatora. To ustawienie jest dostępne, gdy wybrana jest opcja **Szacuj wielkość efektu dla kontrastów**.

Istotna jest kolejność współczynników, ponieważ odpowiada ona rosnącej kolejności wartości kategorii czynnika. Pierwszy współczynnik na liście odpowiada najniższej wartości grupy czynnika, zaś ostatni – najwyższej. Przykładowo: jeśli istnieje sześć kategorii czynnika, to współczynniki -1, 0, 0, 0, 0,5 i 0,5 kontrastują pierwszą grupę z piątą i szóstą. W przypadku większości aplikacji współczynniki powinny być sumowane do wartości 0. Zestawy, które nie sumują się do 0, mogą być również używane, ale wyświetlany jest komunikat ostrzegawczy.

Jednoczynnikowa ANOVA: Wielokrotne porównania post hoc

Gdy już zostanie wykazane istnienie różnic między średnimi, za pomocą testów rozstępów post hoc i wielokrotnych porównań parami można określić, które średnie są różne. Testy rozstępów identyfikują podzbiory jednorodnych średnich nie różniących się od siebie. Wielokrotne porównania parami testują różnicę między każdą parą średnich i generują macierz, na której gwiazdki oznaczają znacząco różne średnie grupowe przy poziomie istotności alfa 0,05.

Założenie o równości wariancji

Testy rzeczywiście znaczącej różnicy Tukey'a, GT2 Hochberga, Gabriela i Scheffé'a są testami wielokrotnych porównań oraz testami rozstępu. Inne dostępne testy to *b* Tukeya, *S-N-K* (Studenta-Newmana-Keulsa), *Duncana*, *F R-E-G-W* (test *F* Ryana-Einota-Gabriela-Welscha), *Q R-E-G-W* (test rozstępów Ryana-Einota-Gabriela-Welscha) oraz *Wallera-Duncana*. Dostępne testy wielokrotnych porównań to: test Bonferroni, test rzeczywiście znaczącej różnicy Tukeya, testy Sidaka, Gabriela, Hochberga, Dunnetta, Scheffé'a oraz *LSD* (najmniej istotnej różnicy).

- *LSD*. używa testów *t* do wykonywania wszystkich porównań parami między środkami grupy. Metoda nie kontroluje wzrostu wskaźnika błędów spowodowanego wykonaniem wielokrotnych porównań.
- *Bonferroni*. Metoda porównywania parami średnich grupowych za pomocą testów *t*, z kontrolą poziomu ogólnego błędów poprzez ustawienie poziomu błędów dla każdego testu na wartość równą poziomowi błędów doświadczenia podzielonego przez liczbę testów. Obserwowany poziom istotności uwzględnia fakt przeprowadzania wielu porównań.
- *Sidak*. Test wielokrotnych porównań Pairwise w oparciu o statystykę *t*. Test Sidaka dostosowuje poziom istotności dla porównań wielokrotnych i szacuje węższe granice niż test Bonferroni.
- *Scheffe*. Wykonuje symultaniczne porównania parami dla wszystkich możliwych kombinacji średnich parami. Wykorzystuje rozkład statystyki *F*. Poza porównywaniem par średnich można go stosować do testowania wszystkich możliwych liniowych kombinacji średnich grupowych.
- *F R-E-G-W*. Wielokrotna krokowa zstępująca procedura Ryana-Einota-Gabriela-Welscha oparta na teście *F*.
- *Q R-E-G-W*. Wielokrotna krokowa zstępująca procedura Ryana-Einota-Gabriela-Welscha oparta na studentyzowanym rozstępie.
- *S-N-K*. Dokonuje wszystkich porównań parami między sposobami korzystania ze studentyzowanego rozkładu zakresu. Dla prób o równej wielkości dokonuje również porównań parami średnich w obrębie jednorodnych podzbiorów, wykorzystując metodę krokową. Średnie są porządkowane od największej do najmniejszej, a jako pierwsze testowane są największe różnice między nimi.
- *Tukey*. Wykorzystuje studentyzowaną statystykę zakresu w celu dokonania wszystkich porównań parami między grupami. Ustala poziom ogólnego błędów doświadczenia dla wszystkich porównań parami na poziomie błędów dla zbioru.
- *B Tukey'a*. korzysta ze studentyzowanego rozkładu zakresu w celu dokonywania porównań parami między grupami. Jego wartość krytyczna to średnia z odpowiednich wartości testu rzeczywiście znaczącej różnicy Tukey'a oraz testu Studenta-Newmana-Keulsa.
- *Duncan*. Dokonuje porównań parami przy użyciu krokowego rzędu porównań identycznych z zamówieniem używanym przez test Student-Newman-Keuls, ale ustawia poziom ochrony dla współczynnika błędów dla kolekcji testów, a nie współczynnik błędów dla poszczególnych testów. Używa statystyki opartej na studentyzowanym rozstępie.
- *Hochberg's GT2*. Test wielokrotnych porównań i zakresu, który korzysta ze studentyzowanego maksimum modulo. Podobny do testu rzeczywiście znaczącej różnicy Tukey'a.

- *Gabriel*. Test porównawczy Pairwise, który użył studentyzowanego maksymalnego modułu i jest generalnie bardziej wydajny niż GT2 firmy Hochberg, gdy wielkości komórek są nierówne. Test Gabriela może stać się liberalny, jeśli rozmiary komórek znacznie się różnią.
- *Waller-Duncan*. Test wielokrotnych porównań w oparciu o statystykę t; używa podejścia bayesowskiego.
- *Dunnett*. Test t wielokrotnych porównań parami, porównujący zbiór wyników zabiegów z jedną średnią kontrolną. Domyślnie, kategorią kontrolną jest ostatnia kategoria. Można również wybrać kategorię pierwszą. Do sprawdzania, czy średnia czynnika na każdym poziomie (poza kategorią kontrolną) jest różna od średniej dla kategorii kontrolnej, wykorzystywać należy test **dwustronny**. Aby sprawdzić, czy średnia na którymkolwiek poziomie czynnika jest mniejsza od średniej kategorii kontrolnej, wybierz opcję **< Kontrolna**. Aby sprawdzić, czy średnia na którymkolwiek poziomie czynnika jest większa od średniej kategorii kontrolnej, wybierz opcję **> Kontrolna**.

Brak założenia o równości wariancji

Testy wielokrotnych porównań, które nie zakładają równych wariancji to: T2 Tamhane'a, T3 Dunnetta, Gamesa-Howella oraz C Dunnetta.

- *Tamhane's T2*. Konserwatywne testy porównawcze parami na podstawie testu t. Odpowiedni w przypadku niejednorodnych wariancji.
- *Dunnetta T3*. Test porównawczy Pairwise oparty na studentyzowanym maksymalnym modułowi. Odpowiedni w przypadku niejednorodnych wariancji.
- *Games-Howell*. Test porównania parami, który czasami jest liberalny. Odpowiedni w przypadku niejednorodnych wariancji.
- *C Dunnetta*. Test porównawczy Pairwise oparty na studentyzowanym zakresie. Odpowiedni w przypadku niejednorodnych wariancji.

Uwaga: Interpretacja wyników testów post hoc może okazać się łatwiejsza, jeśli zostanie odznaczone pole wyboru **Ukryj puste wiersze i kolumny** w oknie dialogowym Właściwości tabeli (w aktywnej tabeli przestawnej należy wybrać polecenie **Właściwości tabeli** z menu Format).

Test hipotezy zerowej

Określa, jak poziom istotności (alfa) jest traktowany w teście post hoc.

Użyj tego samego poziomu istotności (alfa), co ustawiony w Opcjach

Gdy ta opcja jest wybrana, używane jest to samo ustawienie, co określone w oknie dialogowym Opcje.

Określ poziom istotności (alfa) dla testu post hoc

Po wybraniu tej opcji można określić poziom istotności (alfa) w polu **Poziom**.

Wykonywanie testów post hoc dla jednoczynnikowej ANOVA

Jednoczynnikowa ANOVA: Opcje

Statystyki

Należy wybrać co najmniej jedną spośród następujących opcji:

Statystyka opisowa

Oblicza liczbę obserwacji, średnią, odchylenie standardowe, standardowy błąd średniej, minimum, maksimum i przedział ufności o wartości 95% dla każdej zmiennej zależnej dla każdej grupy.

Efekty stałe i losowe

Wyświetla odchylenie standardowe, błąd standardowy i przedział ufności na poziomie 95% dla modelu efektów stałych, oraz błąd standardowy, przedział ufności na poziomie 95% i szacunkową wariancję międzyskładnikową dla modelu efektów losowych.

Test jednorodności wariancji

Oblicza test Levene'a w celu przetestowania równości wariancji grupowych. Test ten nie jest uzależniony od założenia o normalności.

Brown-Forsythe

Oblicza statystykę Browna-Forsythe'a w celu przetestowania równości średnich grupowych. Statystyka ta jest preferowana w stosunku do statystyki F , jeśli założenie o równości wariancji się nie sprawdza.

Welch

Oblicza statystykę Welcha w celu przetestowania równości średnich grupowych. Statystyka ta jest preferowana w stosunku do statystyki F , jeśli założenie o równości wariancji się nie sprawdza.

Braki danych

Kontrola sposobu postępowania z brakami danych.

Wyłączanie obserwacji analiza po analizie

W tej analizie nie wykorzystuje się obserwacji z brakami danych dla zmiennej zależnej lub czynnika wykorzystanych w danej analizie. Ponadto nie wykorzystuje obserwacji spoza zakresu określonego dla czynnika.

Wyłącz wszystkie obserwacje z brakami

Z wszystkich analiz wyłączone są obserwacje z brakami danych dla czynnika lub dla jakiegokolwiek zmiennej zależnej występującej na liście Zmienne zależne w głównym oknie dialogowym. Jeśli nie określono wielu zmiennych zależnych, powyższa zasada nie wywiera żadnego skutku.

Przedział ufności

Domyślnie przedział ufności dla różnicy między średnią i hipotetyczną wartością testową ma wartość 95%. Aby zmienić przedział ufności, wprowadź odpowiednią wartość z przedziału od 1 do 99.

Wykres średnich

Wyświetla wykres przedstawiający średnie podgrup (średnie dla każdej grupy zdefiniowane według wartości czynnika).

Określanie opcji dla jednoczynnikowej analizy ANOVA

Dodatkowe właściwości komendy ONEWAY

Język składni komend umożliwia również:

- Otrzymywanie statystycznych efektów stałych i losowych. Odchylenie standardowe, błąd standardowy średniej oraz przedział ufności o wartości 95% dla modelu efektów stałych. Błąd standardowy oraz przedziały ufności (95%) i szacunkowa wariancja międzyskładnikowa dla modelu efektów losowych (za pomocą parametrów STATISTICS=EFFECTS).
- Określanie poziomów alfa dla najmniejszej istotnej różnicy, testów wielokrotnych porównań: Bonferroniego, Duncana i Scheffé'a (za pomocą opcji komendy RANGES).
- Zapisywanie macierzy średnich, odchyłeń standardowych i częstości lub odczytywanie macierzy średnich, częstości, wariancji wspólnych i stopni swobody dla wariancji wspólnych. Te macierze mogą być wykorzystywane zamiast danych surowych w celu przeprowadzenia analizy jednoczynnikowej wariancji (za pomocą opcji komendy MATRIX).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Analiza OML jednej zmiennej

Procedura OML jednej zmiennej umożliwia dokonywanie analizy regresji oraz analizy wariancji dla jednej zmiennej zależnej względem dowolnej liczby czynników i/lub zmiennych. Czynniki dzielą populację na grupy. Przy użyciu tej procedury ogólnego modelu liniowego można sprawdzić hipotezę zerową dotyczącą wpływu innych zmiennych na średnie grupowe pojedynczej zmiennej zależnej. Sprawdzać można interakcje zachodzące między poszczególnymi czynnikami, jak również wpływy poszczególnych czynników, z których niektóre mogą być losowe. Ponadto uwzględnić można wpływ współzmiennych oraz ich interakcje z czynnikami. Dla potrzeb analizy regresji zmienne niezależne (predyktory) mogą być określone jako współzmiennie.

Testować można zarówno modele zrównoważone, jak i niezrównoważone. Plan jest zrównoważony, gdy każda komórka modelu zawiera tę samą liczbę obserwacji. Analiza OML jednej zmiennej umożliwia nie tylko testowanie hipotez, lecz również uzyskiwanie oszacowań parametrów.

Do testowania hipotez wykorzystuje się kontrasty a priori. Ponadto, po ustaleniu całkowitej istotności testu F , przeprowadzać można testy post hoc w celu obliczenia różnic między średnimi. Szacowane średnie brzegowe są przybliżeniami przewidywanych wartości średnich dla poszczególnych komórek modelu. Niektóre zależności można w prosty sposób przedstawić przy użyciu wykresów profili (wykresów interakcji) tych średnich.

Reszty, wartości przewidywane, Odległość Cooka i wartości wpływu można zapisać jako nowe zmienne w pliku danych dla celów sprawdzenia założeń.

W polu WNK Waga można określić zmienną używaną w celu nadania obserwacjom różnych wag podczas analizy metodą ważonych najmniejszych kwadratów (WNK), co kompensuje różne poziomy dokładności pomiarów.

Przykład. Zebrane zostały dane dotyczące poszczególnych uczestników biegów maratońskich odbywających się w Chicago w okresie kilku lat. Zmienną zależną jest czas ukończenia biegu. Czynniki są: pogoda (zimno, komfortowo, gorąco), liczba miesięcy przygotowań, liczba uprzednio ukończonych maratonów oraz płęć. Współzmienną jest wiek. Zauważyć można, że znaczny wpływ na zmienną zależną ma płęć oraz interakcja płci i pogody.

Metody. Poszczególne hipotezy sprawdzać można przy użyciu sum kwadratów typu I, typu II, typu III i typu IV. Typem domyślnym jest typ III.

Statystyki. Testy post hoc rozstępów i porównania wielokrotne: najmniejsza istotna różnica, testy Bonferroniego, Sidaka, Scheffé, 'a, test wielokrotnego F Ryana-Einota-Gabriela-Welscha, test wielozakresowy Ryana-Einota-Gabriela-Welscha, test Studenta-Newmana-Keulsa, test uczciwie istotnej różnicy Tukeya, test b Tukeya, test Duncana, test GT_2 Hochberga, test Gabriela, test t Wallera-Duncana, test Dunnetta (jedno- i dwustronny), test T_2 Tamhane'a, test T_3 Dunnetta, test Gamesa-Howella i test C Dunnetta. Statystyki opisowe: obserwowane średnie, standardowe odchylenia i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach. Testy Levene'a jednorodności wariancji.

Wykresy. Wykresy rozrzut-poziom, reszt i profili (interakcja).

Wymagania dotyczące danych dla OML jednej zmiennej

Dane. Zmienna zależna jest ilościowa. Czynniki są typu jakościowego. Mogą mieć wartości liczbowe lub łańcuchowe składające się z nie więcej niż ośmiu znaków. Współzmiennne są zmiennymi ilościowymi powiązаныmi ze zmienną zależną.

Założenia. Dane są próbą losową populacji normalnej; w populacji wszystkie wariancje komórek są takie same. Analiza wariancji jest odporna na odstępstwa od rozkładu normalnego, ale dane powinny być symetryczne. Do sprawdzenia założeń wykorzystać można jednorodność testów wariancji oraz wykresy rozrzut-poziom. Można również analizować reszty i wykresy reszt.

Otrzymywanie tabel OML jednej zmiennej

1. Wybierz z menu następującą opcję:

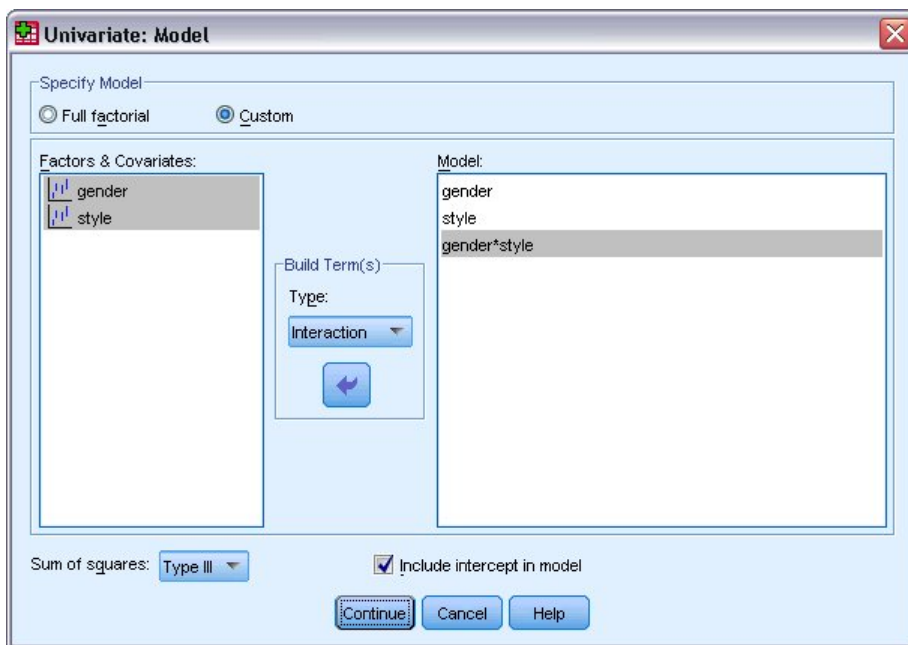
Analiza > Ogólny model liniowy > Jednej zmiennej...

2. Wybierz zmienną zależną.

3. Odpowiednio do swoich danych wybierz zmienne, które mają być Czynnikiami stałymi, Czynnikiami losowymi oraz Współzmiennymi.

4. Opcjonalnie można użyć pola WNK Waga w celu określenia zmiennej ważącej do ważonej analizy najmniejszych kwadratów. Jeśli wartość zmiennej ważącej wynosi zero, jest ujemna lub jest brakującą wartością to dana obserwacja zostaje wyłączona z analizy. Zmienna, która już została włączona do modelu nie może być użyta jako zmienna ważąca.

Model OML



Rysunek 1. Okno dialogowe Model jednej zmiennej

Określ model. Pełny model czynnikowy obejmuje efekty główne wszystkich czynników i współzmiennych oraz interakcje wszystkich czynników. Nie obejmuje interakcji współzmiennych. Aby samodzielnie określić tylko podzbiór interakcji lub interakcje czynnik-współzmienna, wybierz opcję **Użytkownika**. Należy określić wszystkie składniki modelu.

Czynniki i współzmiennie. Lista zawiera czynniki i współzmiennie.

Model. Model zależy od charakteru danych. Po wybraniu opcji **Użytkownika** można wybrać efekty główne oraz interakcje będące przedmiotem zainteresowania w czasie analizy.

Suma kwadratów. Metoda obliczania sum kwadratów. Dla modeli zrównoważonych lub niezrównoważonych bez komórek z brakami danych najczęściej wykorzystywana jest metoda bazująca na sumie kwadratów typu III.

Uwzględnij wyraz wolny w modelu. Wyraz wolny jest zwykle uwzględniany w modelu. Jeśli można założyć, że dane przechodzą przez początek układu współrzędnych, to wyraz wolny można wyłączyć z modelu.

Budowanie składników i składniki użytkownika

Buduj składniki

Tej opcji należy użyć, chcąc uwzględnić niezagnieżdżone składniki konkretnego typu (takie jak efekty główne) dla wszystkich kombinacji wybranego zestawu czynników i współzmiennych.

Buduj składniki użytkownika

Tej opcji należy użyć, chcąc uwzględnić składniki zagnieżdżone lub konstruować składnik jawnie, zmienna po zmiennej. Proces budowania składnika zagnieżdżonego obejmuje następujące kroki:

Suma kwadratów

Dla danego modelu można wybrać typ sumy kwadratów. Najczęściej używanym i jednocześnie domyślnym typem jest typ III.

Typ I. Ta metoda jest znana również jako hierarchiczna dekompozycja metody bazującej na sumie kwadratów. Każdy składnik jest dostosowywany tylko do poprzedzającego go w modelu składnika. Suma kwadratów typu I używana jest zwykle w następujących zastosowaniach:

- Zrównoważony model ANOVA, w którym wszystkie główne efekty określone są przed efektami interakcji pierwszego rzędu, a wszystkie efekty interakcji pierwszego rzędu określone są przed efektami interakcji drugiego rzędu.
- Model regresji wielomianowej, w którym składniki niższego rzędu określone są przed wszystkimi składnikami wyższych rzędów.
- Model w pełni zagnieżdżony, w którym pierwszy z określanych efektów zagnieżdżony jest w drugim, ten z kolei w efekcie określanym jako trzeci itd. (ten sposób zagnieżdżania można określić jedynie za pomocą składni).

Typ II. W metodzie tej obliczana jest suma kwadratów efektu w modelu przy uwzględnieniu wszystkich pozostałych „odpowiednich” efektów. Odpowiedni efekt to taki, który odnosi się do wszystkich efektów niezawierających badanego efektu. Metoda bazująca na sumie kwadratów typu II używana jest zwykle w następujących zastosowaniach:

- Zrównoważony model ANOVA.
- Dowolny model zawierający jedynie efekty główne czynnika.
- Dowolny model regresji.
- Plan w pełni zagnieżdżony (ten sposób zagnieżdżania można określić za pomocą składni).

Typ III. Wartość domyślna. W tej metodzie suma kwadratów efektu jest obliczana jako suma kwadratów uwzględniająca wszelkie inne efekty niezawierające tego efektu i ortogonalne względem wszelkich efektów, które ten efekt zawierają. Sumy kwadratów typu III mają tę zaletę, że są niezmiennicze ze względu na częstości w komórkach, jeśli ogólna forma szacowania jest stała. Suma kwadratów tego typu jest więc przydatna w niezrównoważonych modelach, które nie zawierają brakujących komórek. W planie czynnikowym bez brakujących komórek metoda ta jest równoważna technice ważonych kwadratów średnich Yatesa. Metoda bazująca na sumie kwadratów typu III używana jest zwykle w następujących zastosowaniach:

- Wszystkie modele wymienione dla typu I i typu II.
- Dowolny, zrównoważony lub niezrównoważony, model niezawierający pustych komórek.

Typ IV. Ta metoda została opracowana do wykorzystania w sytuacjach, w których występują brakujące komórki. Dla każdego efektu F w planie, jeśli F nie jest zawarty w żadnym innym efekcie, wówczas typ IV = typ III = typ II. Jeśli czynnik F jest zawarty w innym efekcie, typ IV przekazuje kontrasty między poszczególnymi parametrami czynnika F równomiernie do wszystkich efektów wyższego rzędu. Metoda bazująca na sumie kwadratów typu IV używana jest zwykle w następujących zastosowaniach:

- Wszystkie modele wymienione dla typu I i typu II.
- Dowolny, zrównoważony lub niezrównoważony, model zawierający puste komórki.

OML: Kontrasty

Kontrasty wykorzystywane są do sprawdzania różnic między poziomami czynnika. Dla każdego czynnika modelu określić można kontrast (w modelu powtarzanych pomiarów – dla każdego czynnika międzyobiektowego). Kontrasty przedstawiają liniowe kombinacje parametrów.

OML jednej zmiennej. Testowanie hipotez oparte jest na hipotezie zerowej $\mathbf{LB}=\mathbf{0}$, gdzie \mathbf{L} jest macierzą współczynników kontrastów, a \mathbf{B} jest wektorem parametrów. Po określeniu kontrastu tworzona jest macierz \mathbf{L} . Kolumny macierzy \mathbf{L} reprezentujące czynniki odpowiadają kontrastowi. Pozostałe kolumny dopasowane są w sposób umożliwiający szacowanie macierzy \mathbf{L} .

Raport zawiera statystykę F dla każdego zbioru kontrastów. Dla różnic kontrastów wyświetlane są również symultaniczne przedziały ufności typu Bonferroniego, oparte na rozkładzie t Studenta.

Dostępne kontrasty

Dostępne są następujące kontrasty: kontrast odchylenia, prosty, różnicy, Helmerta, powtórzony i wielomianowy. Dla kontrastu odchylenia i kontrastu prostego można wybrać, czy kategorią odniesienia będzie kategoria ostatnia czy pierwsza.

Typy kontrastów

Odchylenie. Porównuje średnią każdego poziomu (z wyjątkiem kategorii odniesienia) ze średnią wszystkich poziomów (średnią ogólną). Poziomy czynnika mogą mieć dowolną kolejność.

Prosty. Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu ze średnią wybranego poziomu. Ten typ kontrastu jest przydatny szczególnie w przypadku korzystania z grupy kontrolnej. Jako odniesienie wybrać można kategorię pierwszą lub ostatnią.

Różnica. Porównuje średnią każdego poziomu (z wyjątkiem pierwszego) ze średnią poprzednich poziomów. (typ ten określany jest także mianem odwrotnego kontrastu Helmerta).

Helmerta. Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu czynnika (prócz ostatniego) ze średnią poziomów następných.

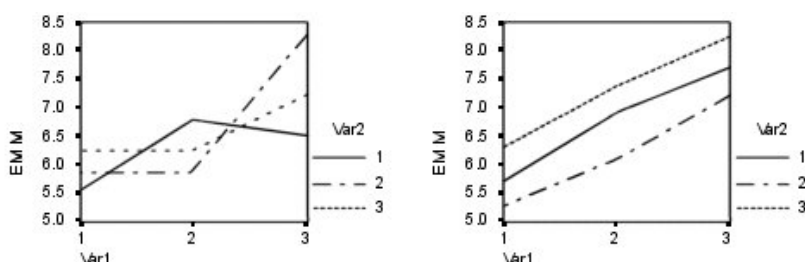
Powtórzone. Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz ostatniego) ze średnią poziomów następných.

Wielomianowy. Porównuje efekt liniowy, efekt kwadratowy, efekt sześcienny itd. Dla wszystkich kategorii efekt liniowy zawarty jest w pierwszym stopniu swobody; efekt kwadratowy - w drugim stopniu swobody itd. Tego typu kontrasty używane są często do szacowania trendów wielomianowych.

OML: Wykresy profili

Wykresy profili (interaktywne) przydatne są do porównywania średnich brzegowych modelu. Wykres profilu jest wykresem liniowym, w którym każdy punkt wskazuje oszacowaną średnią brzegową zmiennej zależnej (skorygowaną o ewentualne wartości współzmiennych) przy jednym poziomie czynnika. Poziomy drugiego czynnika mogą być wykorzystywane do tworzenia osobnych linii. Każdy poziom trzeciego czynnika można wykorzystać do utworzenia osobnego wykresu. Wszystkie czynniki stałe i losowe, o ile istnieją, również mogą być używane do tworzenia wykresów. Podczas analiz wielowymiarowej wykresy profili tworzone są dla każdej zmiennej zależnej. Podczas analizy powtarzanych pomiarów na wykresach profili uwzględnione mogą być zarówno czynniki międzyobiektowe i wewnątrzobiektowe. Analiza OML wielu zmiennych oraz OML powtarzanych pomiarów dostępne są wyłącznie po zainstalowaniu modułu Statystyki zaawansowane.

Wykres profilu jednego czynnika pokazuje, czy oszacowane średnie brzegowe mają tendencje rosnące czy malejące dla poszczególnych poziomów. Dla dwóch lub więcej czynników linie równoległe wskazują, że między czynnikami nie ma interakcji, co oznacza, że badać można poziomy tylko jednego czynnika. Linie nierównoległe wskazują interakcje.



Rysunek 2. Wykres nierównoległy (lewy) i równoległy (prawy)

Po określeniu wykresu przez wybranie czynników dla osi poziomej i, opcjonalnie, czynników dla oddzielnych linii i oddzielnych wykresów, wykres musi być dodany do listy Wykresy.

Typ wykresu

Dostępne są wykresy liniowe lub słupkowe.

Słupki błędów

Można uwzględnić słupki błędów reprezentujące przedziały ufności albo liczbę błędów standardowych. Przedział ufności wyznaczany jest na podstawie poziomu istotności określonego w oknie dialogowym Opcje.

Uwzględniaj linię referencyjną średniej ogólnej

Powoduje uwzględnienie średniej referencyjnej oznaczającej średnią ogólną.

Oś Y rozpoczyna się od 0

W przypadku wykresów liniowych ze wszystkimi dodatnimi lub wszystkimi ujemnymi wartościami wymusza początek osi Y na wartości 0. Wykresy słupkowe zawsze zaczynają się od wartości 0 (lub ją zawierają).

Opcje OML

W tym oknie dialogowym dostępne są statystyki opcjonalne. Statystyki są obliczane przy użyciu modelu efektów stałych.

Pokaż. Zaznaczenie pola **Statystyki opisowe** powoduje wyświetlenie obserwowanych średnich, odchyłeń standardowych i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach. Zaznaczenie pola **Oceny wielkości efektu** powoduje wyświetlenie częściowej wartości eta kwadrat dla każdego efektu i każdej oceny parametru. Eta kwadrat jest statystyką opisującą część całkowitej zmienności, którą można przypisać czynnikowi. Zaznaczenie pola **Obserwowana siła** powoduje wyświetlenie siły testu po ustaleniu hipotezy alternatywnej na podstawie obserwowanej wartości. Zaznaczenie pola **Oceny parametrów** powoduje wyświetlenie ocen parametrów, błędów standardowych, testów t , przedziałów ufności i obserwowanej siły każdego testu. Zaznaczenie pola **Macierz współczynników kontrastów** powoduje otrzymanie macierzy **L**.

Testy jednorodności realizują testy Levene'a jednorodności wariancji dla każdej zmiennej zależnej i każdej kombinacji poziomów czynników międzyobiektowych albo tylko dla czynników międzyobiektowych. Do sprawdzenia założeń dotyczących danych wykorzystać można opcję Wykresy rozrzut-poziom i Wykres reszt. W przypadku braku czynników, opcja ta jest niedostępna. Zaznaczenie opcji **Wykres reszt** powoduje utworzenie wykresu reszt rzeczywistych-szacowanych-standaryzowanych dla każdej zmiennej zależnej. Wykresy te są przydatne podczas sprawdzania założeń równej wariancji. Zaznaczenie pola **Brak dopasowania** powoduje sprawdzenie, czy model nadaje się do prawidłowego opisu relacji zachodzącej między zmienną zależną a zmiennymi niezależnymi. **Ogólne funkcje estymowalne** umożliwiają konstruowanie testów niestandardowych hipotez na podstawie ogólnych funkcji estymowalnych. Wiersze w każdej macierzy współczynników kontrastu są liniowymi kombinacjami ogólnych funkcji estymowalnych.

Dostępne są **testy heteroskedastyczności** służące do sprawdzania, czy wariancja błędów (dla każdej zmiennej zależnej) zależy od wartości zmiennych niezależnych. W przypadku **testu Breuscha-Pagana**, **zmodyfikowanego testu Breuscha-Pagana** i **testu F** można określić model, na którym ma bazować test. Domyślnie model składa się ze składnika stałego, składnika liniowego w wartościach przewidywanych, składnika kwadratowego w wartościach przewidywanych i składnika błędów.

Oszacowania parametrów z odpornymi błędami standardowymi generują tabelę oszacowań parametrów oraz błędy standardowe odporne bądź o spójnej heteroskedastyczności, a także statystyki t , wartości istotności oraz przedziały ufności wyznaczone przy użyciu odpornych błędów standardowych. Dostępnych jest pięć różnych metod estymacji odpornej macierzy kowariancji.

HC0

Oparta na oryginalnym asymptotycznym lub bazującym na dużej próbie, odpornym, empirycznym lub „kanapkowym” estymatorze macierzy kowariancji oszacowań parametrów. Środkowa część „kanapki” zawiera reszty kwadratowe wyznaczone zwykłą metodą najmniejszych kwadratów (OLS) lub metodą ważonych najmniejszych kwadratów (WLS).

HC1

Modyfikacja HC0 z próbą skończoną, przemnożoną przez $N/(N-p)$, gdzie N jest wielkością próby, a p jest liczbą nienadmiarowych parametrów w modelu.

HC2

Modyfikacja HC0, w której reszty kwadratowe dzielone są przez $1-h$, gdzie h jest wartością wpływu obserwacji.

HC3

Modyfikacja HC0 będąca przybliżeniem estymatora scyzorykowego (jackknife). Reszty kwadratowe dzielone są przez kwadrat $1-h$.

HC4

Modyfikacja HC0, w której reszty kwadratowe są dzielone przez 1-h do potęgi, której wykładnik zależy od h, N i p, ale nie może być wyższy niż 4.

Poziom istotności. Może zająć potrzeba skorygowania poziomu istotności używanego w testach post hoc oraz poziomu ufności wykorzystywanego do określania przedziałów ufności. Określona wartość jest również używana do obliczenia obserwowanej siły testu. Podczas określania poziomu istotności powiązany z nim poziom przedziałów ufności wyświetlany jest w oknie dialogowym.

Dodatkowe właściwości komendy UNIANOVA

Język składni komend umożliwia również:

- Określanie zagnieżdżonych efektów w planie (za pomocą opcji DESIGN).
- Określanie testów efektów w odniesieniu do liniowej kombinacji efektów lub wartości (za pomocą opcji komendy TEST).
- Określanie wielu kontrastów (za pomocą opcji komendy CONTRAST).
- Uwzględnianie braków danych zdefiniowanych przez użytkownika (za pomocą opcji komendy MISSING).
- Określanie kryteriów EPS (za pomocą opcji komendy CRITERIA).
- Tworzenie macierzy **L**, **M** lub **K** użytkownika (za pomocą opcji komendy LMATRIX, MMATRIX i KMATRIX).
- Określanie pośredniej kategorii odniesienia dla kontrastów odchyłeń lub kontrastów prostych (za pomocą opcji komendy CONTRAST).
- Określanie metryk kontrastów wielomianowych (za pomocą opcji komendy CONTRAST).
- Określanie składników błędu dla porównań post hoc (za pomocą opcji komendy POSTHOC).
- Obliczanie szacowanych średnich brzegowych dla czynników bądź interakcji między czynnikami umieszczonymi na liście czynników (za pomocą opcji komendy EMMEANS).
- Określanie nazw zmiennych tymczasowych (za pomocą opcji komendy SAVE).
- Tworzenie pliku danych zawierającego macierz korelacji (za pomocą opcji komendy OUTFILE).
- Tworzenie pliku macierzowego zawierającego statystyki z międzyobiektowej tabeli ANOVA (za pomocą opcji komendy OUTFILE).
- Zapisywanie macierzy planu w nowym pliku danych (za pomocą opcji komendy OUTFILE).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

OML: Porównania post hoc

Testy wielokrotnych porównań post hoc. Gdy już zostanie wykazane istnienie różnic między średnimi, za pomocą testów rozstępów post hoc i wielokrotnych porównań parami można określić, które średnie są różne. Porównania dotyczą wartości nieskorygowanych. Testy te używane są wyłącznie dla stałych czynników międzyobiektowych. W OML Powtarzane pomiary testy takie są niedostępne, jeśli nie występują czynniki międzyobiektowe, a testy wielokrotnych porównań post hoc są wykonywane dla średniej na wszystkich poziomach czynników wewnątrzobiektowych. W przypadku OML Wielu zmiennych testy post hoc są wykonywane oddzielnie dla każdej zmiennej zależnej. Analiza OML wielu zmiennych oraz OML powtarzanych pomiarów dostępne są wyłącznie po zainstalowaniu modułu Statystyki zaawansowane.

Często stosowanymi testami wielokrotnych porównań są testy rzeczywiście znaczących różnic Tukeya i Bonferroniego. **Test Bonferroniego**, oparty na statystyce *t* Studenta, koryguje obserwowany poziom istotności ze względu na fakt realizacji porównań wielokrotnych. **Test t Sidaka** dostosowuje również poziom istotności i zapewnia węższe granice niż test Bonferroniego. **Test rzeczywiście znaczącej różnicy Tukeya** wykorzystuje studentyzowaną statystykę rozstępu do dokonywania wszystkich porównań parami pomiędzy grupami. Ustala poziom ogólnego błędu doświadczenia na poziomie błędu dla zbioru, dla wszystkich porównań parami. Podczas testowania dużej liczby par średnich test rzeczywiście znaczącej różnicy Tukeya posiada większą moc od testu Bonferroniego. Dla niewielkiej liczby par test Bonferroniego posiada większą moc.

Test **GT2 Hochberga** jest podobny do testu rzeczywiście znaczącej różnicy Tukeya, ale używany jest w nim studentyzowany największy moduł. Zwykle test Tukey'a posiada większą moc. **Test porównań parami Gabriela** również wykorzystuje studentyzowany największy moduł i posiada zwykle większą moc niż test GT2 Hochberga w przypadku nierównych rozmiarów komórek. Test Gabriela może stać się liberalny, jeśli rozmiary komórek znacznie się różnią.

Test t wielokrotnych porównań parami Dunnetta porównuje zestaw wyników działania czynników z pojedynczą średnią kontrolną. Domyślnie, kategorią kontrolną jest ostatnia kategoria. Można również wybrać kategorię pierwszą. Dostępny jest test dwustronny i jednostronny. Do sprawdzania, czy średnia czynnika na każdym poziomie (poza kategorią kontrolną) jest różna od średniej dla kategorii kontrolnej, wykorzystywać należy test dwustronny. Aby sprawdzić, czy średnia czynnika na każdym poziomie jest mniejsza od średniej dla kategorii kontrolnej, wybierz opcję **< Kontrola**. Podobnie, aby sprawdzić, czy średnia czynnika na każdym poziomie jest większa od średniej dla kategorii kontrolnej, wybierz opcję **> Kontrola**.

Ryan, Einot, Gabriel i Welsch (R-E-G-W) opracowali dwa wielokrotne zstępujące testy rozstępów. Wielokrotne procedury zstępujące najpierw sprawdzają, czy wszystkie średnie są równe. Jeśli nie, to ze względu na równość sprawdzane są podzbiory średnich. Test **F R-E-G-W** jest oparty na teście F , a test **Q R-E-G-W** jest oparty na studentyzowanym rozstępie. Testy te mają większe możliwości niż test wielozakresowy Duncana i test Studenta-Newmana-Keulsa (które także są procedurami obejmującymi wiele kroków), ale nie zaleca się ich w przypadku nierównych rozmiarów komórek.

W przypadku niejednorodności wariancji korzystać można z testu **T2 Tamhane'a** (konserwatywny test porównań parami oparty na teście t), **T3 Dunnetta** (porównań parami oparty na studentyzowanym największym module), testu **Gamesa-Howella porównań parami** (czasem liberalny) lub **testu C Dunnetta** (porównań parami oparty na studentyzowanym rozstępie). Należy zwrócić uwagę, że testy te są nieprawidłowe i nie zostaną wygenerowane, jeśli model zawiera wiele czynników.

Test **Duncana wielokrotnych rozstępów**, Studenta-Newmana-Keulsa (**S-N-K**) oraz **b Tukeya** to testy rozstępów rangujące średnie grupowe i obliczające wartość rozstępu. Testy te nie są używane tak często, jak testy omówione uprzednio.

Test t Wallera-Duncana wykorzystuje podejście bayesowskie. Test ten korzysta ze średniej harmonicznej wielkości próby, kiedy rozmiary prób nie są równe.

Poziom istotności testu **Scheffégo** ma za zadanie umożliwić przetestowanie wszystkich możliwych liniowych kombinacji średnich grup, nie tylko porównań parami udostępnianych przez te funkcje. W rezultacie test Scheffégo daje często wyniki bardziej zachowawcze niż inne testy, co oznacza, że dla istotności wymagana jest większa różnica między średnimi.

Test najmniejszej istotnej różnicy (**NIR**) wielokrotnych porównań parami jest równoznaczny wielu osobnym testom t pomiędzy wszystkimi parami grup. Wada tej metody polega na tym, że nie dostosowuje ona obserwowanego poziomu istotności dla wielokrotnych porównań.

Wyświetlane testy. Porównania parami są wykonywane w przypadku testów LSD, Sidaka, Bonferroniego, Gamesa-Howella, testów T2 i T3 Tamhane'a oraz testów C i T3 Dunnetta. Podzbiory jednorodne dla testów rozstępów dostępne są dla testów S-N-K, b Tukeya, Duncana, F R-E-G-W, Q R-E-G-W i Wallera. Testy rzeczywiście znaczącej różnicy Tukey'a, GT2 Hochberga, Gabriela i Scheffé'a są zarówno testami wielokrotnych porównań, jak i testami rozstępu.

Opcje OML

W tym oknie dialogowym dostępne są statystyki opcjonalne. Statystyki są obliczane przy użyciu modelu efektów statych.

Pokaż. Zaznaczenie pola **Statystyki opisowe** powoduje wyświetlenie obserwowanych średnich, odchyłeń standardowych i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach. Zaznaczenie pola **Oceny wielkości efektu** powoduje wyświetlenie częściowej wartości eta kwadrat dla każdego efektu i każdej oceny parametru. Eta kwadrat jest statystyką opisującą część całkowitej zmienności, którą można przypisać czynnikowi. Zaznaczenie pola **Obserwowana siła** powoduje wyświetlenie siły testu po ustaleniu hipotezy alternatywnej na podstawie obserwowanej wartości. Zaznaczenie pola **Oceny parametrów** powoduje wyświetlenie ocen parametrów, błędów standardowych, testów t , przedziałów

ufności i obserwowanej siły każdego testu. Zaznaczenie pola **Macierz współczynników kontrastów** powoduje otrzymanie macierzy **L**.

Testy jednorodności realizują testy Levene'a jednorodności wariancji dla każdej zmiennej zależnej i każdej kombinacji poziomów czynników międzyobiektowych albo tylko dla czynników międzyobiektowych. Do sprawdzenia założeń dotyczących danych wykorzystać można opcję Wykresy rozrzut-poziom i Wykres reszt. W przypadku braku czynników, opcja ta jest niedostępna. Zaznaczenie opcji **Wykres reszt** powoduje utworzenie wykresu reszt rzeczywistych-szacowanych-standaryzowanych dla każdej zmiennej zależnej. Wykresy te są przydatne podczas sprawdzania założeń równej wariancji. Zaznaczenie pola **Brak dopasowania** powoduje sprawdzenie, czy model nadaje się do prawidłowego opisu relacji zachodzącej między zmienną zależną a zmiennymi niezależnymi. **Ogólne funkcje estymowalne** umożliwiają konstruowanie testów niestandardowych hipotez na podstawie ogólnych funkcji estymowalnych. Wiersze w każdej macierzy współczynników kontrastu są liniowymi kombinacjami ogólnych funkcji estymowalnych.

Dostępne są **testy heteroskedastyczności** służące do sprawdzania, czy wariancja błędów (dla każdej zmiennej zależnej) zależy od wartości zmiennych niezależnych. W przypadku **testu Breuscha-Pagana**, **zmodyfikowanego testu Breuscha-Pagana** i **testu F** można określić model, na którym ma bazować test. Domyślnie model składa się ze składnika stałego, składnika liniowego w wartościach przewidywanych, składnika kwadratowego w wartościach przewidywanych i składnika błędu.

Oszacowania parametrów z odpornymi błędami standardowymi generują tabelę oszacowań parametrów oraz błędy standardowe odporne bądź o spójnej heteroskedastyczności, a także statystyki *t*, wartości istotności oraz przedziały ufności wyznaczone przy użyciu odpornych błędów standardowych. Dostępnych jest pięć różnych metod estymacji odpornej macierzy kowariancji.

HC0

Oparta na oryginalnym asymptotycznym lub bazującym na dużej próbie, odpornym, empirycznym lub „kanapkowym” estymatorze macierzy kowariancji oszacowań parametrów. Środkowa część „kanapki” zawiera reszty kwadratowe wyznaczone zwykłą metodą najmniejszych kwadratów (OLS) lub metodą ważonych najmniejszych kwadratów (WLS).

HC1

Modyfikacja HC0 z próbą skończoną, przemnożoną przez $N/(N-p)$, gdzie *N* jest wielkością próby, a *p* jest liczbą nienadmiarowych parametrów w modelu.

HC2

Modyfikacja HC0, w której reszty kwadratowe dzielone są przez $1-h$, gdzie *h* jest wartością wpływu obserwacji.

HC3

Modyfikacja HC0 będąca przybliżeniem estymatora scyzorykowego (jackknife). Reszty kwadratowe dzielone są przez kwadrat $1-h$.

HC4

Modyfikacja HC0, w której reszty kwadratowe są dzielone przez $1-h$ do potęgi, której wykładnik zależy od *h*, *N* i *p*, ale nie może być wyższy niż 4.

Poziom istotności. Może zająć potrzeba skorygowania poziomu istotności używanego w testach post hoc oraz poziomu ufności wykorzystywanego do określania przedziałów ufności. Określona wartość jest również używana do obliczenia obserwowanej siły testu. Podczas określania poziomu istotności powiązany z nim poziom przedziałów ufności wyświetlany jest w oknie dialogowym.

Dodatkowe właściwości komendy UNIANOVA

Język składni komend umożliwia również:

- Określanie zagnieżdżonych efektów w planie (za pomocą opcji DESIGN).
- Określanie testów efektów w odniesieniu do liniowej kombinacji efektów lub wartości (za pomocą opcji komendy TEST).
- Określanie wielu kontrastów (za pomocą opcji komendy CONTRAST).
- Uwzględnianie braków danych zdefiniowanych przez użytkownika (za pomocą opcji komendy MISSING).

- Określanie kryteriów EPS (za pomocą opcji komendy CRITERIA).
- Tworzenie macierzy **L**, **M** lub **K** użytkownika (za pomocą opcji komendy LMATRIX, MMATRIX i KMATRIX).
- Określanie pośredniej kategorii odniesienia dla kontrastów odchyłeń lub kontrastów prostych (za pomocą opcji komendy CONTRAST).
- Określanie metryk kontrastów wielomianowych (za pomocą opcji komendy CONTRAST).
- Określanie składników błędu dla porównań post hoc (za pomocą opcji komendy POSTHOC).
- Obliczanie szacowanych średnich brzegowych dla czynników bądź interakcji między czynnikami umieszczonymi na liście czynników (za pomocą opcji komendy EMMEANS).
- Określanie nazw zmiennych tymczasowych (za pomocą opcji komendy SAVE).
- Tworzenie pliku danych zawierającego macierz korelacji (za pomocą opcji komendy OUTFILE).
- Tworzenie pliku macierzowego zawierającego statystyki z międzyobiektywnej tabeli ANOVA (za pomocą opcji komendy OUTFILE).
- Zapisywanie macierzy planu w nowym pliku danych (za pomocą opcji komendy OUTFILE).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

OML: Zapisz

Pozwala na zapisanie wartości przewidywanych przez model, wartości reszt regresji i innych powiązanych z nimi miar jako nowych zmiennych w Edytorze danych. Wiele z tych zmiennych można wykorzystać do sprawdzania założeń dotyczących danych. Aby zapisać wartości do wykorzystania w innej sesji programu IBM SPSS Statistics, należy zapisać bieżący plik danych.

Wartości przewidywane. Wartości, które model przewiduje dla każdej obserwacji.

- *Niestandardyzowane.* Wartość zmiennej zależnej przewidywana przez model.
- *Ważone.* Ważone niestandardyzowane wartości przewidywane. Dostępne tylko wtedy, gdy wcześniej została wybrana zmienna WNK.
- *Błąd standardowy.* Oszacowanie odchylenia standardowego średniej wartości zmiennej zależnej dla obserwacji, które mają takie same wartości zmiennych niezależnych.

Diagnostyka. Narzędzia do identyfikowania obserwacji z niezwykłą kombinacją wartości zmiennych niezależnych oraz obserwacji, które mogą w znacznym stopniu wpływać na model.

- *Odległość Cooka.* Miara stopnia, w jakim zmieniłyby się wskaźniki reszt dla wszystkich obserwacji przy wykluczeniu poszczególnych obserwacji z obliczeń współczynników regresji. Duże wartości odległości Cooka wskazują na to, że usunięcie obserwacji z obliczeń statystyk regresyjnych zmienia istotnie wielkość tych współczynników.
- *Wartości wpływu.* Niecentrowane wartości dźwigni. Względny wpływ każdej obserwacji na dopasowanie modelu.

Reszty. Reszta niestandardyzowana jest faktyczną wartością zmiennej zależnej pomniejszoną o wartość przewidywaną przez model. Dostępne są również reszty standaryzowane, studentyzowane i usunięte. Jeśli wybrana została zmienna WNK, to dostępne są również ważone reszty niestandardyzowane.

- *Niestandardyzowane.* Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- *Ważone.* Ważone reszt niestandardyzowanych. Dostępne tylko wtedy, gdy wcześniej została wybrana zmienna WNK.
- *Standaryzowane.* Reszta podzielona przez oszacowanie jego odchylenia standardowego. Standaryzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Studentyzowane.* Reszta podzielona przez oszacowanie jego odchylenia standardowego, które różni się w zależności od przypadku, w zależności od odległości między wartościami poszczególnych obserwacji w zmiennych niezależnych od średnich zmiennych niezależnych. Czasami nazywane wewnętrznymi resztami studentyzowanymi.

- *Usunięte*. Rezydualna dla przypadku, gdy ta sprawa jest wykluczona z obliczenia współczynników regresji. Jest to różnica pomiędzy wartością zmiennej zależnej, a jej skorygowaną wartością przewidywaną.

Statystyki współczynników. Tworzy macierz wariancji-kowariancji ocen parametrów modelu w nowym zbiorze danych w ramach bieżącej sesji lub w zewnętrznym pliku danych IBM SPSS Statistics. Ponadto dla każdej zmiennej zależnej istnieje wiersz oszacowań parametrów, wiersz błędów standardowych oszacowań parametrów, wiersz wartości istotności dla statystyk *t* odpowiadających oszacowaniom parametrów i wiersz stopni swobody reszt. W modelach o wielu zmiennych tego typu wiersze tworzone są dla każdej zmiennej zależnej. Gdy wybrana jest statystyka o spójnej heteroskedastyczności (dostępna tylko dla modeli jednej zmiennej), macierz wariancji-kowariancji obliczana jest przy użyciu estymatora odpornego, wiersz błędów standardowych zawiera odporne błędy standardowe, a wartości istotności odzwierciedlają błędy odporne. Macierzy tej można użyć w innych procedurach, które odczytują pliki macierzy.

Oszacowane średnie brzegowe OML

Wybierz czynniki i interakcje, dla których utworzone mają być oceny średnich brzegowych populacji w komórkach. Średnie te są korygowane o wartości współzmiennych, o ile jakiegokolwiek występują.

Porównaj efekty główne

Dostarcza nieskorygowanych porównań parami szacowanych średnich brzegowych dla dowolnego efektu głównego modelu, zarówno dla czynników międzyobiektowych, jak i wewnątrzobiektowych. Pole to można zaznaczyć tylko jeśli na liście Pokaż średnie dla wybrane są efekty główne.

Porównaj proste efekty główne

To ustawienie jest włączone, zawsze gdy lista docelowa zawiera co najmniej jeden iloczyn lub efekt interakcji (na przykład $A \times B$, $A \times B \times C$). Ustawienie to obsługuje specyfikację porównań między prostymi efektami głównymi, które są głównymi efektami zagnieżdżonymi w obrębie poziomów innych czynników.

Korekta przedziału ufności

Dostępne są następujące metody korygowania przedziałów ufności oraz istotności: NIR (najmniejszej istotnej różnicy), Bonferroniego i Sidaka. Ten element jest dostępny tylko wtedy, gdy wybrana jest opcja **Porównaj efekty główne** i/lub **Porównaj proste efekty główne**.

Określanie oszacowanych średnich brzegowych

1. Z menu wybierz jedną z procedur dostępnych po wybraniu opcji > **Analiza** > **Ogólny model liniowy**.
2. W głównym oknie dialogowym kliknij opcję **Średnie EM**.

Opcje OML

W tym oknie dialogowym dostępne są statystyki opcjonalne. Statystyki są obliczane przy użyciu modelu efektów statycznych.

Pokaż. Zaznaczenie pola **Statystyki opisowe** powoduje wyświetlenie obserwowanych średnich, odchyłeń standardowych i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach. Zaznaczenie pola **Oceny wielkości efektu** powoduje wyświetlenie częściowej wartości eta kwadrat dla każdego efektu i każdej oceny parametru. Eta kwadrat jest statystyką opisującą część całkowitej zmienności, którą można przypisać czynnikowi. Zaznaczenie pola **Obserwowana siła** powoduje wyświetlenie siły testu po ustaleniu hipotezy alternatywnej na podstawie obserwowanej wartości. Zaznaczenie pola **Oceny parametrów** powoduje wyświetlenie ocen parametrów, błędów standardowych, testów *t*, przedziałów ufności i obserwowanej siły każdego testu. Zaznaczenie pola **Macierz współczynników kontrastów** powoduje otrzymanie macierzy **L**.

Testy jednorodności realizują testy Levene'a jednorodności wariancji dla każdej zmiennej zależnej i każdej kombinacji poziomów czynników międzyobiektowych albo tylko dla czynników międzyobiektowych. Do sprawdzenia założeń dotyczących danych wykorzystać można opcję Wykresy rozrzut-poziom i Wykres reszt. W przypadku braku czynników, opcja ta jest niedostępna. Zaznaczenie opcji **Wykres reszt** powoduje utworzenie wykresu reszt rzeczywistych-szacowanych-standaryzowanych

dla każdej zmiennej zależnej. Wykresy te są przydatne podczas sprawdzania założeń równej wariancji. Zaznaczenie pola **Brak dopasowania** powoduje sprawdzenie, czy model nadaje się do prawidłowego opisu relacji zachodzącej między zmienną zależną a zmiennymi niezależnymi. **Ogólne funkcje estymowalne** umożliwiają konstruowanie testów niestandardowych hipotez na podstawie ogólnych funkcji estymowalnych. Wiersze w każdej macierzy współczynników kontrastu są liniowymi kombinacjami ogólnych funkcji estymowalnych.

Dostępne są **testy heteroskedastyczności** służące do sprawdzania, czy wariancja błędów (dla każdej zmiennej zależnej) zależy od wartości zmiennych niezależnych. W przypadku **testu Breuscha-Pagana**, **zmodyfikowanego testu Breuscha-Pagana** i **testu F** można określić model, na którym ma bazować test. Domyślnie model składa się ze składnika stałego, składnika liniowego w wartościach przewidywanych, składnika kwadratowego w wartościach przewidywanych i składnika błędu.

Oszacowania parametrów z odpornymi błędami standardowymi generują tabelę oszacowań parametrów oraz błędy standardowe odporne bądź o spójnej heteroskedastyczności, a także statystyki t , wartości istotności oraz przedziały ufności wyznaczone przy użyciu odpornych błędów standardowych. Dostępnych jest pięć różnych metod estymacji odpornej macierzy kowariancji.

HC0

Oparta na oryginalnym asymptotycznym lub bazującym na dużej próbie, odpornym, empirycznym lub „kanapkowym” estymatorze macierzy kowariancji oszacowań parametrów. Środkowa część „kanapki” zawiera reszty kwadratowe wyznaczone zwykłą metodą najmniejszych kwadratów (OLS) lub metodą ważonych najmniejszych kwadratów (WLS).

HC1

Modyfikacja HC0 z próbą skończoną, przemnożoną przez $N/(N-p)$, gdzie N jest wielkością próby, a p jest liczbą nienadmiarowych parametrów w modelu.

HC2

Modyfikacja HC0, w której reszty kwadratowe dzielone są przez $1-h$, gdzie h jest wartością wpływu obserwacji.

HC3

Modyfikacja HC0 będąca przybliżeniem estymatora scyzorykowego (jackknife). Reszty kwadratowe dzielone są przez kwadrat $1-h$.

HC4

Modyfikacja HC0, w której reszty kwadratowe są dzielone przez $1-h$ do potęgi, której wykładnik zależy od h , N i p , ale nie może być wyższy niż 4.

Poziom istotności. Może zająć potrzeba skorygowania poziomu istotności używanego w testach post hoc oraz poziomu ufności wykorzystywanego do określania przedziałów ufności. Określona wartość jest również używana do obliczenia obserwowanej siły testu. Podczas określania poziomu istotności powiązany z nim poziom przedziałów ufności wyświetlany jest w oknie dialogowym.

Pomocniczy model regresji OML

W oknie dialogowym Pomocniczy model regresji określa się model służący do testowania heteroskedastyczności.

Użyj wartości przewidywanych

Powoduje użycie modelu składającego się ze składnika stałego, składnika liniowego w wartościach przewidywanych, składnika kwadratowego w wartościach przewidywanych i składnika błędu.

Użyj modelu jednej zmiennej

Powoduje użycie modelu określonego w podrzędnym oknie dialogowym Model. Jeśli wskazany model nie zawiera składnika stałej, to jest on dodawany.

Model użytkownika

Powoduje użycie modelu wskazanego wprost przez użytkownika.

Buduj składniki

Tej opcji należy użyć, chcąc uwzględnić niezagnieżdżone składniki konkretnego typu (takie jak efekty główne) dla wszystkich kombinacji wybranego zestawu czynników i współzmiennych.

Buduj składniki użytkownika

Tej opcji należy użyć, chcąc uwzględnić składniki zagnieżdżone lub konstruować składnik jawnie, zmienna po zmiennej. Proces budowania składnika zagnieżdżonego obejmuje następujące kroki:

Dodatkowe właściwości komendy UNIANOVA

Język składni komend umożliwia również:

- Określanie zagnieżdżonych efektów w planie (za pomocą opcji DESIGN).
- Określanie testów efektów w odniesieniu do liniowej kombinacji efektów lub wartości (za pomocą opcji komendy TEST).
- Określanie wielu kontrastów (za pomocą opcji komendy CONTRAST).
- Uwzględnianie braków danych zdefiniowanych przez użytkownika (za pomocą opcji komendy MISSING).
- Określanie kryteriów EPS (za pomocą opcji komendy CRITERIA).
- Tworzenie macierzy **L**, **M** lub **K** użytkownika (za pomocą opcji komendy LMATRIX, MMATRIX i KMATRIX).
- Określanie pośredniej kategorii odniesienia dla kontrastów odchyłeń lub kontrastów prostych (za pomocą opcji komendy CONTRAST).
- Określanie metryk kontrastów wielomianowych (za pomocą opcji komendy CONTRAST).
- Określanie składników błędów dla porównań post hoc (za pomocą opcji komendy POSTHOC).
- Obliczanie szacowanych średnich brzegowych dla czynników bądź interakcji między czynnikami umieszczonymi na liście czynników (za pomocą opcji komendy EMMEANS).
- Określanie nazw zmiennych tymczasowych (za pomocą opcji komendy SAVE).
- Tworzenie pliku danych zawierającego macierz korelacji (za pomocą opcji komendy OUTFILE).
- Tworzenie pliku macierzowego zawierającego statystyki z międzyobiektywnej tabeli ANOVA (za pomocą opcji komendy OUTFILE).
- Zapisywanie macierzy planu w nowym pliku danych (za pomocą opcji komendy OUTFILE).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Korelacje parami

Procedura korelacji parami oblicza współczynnik korelacji Pearsona, wartość rho Spearmana i wartość tau-*b* Kendalla wraz z ich poziomami istotności. Korelacje mierzą zależności pomiędzy zmiennymi lub rangami. Przed przystąpieniem do obliczania współczynnika korelacji, należy przejrzeć dane w poszukiwaniu wartości odstających (które mogą powodować błędne wyniki) i dowodów zależności liniowej. Współczynnik korelacji Pearsona jest miarą powiązań liniowych. Dwie zmienne mogą być idealnie powiązane, lecz jeśli relacja nie jest liniowa, współczynnik korelacji Pearsona nie jest odpowiednią statystyką do pomiaru ich powiązania.

Ustawienia przedziału ufności są dostępne dla testów Pearsona i Spearmana.

Przykład

Czy liczba meczów wygranych przez drużynę koszykówki jest skorelowana ze średnią liczbą punktów zdobywanych w każdym meczu? Wykres rozrzutu wskazuje, że zależność jest liniowa. Analiza danych z rozgrywek NBA w sezonie 1994-1995 wykazała, że współczynnik korelacji Pearsona (0,581) jest istotny na poziomie 0,01. Można podejrzewać, że im więcej meczów wygranych w sezonie, tym mniej punktów zdobywają przeciwnicy. Zmienne te są skorelowane ujemnie (0,401), a korelacja jest istotna na poziomie 0,05.

Statistics

W przypadku każdej zmiennej: liczby obserwacji bez brakujących wartości, średnia oraz odchylenie standardowe. W przypadku każdej pary zmiennych: współczynnik korelacji Pearsona, rho Spearmana, tau-*b* Kendalla, iloczyn wektorowy odchyłeń i kowariancja.

Zagadnienia dotyczące danych

Dane

Do obliczenia współczynnika korelacji Pearsona należy wykorzystać symetryczne zmienne ilościowe, a do obliczenia rho Spearmana i tau-*b* Kendalla zmienne ilościowe lub zmienne z uporządkowanymi kategoriami.

Założenia

We współczynniku korelacji Pearsona zakłada się, że każda para zmiennych jest parami normalna.

Uzyskiwanie korelacji parami

Wybierz z menu następującą opcję:

Analiza > Korelacje > Parami...

1. Wybierz co najmniej dwie zmienne numeryczne.

Dostępne są także następujące opcje:

Współczynniki korelacji

Dla zmiennych ilościowych o rozkładzie normalnym należy wybrać współczynnik korelacji **Pearsona**. Jeśli dane nie mają rozkładu normalnego lub mają uporządkowane kategorie, to należy wybrać współczynniki **tau-b Kendalla** lub **Spearman**, które są miarą powiązań pomiędzy rangami. Współczynniki korelacji przyjmują wartości od -1 (idealna relacja ujemna) do +1 (idealna relacja dodatnia). Wartość równa 0 oznacza brak związku liniowego. Interpretując wyniki nie należy wyciągać wniosków przyczynowo-skutkowych ze względu na korelację istotną.

Test istotności

Można wybrać prawdopodobieństwa dwustronne lub jednostronne. Jeśli kierunek powiązania jest z góry znany, należy zaznaczyć opcję **Jednostronna**. W przeciwnym wypadku należy zaznaczyć opcję **Dwustronna**.

Oznacz korelacje istotne

Współczynniki korelacji istotne na poziomie 0,05 są oznaczane jedną gwiazdką, istotne na poziomie 0,01 są oznaczane dwoma gwiazdkami.

Pokaż tylko dolny trójkąt

Gdy ta opcja jest zaznaczona, w danych wyjściowych przedstawiony jest tylko dolny trójkąt macierzy korelacji. Gdy ta opcja nie jest zaznaczona, w danych wyjściowych przedstawiana jest cała macierz korelacji. To ustawienie umożliwia uzyskanie tabeli zgodnej z wytycznymi APA.

Pokaż przekątną

Po wybraniu tej opcji w wynikach prezentowany jest dolny trójkąt tabeli macierzy korelacji wraz z wartościami przekątnych. To ustawienie umożliwia uzyskanie tabeli zgodnej z wytycznymi APA.

2. Opcjonalnie można wybrać następujące opcje:

- Kliknij przycisk **Opcje...**, aby określić statystyki korelacji Pearsona i ustawienia braków danych.
- Kliknij przycisk **Styl**, aby określić warunki automatycznego zmieniania właściwości tabel przestawnych na podstawie określonych warunków.
- **Bootstrap** to metoda uzyskiwania mocnych oszacowań błędów standardowych i przedziałów ufności dla ocen, takich jak średnia, mediana, proporcja, iloraz szans, współczynnik korelacji albo współczynnik regresji.
- Kliknij przycisk **Przedział ufności...**, aby określić opcje szacowania przedziałów ufności.

Korelacje parami: Opcje

Statystyki

W przypadku korelacji Pearsona można wybrać jedną lub obie z następujących opcji:

Średnie i odchylenia standardowe

Wyświetlane dla każdej zmiennej. Pokazywana jest także liczba obserwacji bez braków danych. Braki danych są obsługiwane indywidualnie dla każdej zmiennej, bez względu na ustawienia dla braków danych.

Iloczynny wektorowe odchyłeń i kowariancji

Wyświetlane dla każdej pary zmiennych. Iloczyn wektorowy odchyłeń jest równy sumie iloczynów zmiennych skorygowanych przez odjęcie ich średnich. Jest to licznik współczynnika korelacji Pearsona. Kowariancja jest niestandardyzowaną miarą związku pomiędzy dwoma zmiennymi, równą iloczynowi wektorowego odchylenia podzielonemu przez $N-1$.

Braki danych

Można wybrać jedną z następujących opcji:

Wyłączanie obserwacji parami

Obserwacje z brakami danych dla jednej lub obu z pary zmiennych do współczynnika korelacji są wyłączone z analizy. Ponieważ każdy współczynnik oparty jest na wszystkich obserwacjach mających prawidłowe kody w tej szczególnej parze zmiennych, to w każdym obliczeniu wykorzystywana jest maksymalna ilość dostępnych informacji. Wynikiem tego może być zbiór współczynników opartych na różnych liczbach obserwacji.

Wyłączanie wszystkich obserwacji z brakami

Obserwacje z brakami danych dla dowolnej ze zmiennych są wyłączone ze wszystkich korelacji.

Korelacje parami: przedział ufności

Okno dialogowe Przedział ufności udostępnia opcje szacowania przedziałów ufności. Okno dialogowe jest dostępne, gdy w oknie dialogowym Korelacje parami wybrano opcję **Pearsona**, **Tau-b Kendalla** lub **Spearmana**.

Szacuje przedział ufności parametru korelacja parami

Kontroluje szacowanie przedziału ufności parametru korelacja parami. Po wybraniu wykonywane jest szacowanie przedziału ufności.

Przedział ufności (%)

Określa poziom ufności dla wszystkich utworzonych przedziałów ufności. Podaj wartość liczbową z zakresu od 0 do 100. Wartością domyślną jest 95.

Korelacja Pearsona

Ustawienie **Zastosuj korektę obciążenia** określa, czy wykonywana jest korekta obciążenia. Domyślnie ustawienie to nie jest wybrane, a korekta obciążenia nie jest wykonywana. Po wybraniu tego ustawienia wykonywana jest korekta obciążenia szacowanych przedziałów ufności. To ustawienie jest dostępne, gdy w oknie dialogowym Korelacje parami wybrano opcję **Pearsona**.

Korelacja Spearmana

To ustawienie jest dostępne, gdy w oknie dialogowym Korelacje parami wybrano opcję **Spearmana** i określono opcje szacowania wariancji korelacji Spearmana przy użyciu następujących metod:

- **Fieller, Hartley i Pearson**
- **Bonett i Wright**
- **Carusa i Clifffa**

Dodatkowe właściwości komend CORRELATIONS (KORELACJE) oraz NONPAR CORR (KORELACJE NIEPARAMETRYCZNE)

Język składni komend umożliwia również:

- Tworzenie macierzy korelacji dla korelacji Pearsona, która może być wykorzystana zamiast danych surowych w celu przeprowadzenia innych analiz, jak np. analiza czynnikowa (za pomocą opcji komendy MATRIX).

- Uzyskanie korelacji każdej zmiennej na liście z każdą zmienną na drugiej liście (przy użyciu słowa kluczowego WITH w opcji komendy VARIABLES).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Korelacje cząstkowe

Procedura Korelacje cząstkowe umożliwia wyliczenie współczynników korelacji cząstkowej, opisujących liniową relację między dwiema zmiennymi, przy uwzględnieniu wpływu jednej lub wielu dodatkowych zmiennych. Korelacje są miarami powiązania liniowego. Dwie zmienne mogą być idealnie powiązane, lecz jeśli relacja nie jest liniowa, współczynnik korelacji Pearsona nie jest odpowiednią statystyką do mierzenia stopnia ich powiązania.

Przykład

Czy występuje relacja między środkami na ochronę zdrowia a wskaźnikami zachorowań? Chociaż można oczekiwać istnienia zależności odwrotnie proporcjonalnej, z analiz wynika istnienie znaczącej korelacji *dodatniej*: gdy wzrastają środki na ochronę zdrowia, wzrasta wskaźnik zachorowań. Jeśli jednak skontroluje się wskaźnik wizyt w placówkach służby zdrowia, obserwowana korelacja dodatnia zostaje niemal wyeliminowana. Środki na ochronę zdrowia wydają się być dodatnio skorelowane ze wskaźnikiem zachorowań tylko dlatego, że wzrost tych środków umożliwia dostęp do opieki zdrowotnej większej liczbie osób, co z kolei powoduje wzrost liczby chorób rejestrowanych przez lekarzy i szpitale.

Statystyka

W przypadku każdej zmiennej: liczby obserwacji bez brakujących wartości, średnia oraz odchylenie standardowe. Macierze korelacji cząstkowych i korelacji rzędu zerowego, ze stopniami swobody oraz poziomami istotności.

Zagadnienia dotyczące danych

Dane

Należy używać symetrycznych zmiennych ilościowych.

Założenia

Procedura Korelacje cząstkowe zakłada, że wszystkie zmienne są parami normalne.

Uzyskiwanie korelacji cząstkowych

1. Wybierz z menu następujące opcje:

Analizuj > Korelacja > Częściowa...

2. Wybierz co najmniej dwie zmienne liczbowe, dla których mają być wyliczone korelacje cząstkowe.
3. Wybierz co najmniej jedną zmienną sterującą.

Dostępne są także następujące opcje:

Test istotności

Można wybrać prawdopodobieństwa dwustronne lub jednostronne. Jeśli kierunek powiązania jest z góry znany, należy zaznaczyć opcję **Jednostronna**. W przeciwnym wypadku należy zaznaczyć opcję **Dwustronna**.

Pokaż rzeczywisty poziom istotności

Domyślnie dla każdego współczynnika korelacji pokazane jest prawdopodobieństwo i stopnie swobody. Jeśli zaznaczenie tego pola wyboru zostanie usunięte, współczynniki o poziomie istotności 0,05 będą wyróżnione gwiazdkami, współczynniki istotne na poziomie 0,01 będą wyróżnione podwójnymi gwiazdkami, a stopnie swobody nie będą pokazywane. To ustawienie dotyczy zarówno macierzy korelacji cząstkowych jak i macierzy rzędu zerowego.

Korelacje cząstkowe: Opcje

Statystyki. Można wybrać jedną lub obie z następujących opcji:

- **Średnie i odchylenia standardowe.** Wyświetlane dla każdej zmiennej. Pokazywana jest także liczba obserwacji bez braków danych.
- **Korelacje rzędu zerowego.** Wyświetlana jest macierz prostych korelacji między wszystkimi zmiennymi, włącznie ze zmiennymi sterującymi.

Braki danych. Można wybrać jedną z poniższych alternatyw:

- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej (w tym także zmiennej sterującej) są wyłączone ze wszystkich obliczeń.
- **Wyłączanie obserwacji parami.** Przy obliczeniach korelacji rzędu zerowego, na których opierają się korelacje cząstkowe, obserwacje z brakami danych dla jednej lub obu par zmiennych nie są wykorzystywane. Usuwanie parami wykorzystuje wszystkie dane, których można użyć. Liczba obserwacji może jednak być różna dla różnych współczynników. Podczas usuwania parami liczba stopni swobody danego współczynnika cząstkowego zależy od najmniejszej liczby obserwacji użytej przy obliczaniu korelacji rzędu zerowego.

Dodatkowe właściwości komendy KORELACJE NIEPARAMETRYCZNE

Język składni komend umożliwia również:

- Odczytywanie macierzy korelacji rzędu zerowego lub zapisywanie macierzy korelacji cząstkowych (za pomocą opcji komendy MATRIX).
- Uzyskanie korelacji każdej zmiennej na liście z każdą zmienną na drugiej liście (przy użyciu słowa kluczowego WITH w opcji komendy VARIABLES).
- Uzyskiwanie wielu analiz (za pomocą wielu opcji komend VARIABLES).
- Określanie żądanych wartości porządkowych (na przykład korelacje cząstkowe pierwszej i drugiej kolejności) przy dwóch zmiennych sterujących (za pomocą opcji komend VARIABLES).
- Usuwanie zbędnych współczynników (za pomocą opcji komendy FORMAT).
- Wyświetlanie macierzy prostych korelacji, kiedy niektóre współczynniki nie mogą zostać obliczone (za pomocą opcji komendy STATISTICS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Odległości

Procedura ta umożliwia obliczenie każdej z szerokiej gamy statystyk służących do pomiaru podobieństwa lub niepodobieństwa (odległości) między parami zmiennych lub obserwacji. Te miary podobieństwa lub odległości mogą następnie zostać wykorzystane do innych procedur, takich jak analiza czynnikowa, analiza skupień, czy skalowanie wielowymiarowe, w celu ułatwienia analizy złożonych zbiorów danych.

Przykład. Czy możliwy jest pomiar podobieństw między parami samochodów w oparciu o pewne charakterystyki, takie jak pojemność silnika, zużycie paliwa i moc wyrażona w koniach mechanicznych? Obliczając podobieństwa między samochodami można wyrobić sobie pojęcie na temat tego, które z nich są do siebie podobne, a które się różnią. Aby przeprowadzić bardziej formalną analizę, można zastosować do podobieństw hierarchiczną analizę skupień lub skalowanie wielowymiarowe, aby odkryć ich strukturę.

Statystyki. Miary niepodobieństwa (odległości) dla danych przedziałowych to: odległość euklidesowa, kwadrat odległości euklidesowej, odległość Czebyszewa, odległość miejska, odległość Minkowskiego lub odległość użytkownika; dla danych będących liczebnościami: chi-kwadrat lub phi-kwadrat; dla danych binarnych: odległość euklidesowa, kwadrat odległości euklidesowej, różnica wielkości, różnica wzoru, wariancja, kształt lub miara Lance'a i Williama. Miary podobieństwa dla danych przedziałowych to: korelacja Pearsona lub cosinus; dla danych binarnych: miara Russela i Rao, proste zgodności, miara Jaccarda, miara Dice'a, miara Rogersa i Tanimoto, miara Sokala i Sneatha 1, miara Sokala i Sneatha 2, miara Sokala i Sneatha 3, miara Kulczyńskiego 1, miara Kulczyńskiego 2, miara Sokala i Sneatha 4, miara Hamanna, lambda, D Anderberga, Y Yule'a, Q Yule'a, miara Ochiai, miara Sokala i Sneatha 5, phi korelacja 4-punktowa lub rozproszenie.

Obliczanie macierzy odległości

1. Wybierz z menu następującą opcję:

Analiza > Korelacje > Odległości...

2. Wybierz co najmniej jedną zmienną numeryczną do wyliczenia odległości między obserwacjami lub co najmniej dwie zmienne numeryczne do wyliczenia odległości między zmiennymi.
3. W grupie Oblicz odległości wybierz opcję wyliczenia odległości pomiędzy obserwacjami lub pomiędzy zmiennymi.

Odległości: Miary niepodobieństwa

Z grupy Miara dla danych wybierz opcję odpowiadającą wybranemu typowi danych (interwałowe, liczebności lub binarne), a następnie z listy rozwijanej wybierz jedną miarę, która odpowiada temu typowi danych. Dostępne miary, według typu danych, to:

- **Dane przedziałowe.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Odległość Czebyszewa, Odległość miejska, Odległość Minkowskiego lub Odległość użytkownika.
- **Dane liczebnościowe.** Odległość chi-kwadrat lub Odległość phi-kwadrat.
- **Dane binarne.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Różnica wielkości, Różnica wzoru, Miara wariancyjna, Kształt lub Miara Lance'a i Williamsa (w polach Występuje i Nie występuje należy wprowadzić wartości określające, które dwie wartości są znaczące; wszystkie pozostałe wartości zostaną zignorowane).

Grupa Przekształcanie wartości umożliwia standaryzację wartości danych dla obserwacji lub dla zmiennych *przed* wyliczeniem odległości. Przekształcenia te nie mają zastosowania do danych binarnych. Dostępne metody standaryzacji to: wartości statystyki z, zakres od -1 do 1, zakres od 0 do 1, maksymalna wartość równa 1, średnia równa 1 lub odchylenie standardowe równe 1.

Grupa Transformacja miar umożliwia przekształcenie wartości generowanych przez miarę odległości. Są one stosowane po wyliczeniu miary odległości. Dostępne opcje to: wartości bezwzględne, zmiana znaku i przeskalowanie do zakresu od 0 do 1.

Odległości: Miary podobieństwa

Z grupy Miara dla danych wybierz opcję odpowiadającą wybranemu typowi danych (interwałowe lub binarne), a następnie z listy rozwijanej wybierz jedną miarę, która odpowiada temu typowi danych. Dostępne miary, według typu danych, to:

- **Dane przedziałowe.** Korelacja Pearsona lub cosinus.
- **Dane binarne.** Miara Russela i Rao, proste zgodności, miara Jaccarda, miara Dice'a, miara Rogersa i Tanimoto, miara Sokala i Sneatha 1, miara Sokala i Sneatha 2, miara Sokala i Sneatha 3, miara Kulczyńskiego 1, miara Kulczyńskiego 2, miara Sokala i Sneatha 4, miara Hamanna, lambda, *D* Anderberga, *Y* Yule'a, *Q* Yule'a, miara Ochiai, miara Sokala i Sneatha 5, phi korelacja 4-punktowa lub rozproszenie. (w polach Występuje i Nie występuje należy wprowadzić wartości określające, które dwie wartości są znaczące; wszystkie pozostałe wartości zostaną zignorowane).

Grupa Przekształcanie wartości umożliwia standaryzację wartości danych dla obserwacji lub dla zmiennych *przed* wyliczeniem odległości. Przekształcenia te nie mają zastosowania do danych binarnych. Dostępne metody standaryzacji to: wartości statystyki z, zakres od -1 do 1, zakres od 0 do 1, maksymalna wartość równa 1, średnia równa 1 i odchylenie standardowe równe 1.

Grupa Transformacja miar umożliwia przekształcenie wartości generowanych przez miarę odległości. Są one stosowane po wyliczeniu miary odległości. Dostępne opcje to: wartości bezwzględne, zmiana znaku i przeskalowanie do zakresu od 0 do 1.

Dodatkowe właściwości komendy PROXIMITIES

Procedura hierarchiczna skupień stosuje składnię komend PROXIMITIES. Język składni komend umożliwia również:

- Określ liczbę całkowitą jako potęgę dla miary odległości Minkowskiego.

- Określ liczby całkowite jako potęgę i pierwiastek dla definiowanej miary odległości.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Modele liniowe

Modele liniowe przewidują przewidywaną zmienną ilościową na podstawie liniowych relacji między przewidywaną a jednym lub większą liczbą predyktorów.

Modele liniowe są stosunkowo proste i zapewniają łatwy w interpretacji wzór matematyczny do oceny. W przeciwieństwie do innych typów modeli dla tego samego zbioru danych (takich jak sieci neuronowe czy drzewa decyzyjne), właściwości tych modeli łatwo zrozumieć i zwykle można się ich szybko nauczyć.

Przykład. Firma ubezpieczeniowa o ograniczonych środkach na sprawdzenie roszczeń ubezpieczeniowych właścicieli domów chce stworzyć model przybliżający koszty roszczeń. Po wdrożeniu tego modelu w centrach usługowych przedstawiciele będą mogli wprowadzać informacje na temat roszczeń podczas rozmów telefonicznych z klientem i natychmiast otrzymać „przybliżony” koszt roszczenia bazujący na wcześniejszych danych.

Wymagania dotyczące zmiennych. Musi istnieć zmienna przewidywana i co najmniej jedna wejściowa. Domyślnie zmienne ze wstępnie zdefiniowanymi rolami Obie lub Żadna nie są używane. Docelowa musi być zmienną ciągłą (ilościową). Nie ma żadnych ograniczeń poziomu pomiaru dla predyktorów (wejścia); zmienne jakościowe (nominalne oraz porządkowe) są używane jako czynniki w modelu, a zmienne ciągłe używane są jako współzmiennie.

Uwaga: Jeśli zmienna jakościowa zawiera więcej niż 1000 kategorii, procedura nie uruchamia się i nie jest tworzony żaden model.

Otrzymywanie modelu liniowego

Ta zmienna wymaga opcji Statistics Base.

Z menu wybierz:

Analiza > Regresja > Automatyczne modelowanie liniowe...

1. Upewnij się, że istnieje co najmniej jedna docelowa i jedno wejście.
2. Kliknij **Opcje budowania**, aby podać opcjonalne ustawienia budowania i modelu.
3. Kliknij **Opcje modelu**, aby zapisać wyniki w aktywnym zbiorze danych i wyeksportować ten model do pliku zewnętrznego.
4. Kliknij opcję **Uruchom**, aby uruchomić procedurę i utworzyć Obiektu modelu.

Cele

Jaki chcesz osiągnąć cel? Zaznacz odpowiedni cel.

- **Zbudować model standardowy.** Ta metoda tworzy pojedynczy model do przewidywania przy pomocy predyktorów. Ogólnie rzecz biorąc standardowe modele są łatwiejsze w interpretacji i można je szybciej ocenić w porównaniu ze wzmocnionymi, spakowanymi lub dużymi zestawami zbiorów danych.
- **Zwiększyć dokładność modelu (boosting).** Metoda ta tworzy model zespolony przy pomocy wzmocnienia, który generuje sekwencję modeli w celu uzyskania bardziej precyzyjnych predykcji. Tworzenie i ocena zestawów mogą trwać dłużej niż w przypadku standardowego modelu.

Wzmocnienie tworzy kolejność „modeli składników”, z których każdy został skompilowany na podstawie całego zbioru danych. Przed skompilowaniem każdego kolejnego modelu składników, rekordy są ważone na podstawie reszt po poprzednich modelach składników. Obserwacje o dużej wartości reszt dają stosunkowo wyższe wagi analizy tak, że kolejny model składników będzie się skupiał na dobrym przewidywaniu tych rekordów. Te modele składników tworzą razem model zespolony. Model zespolony ocenia nowe rekordy przy pomocy reguły łączenia; dostępne reguły zależą od poziomu pomiaru celu.

- **Wzmocnić stabilność modelu (agregacja bootstrapowa).** Metoda ta tworzy model zespolony przy pomocy spakowania (agregacja metodą bootstrap), które generuje wiele modeli w celu uzyskania

bardziej wiarygodnych predykcji. Tworzenie i ocena zestawów mogą trwać dłużej niż w przypadku standardowego modelu.

Agregacja metodą bootstrap (bagging) powiela zespół danych z przyuczenia, tworząc próbkowanie poprzez zastąpienie oryginalnego zbioru danych. W wyniku tego powstają próby bootstrap, które mają taki sam rozmiar, jak oryginalny zbiór danych. Następnie na podstawie każdego powielania kompilowany jest „model składników”. Te modele składników tworzą razem model zespolony. Model zespolony ocenia nowe rekordy przy pomocy reguły łączenia; dostępne reguły zależą od poziomu pomiaru celu.

- **Utworzyć model dla bardzo dużych zbiorów danych (wymaga produktu IBM SPSS Statistics Server).** Metoda ta tworzy model zespolony przez podział zbioru danych na oddzielne bloki danych. Wybierz tę opcję, jeśli Twój zbiór danych jest zbyt duży do utworzenia któregośkolwiek z powyższych modeli, lub aby utworzyć model przyrostowy. Tworzenie tej opcji może być szybsze, ale ocena może potrwać dłużej niż w przypadku standardowego modelu. Ta opcja wymaga IBM SPSS Statistics połączenia z serwerem.

Patrz temat „Zestawy” na stronie 142 w celu zapoznania się z ustawieniami związanymi ze wzmocnieniem, agregacją i bardzo dużymi zbiorami danych ze wzmocnieniem o agregacją.

Podstawy

Automatycznie przygotuj dane. Opcja ta umożliwia przekształcenie docelowej i predyktorów przez tą procedurę w celu maksymalizacji siły predykcji modelu. Wszystkie transformacje są zapisywane razem z modelem i zastosowane dla nowych danych do oceny. Oryginalne wersje przekształconych zmiennych są wyłączone z modelu. Domyślnie odbywa się następujące, automatyczne przygotowanie danych.

- **Obsługa daty i czasu.** Każdy predyktor daty jest przekształcany na nowy predyktor ciągły zawierający czas, który upłynął od daty odniesienia (1970-01-01). Każdy predyktor czasu jest przekształcany na nowy predyktor ciągły, zawierający czas, który upłynął od godziny odniesienia (00:00:00).
- **Korekta poziomu pomiaru.** Predyktory ciągłe zawierające mniej niż 5 odrębnych wartości są uznane za predyktory porządkowe. Predyktory porządkowe zawierające więcej niż 10 odrębnych wartości są uznane za predyktory ciągłe.
- **Obsługa wartości odstających.** Wartości predyktorów ciągłych znajdujące się poza wartością odcięcia (3 standardowe odchylenia od średniej) są ustawione na wartości odcięcia.
- **Traktowanie braków danych.** Brakujące wartości nominalnych predyktorów są zastępowane trybem podziału szkoleniowego. Brakujące wartości porządkowych predyktorów są zastępowane medianą podziału szkoleniowego. Brakujące wartości predyktorów ciągłych są zastępowane średnią podziału szkoleniowego.
- **Nadzorowane scalanie.** Model staje się skromniejszy poprzez zmniejszenie liczby zmiennych do przetworzenia w powiązaniu z docelową. Podobne kategorie identyfikuje się na podstawie relacji między wejściem a zmienną przewidywaną. Scalane są kategorie, które znacząco się nie różnią (to znaczy takie, których wartość p jest większa niż 0,1). Jeśli wszystkie kategorie są scalone w jedną, oryginalne i wyliczone wersje zmiennej są wyłączone z modelu, ponieważ nie mają żadnej wartości jako predyktora.

Poziom ufności. Jest to poziom ufności używany do wyliczania oszacowań przedziałów współczynników modelu w widoku Współczynniki. Należy podać wartość większą od 0 i mniejszą od 100. Domyślną wartością jest 95.

Wybór modelu

Metoda selekcji modelu. Wybierz jedną z metod wyboru modelu (szczegóły znajdują się poniżej) lub **Uwzględnij wszystkie predyktory**, co powoduje wprowadzenie wszystkich dostępnych predyktorów jako składniki modelu efektów głównych. Domyślnie używana jest opcja **Krokowa postępująca**.

Wybór metody krokowej postępującej. Działanie to rozpoczyna się bez efektów w modelu i dodaje oraz usuwa każdorazowo po jednym efekcie do momentu, aż nie można już nic dodać ani usunąć żadnego efektu zgodnie z kryterium krokowej.

- **Kryteria wprowadzania/usuwania.** Jest to statystyka używana do określenia tego, czy należy dodać lub usunąć efekt z modelu. **Kryterium informacyjne (AICC)** bazuje na prawdopodobieństwie zestawu

uczenia przy założeniu, że model jest przystosowany do personalizacji nadmiernie złożonych modeli. Kryterium **Statystyki F** bazuje na teście statystycznym poprawy błędu modelu. **Poprawione R-kwadrat** bazuje na dopasowaniu zestawu uczenia i jest przystosowany do personalizacji nadmiernie złożonych modeli. **Kryterium zabezpieczające przed przeuczeniem (ASE)** bazuje na dopasowaniu (średniego kwadratu błędu lub ASE) zbioru zabezpieczającego przed przeuczeniem. Zbiór zabezpieczający przed przeuczeniem jest losową podpróbą około 30% oryginalnego zbioru danych, który nie został użyty do uczenia modelu.

W przypadku wybrania kryterium innego niż **Statystyki F**, na każdym kroku do modelu dodawany jest efekt, który odpowiada największemu, dodatniemu wzrostowi kryterium. Jakikolwiek efekty w tym modelu, które odpowiadają spadkowi kryterium są usuwane.

W przypadku wybrania **Statystyki F** jako kryterium, na każdym kroku do modelu dodawany jest efekt, który ma najniższą wartość p , mniejszą niż określony próg, **Uwzględnij efekty z wartościami p mniejszymi niż**. Domyślną wartością jest 0,05. Jakikolwiek efekty w modelu o wartości p większej niż określony próg, **Usuń efekty z wartościami p większymi od** są usuwane. Domyślną wartością jest 0,10.

- **Maksimum efektów w modelu ostatecznym.** Domyślnie wszystkie efekty można wprowadzić do modelu. Alternatywnie, jeśli algorytm krokowy zakończy krok z określoną, maksymalną liczbą efektów, algorytm zatrzymuje się z bieżącym zestawem efektów.
- **Określ maksymalną liczbę kroków.** Algorytm krokowy zatrzymuje się po wykonaniu określonej liczby kroków. Domyślnie jest to 3-krotność liczby dostępnych efektów. Alternatywnie podaj dodatnią liczbę całkowitą w maksymalnej liczbie kroków.

Wybór metody najlepszych podzbiorów Opcja ta zaznacza „wszystkie możliwe” modele lub co najmniej większy podzbiór możliwych modeli, niż krokowa postępująca, w celu wybrania najlepszego, zgodnie z kryterium najlepszego podzbioru. **Information Criterion (AICC)** bazuje na prawdopodobieństwie zestawu uczenia przy założeniu, że model jest przystosowany do personalizacji nadmiernie złożonych modeli. **Poprawione R-kwadrat** bazuje na dopasowaniu zestawu uczenia i jest przystosowany do personalizacji nadmiernie złożonych modeli. **Kryterium zabezpieczające przed przeuczeniem (ASE)** bazuje na dopasowaniu (średniego kwadratu błędu lub ASE) zbioru zabezpieczającego przed przeuczeniem. Zbiór zabezpieczający przed przeuczeniem jest losową podpróbą około 30% oryginalnego zbioru danych, który nie został użyty do uczenia modelu.

Model o największej wartości kryterium jest wybierany jako najlepszy model.

Uwaga: Wybór najlepszych podzbiorów może wymagać większej liczby obliczeń niż wybór krokowy, postępujący. Jeśli zadanie najlepszych podzbiorów jest wykonywane w połączeniu z wzmocnieniem, agregacją metodą bootstrap lub bardzo dużymi zbiorami danych, tworzenie modelu z wykorzystaniem wyboru krokowego, postępującego może zająć dużo więcej czasu, niż tworzenie standardowego modelu.

Zestawy

Ustawienia te determinują zachowanie tworzenia zespołów, które występuje, gdy w Celach pożądane jest wspomaganie, agregacja metodą bootstrap lub bardzo duże zbiory danych. Opcje, które nie mają zastosowania do wybranego celu są ignorowane.

Agregacja metodą bootstrap i bardzo duże zbiory danych. Podczas wybierania zestawu jest to reguła służąca do łączenia przewidywanych wartości z modeli podstawowych w celu wyliczenia wartości oceny zestawu.

- **Domyślna reguła zespolenia dla docelowych wartości ilościowych.** Przewidywane wartości zestawu dla jakościowych zmiennych docelowych można połączyć przy pomocy średniej lub mediany przewidywanych wartości z modeli podstawowych.

Należy zwrócić uwagę, że gdy celem jest zwiększenie dokładności modelu, wybory reguły łączenia są ignorowane. Wzmocnienie zawsze wykorzystuje głos ważonej większości do oceny jakościowych zmiennych docelowych i ważonej mediany do oceny jakościowych zmiennych docelowych.

Boosting i agregacja bootstrapowa. Podaj liczbę modeli podstawowych do utworzenia, gdy celem jest zwiększenie dokładności lub stabilności modelu; dla agregacji metodą bootstrap jest to liczba prób agregacji metodą bootstrap. Powinna to być dodatnia liczba całkowita.

Zaawansowane

Replikacja wyników. Ustawienie wartości początkowej generatora liczb losowych umożliwia powielenie analizy. Generator liczb pseudolosowych służy do wyboru rekordów, które znajdują się w zbiorze zabezpieczającym przed przeuczeniem. Podaj liczbę całkowitą lub kliknij przycisk **Generuj**, co spowoduje utworzenie pseudolosowej liczby całkowitej między 1 a 2147483647, włącznie. Domyślną wartością jest 54752075.

Opcje modelu

Zapisz wartości przewidywane w zbiorze danych. Domyślna nazwa zmiennej to *PredictedValue*.

Eksportuj model. Powoduje to powoduje zapisanie modelu w zewnętrznym pliku .zip . Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. Podaj niepowtarzalną prawidłową nazwę pliku. Jeśli specyfikacja pliku odpowiada istniejącemu plikowi, zostanie on nadpisany.

Podsumowanie modelu

Widok Podsumowanie modelu to szybkie podsumowanie modelu i jego dopasowania.

Tabela. Tabela określa niektóre ustawienia modelu wysokiego poziomu, łącznie z:

- nazwa elementu docelowego określona na zakładce Zmienne,
- czy zostało wykonane automatyczne przygotowanie danych zgodnie z ustawieniem w obszarze Podstawowe,
- metoda selekcji modelu oraz kryteria określone w ustawieniach Wybór modelu. Wyświetlana jest także wartość wyboru kryterium dla modelu finalnego, przedstawiona w mniejszym i lepszym formacie.

Wykres. Na wykresie przedstawiono dokładność modelu finalnego, który jest przedstawiony w większym i lepszym formacie. Wartość wynosi 100 x skorygowana R^2 dla modelu finalnego.

Automatyczne przygotowanie danych

Widok ten przedstawia informacje o tym, które zmienne zostały wyłączone i w jaki sposób przekształcone zmienne zostały uwzględnione w kroku automatycznego przygotowania danych (ADP). Dla każdej zmiennej, która została przekształcona lub wyłączona, tabela zawiera nazwę zmiennej, jej rolę w analizie i działanie podjęte przez krok ADP. Zmienne są posortowane alfabetycznie, w kolejności rosnącej, według nazw zmiennych. Działania, które można podjąć dla każdego z pól to:

- **Czas trwania wyliczenia: miesiące** wylicza czas (w miesiącach), który upłynął od wartości pola zawierającego daty do bieżącej daty systemowej.
- **Czas trwania wyliczenia: godziny** wylicza czas (w godzinach), który upłynął od wartości pola zawierającego daty do bieżącego czasu systemu.
- **Zmień poziom pomiaru z ilościowego na porządkowy** konwertuje zmienne ciągłe zawierające mniej niż 5 różnych wartości na zmienne porządkowe.
- **Zmień poziom pomiaru z porządkowego na ciągły** konwertuje zmienne porządkowe zawierające więcej niż 10 różnych wartości na zmienne ciągłe.
- **Obcięcie wartości odstających** ustawia wartości predyktorów ciągłych znajdujące się poza wartością odcięcia (3 standardowe odchylenia od średniej) na wartość odcięcia.
- **Zastąp braki danych** zastępuje brakujące wartości zmiennych nominalnych trybem, zmiennych porządkowych — medianą, a zmiennych ilościowych — średnią.

- **Łączenie kategorii w celu maksymalizacji związku ze zmienną przewidywaną** identyfikuje „podobne” kategorie predyktorów bazujące na relacji między wejściem a docelową. Scalane są kategorie, które znacząco się nie różnią (to znaczy takie, których wartość p jest większa niż 0,05).
- **Wyklucz predyktor o stałych wartościach / po obsłudze wartości odstających / po scaleniu kategorii** usuwa predyktory o pojedynczej wartości po (możliwym) wykonaniu innych działań ADP.

Ważność predyktorów

Zazwyczaj działania modelujące mają koncentrować się na zmiennych predyktorów, które są najważniejsze, a opuszczane lub ignorowane mają być te zmienne, które są najmniej ważne. Wykres ważności predyktorów pomaga osiągnąć ten cel przez wskazanie względnej ważności każdego predyktora przy szacowaniu modelu. Ponieważ wartości są względne, suma wartości wszystkich wyświetlanych predyktorów wynosi 1,0. Ważność predyktora nie jest powiązana z dokładnością modelu. Jest powiązana z ważnością każdego predyktora przy prognozach, a nie z tym, czy taka prognoza jest dokładna.

Przewidywane przez Obserwowane

Przedstawia on wykres rozrzutu z kategoryzacją przewidywanych wartości na osi pionowej przez obserwowane wartości na osi poziomej. W idealnym przypadku punkty te powinny leżeć na prostej nachylonej pod kątem 45 stopni; widok ten może stwierdzić, czy którekolwiek z wyników zostały przewidziane przez model w sposób oczywisty.

Reszty

Widok ten przedstawia wykres diagnostyczny reszt modelu.

Style wykresu. Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Histogram.** Jest to podzielony histogram studentyzowanych reszt z nakładaniem normalnego rozkładu. Modele liniowe zakładają, że reszty mają normalny rozkład tak, że histogram w idealnych warunkach powinien znajdować się maksymalnie blisko gładkiej linii.
- **Wykres P-P.** Jest to podzielony wykres prawdopodobieństwo-prawdopodobieństwo porównujący studentyzowane reszty z rozkładem normalnym. Jeśli nachylenie naniesionych punktów jest mniej strome niż normalna linia, reszta będzie bardziej różnorodna niż normalny rozkład; jeśli nachylenie będzie bardziej strome, reszta będzie mniej różnorodna niż normalny rozkład. Jeśli naniesione punkty mają krzywą w kształcie litery S, wówczas rozkład reszt jest skośny.

Odstające

Tabela ta zestawia rekordy, które wywierają nadmierny wpływ na model i przedstawia identyfikator rekordu (jeśli został podany w zakładce Zmienne), wartość docelową i odległość Cooka. Odległość Cooka jest miarą stopnia, w jakim zmieniłyby się reszty dla wszystkich rekordów, przy wykluczeniu poszczególnych rekordów z obliczeń współczynników modelu. Duża odległość Cooka wskazuje, że wykluczenie z rekordu znacząco zmienia współczynnik i dlatego powinno być uznawane za wpływowe.

Wpływowe rekordy należy uważnie zbadać w celu określenia, czy można im nadać mniejszą wagę podczas oceny modelu lub obciąć wartości odstające do jakiegoś dopuszczalnego progu, lub całkowicie usunąć wpływowe rekordy.

Efekty

Widok ten przedstawia rozmiar każdego efektu w modelu.

Style. Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Diagram.** Jest to wykres z efektami posortowanymi od góry do dołu wg malejącej ważności predyktora. Linie łączące w diagramie są wazone na podstawie istotności efektu, gdzie większa szerokość linii odpowiada bardziej istotnym efektom (niższe wartości p). Umieszczenie kursora nad linią łączącą powoduje wyświetlenie podpowiedzi wskazującej wartość p oraz ważność efektu. Jest to wartość domyślna.

- **Tabela.** Jest to tabela ANOVA dla ogólnego całkowitych i pojedynczych efektów modelu. Pojedyncze efekty są posortowane od góry do dołu ze zmniejszającą się ważnością predyktora. Weź pod uwagę, że domyślnie tabela jest zwinięta i ukazuje tylko wyniki modelu ogólnego. Aby zobaczyć wyniki dla pojedynczego efektu modelu, kliknij w tabeli komórkę **Model skorygowany**.

Ważność predyktora. Dostępny jest suwak ważności predyktora, który steruje widocznością predyktorów w widoku. Nie zmienia to modelu, ale pozwala po prostu na skoncentrowaniu się na najistotniejszych predyktorach Domyślnie wyświetlanych jest 10 najistotniejszych efektów.

Istotność. Dostępny jest suwak istotności, który dalej steruje widocznością efektów w widoku, poza efektami pokazanymi na podstawie ważności predyktora. Efekty o wartościach istotności większych niż wartości suwaka, pozostają ukryte. Nie zmienia to modelu, ale pozwala po prostu na skoncentrowaniu się na najistotniejszych efektach Domyślną wartością jest 1,00 tak, że na podstawie istotności żadne efekty nie są filtrowane.

Współczynniki

Widok ten przedstawia wartość każdego współczynnika w modelu. Należy zwrócić uwagę, że czynniki (predyktory jakościowe) są kodowane wskaźnikami w ramach modelu tak, że **efekty** zawierające czynniki będą miały generalnie wiele powiązanych **współczynników**; po jednym dla każdej kategorii z wyjątkiem kategorii odpowiadającej parametrowi nadmiarowemu (odniesienia).

Style. Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Diagram.** Jest to wykres przedstawiający najpierw wyraz wolny, następnie sortujący efekty z góry do dołu wg malejącej ważności predyktora. W efektach zawierających czynniki, współczynniki są posortowane rosnąco według wartości danych. Linie łączące w diagramie są pokolorowane na podstawie znaku współczynnika (patrz klucz diagramu) i i ważone na podstawie istotności współczynnika, gdzie większa szerokość linii odpowiada bardziej istotnym współczynnikiem (niższe wartości p). Umieszczenie kursora nad linią łączącą powoduje wyświetlenie podpowiedzi wskazującej wartość współczynnika, jego wartości p oraz ważność efektu, z którym parametr jest powiązany. Jest to domyślny styl.
- **Tabela.** Pokazuje ona wartości, testy istotności i przedziały ufności dla poszczególnych współczynników modelu. Po wolnym wyrazie, efekty są posortowane od góry do dołu ze zmniejszającą się ważnością predyktora. W efektach zawierających czynniki, współczynniki są posortowane rosnąco według wartości danych. Zwróć uwagę, że domyślnie tabela jest zwinięta, aby pokazywać tylko współczynnik, istotność i ważność każdego z parametrów modelu. Aby zobaczyć błąd standardowy, statyczne t i przedział ufności, kliknij w tabeli komórkę **Współczynnik**. Umieszczenie kursora nad nazwą parametru modelu znajdującego się w tabeli powoduje wyświetlenie podpowiedzi wskazującej nazwę parametru, efektu powiązanego z parametrem oraz, dla predyktorów jakościowych, wartości etykiet związanych z parametrem modelu. Może to być szczególnie pomocne przy sprawdzaniu nowych kategorii stworzonych w czasie, gdy automatyczne przygotowanie danych scala podobne kategorie predyktora jakościowego.

Ważność predyktora. Dostępny jest suwak ważności predyktora, który steruje widocznością predyktorów w widoku. Nie zmienia to modelu, ale pozwala po prostu na skoncentrowaniu się na najistotniejszych predyktorach Domyślnie wyświetlanych jest 10 najistotniejszych efektów.

Istotność. Dostępny jest suwak istotności, który dalej steruje widocznością współczynników w widoku, poza współczynnikami pokazanymi na podstawie ważności predyktora. Współczynniki o wartościach istotności większych niż wartości suwaka, pozostają ukryte. Nie zmienia to modelu, ale pozwala po prostu na skoncentrowaniu się na najistotniejszych współczynnikach. Domyślną wartością jest 1,00 tak, że na podstawie istotności żadne współczynniki nie są filtrowane.

Oszacowanie średnie

Są to wykresy wyświetlane dla predyktorów istotności. Wykres ten przedstawia na osi pionowej wartości docelowej, szacowane z modelu, dla każdej wartości predyktora na osi poziomej, utrzymując wszystkie inne predyktory na stałym poziomie. Zapewnia on przydatną wizualizację efektów współczynników docelowej każdego predyktora.

Uwaga: jeśli żaden predyktor nie jest istotny, nie powstaje żadna oszacowana średnia.

Podsumowanie tworzenia modelu

Gdy w Ustawieniach wyboru modelu wybierze się algorytm wyboru modelu inny niż **Brak**, spowoduje to dostarczenie niektórych szczegółów procesu tworzenia modelu.

Krokowa postępująca. Gdy algorytmem wyboru jest krokowa postępująca, tabela przedstawia 10 ostatnich kroków w algorytmie krokowym. Pokazana jest wartość kryterium wyboru i efekty w modelu na tym kroku. Dzięki temu użytkownik ma świadomość w jakim stopniu każdy krok wpływa na model. Każda kolumna pozwala posortowanie wierszy tak, aby można było łatwo zobaczyć, które efekty znajdują się w modelu na danym kroku.

Najlepsze podzbiory. Gdy algorytmem wyboru są najlepsze podzbiory, tabela przedstawia 10 najlepszych modeli. Dla każdego modelu pokazana jest wartość kryterium wyboru i efekty w modelu. Obrazuje to stabilność najlepszych modeli; czy mają tendencję do posiadania wielu podobnych efektów o niewielkich różnicach, wówczas użytkownik może mieć względną pewność w „najlepszym” modelu; jeśli mają one tendencję do posiadania bardzo różnych efektów, wówczas niektóre efekty mogą być zbyt proste i powinny być połączone (lub jeden usunięty). Każda kolumna pozwala posortowanie wierszy tak, aby można było łatwo zobaczyć, które efekty znajdują się w modelu na danym kroku.

Regresja liniowa

Regresja liniowa służy oszacowaniu współczynników równania liniowego wymagającego co najmniej jednej zmiennej niezależnej, które najlepiej przybliży wartość zmiennej zależnej. Na przykład na podstawie zmiennych niezależnych, takich jak wiek, wykształcenie, lata pracy można próbować prognozować wartość rocznej sprzedaży przez danego sprzedawcę (zmienną zależną).

Przykład. Czy liczba meczy wygranych w ciągu sezonu przez drużynę koszykarską jest związana ze średnią liczbą punktów zdobywanych przez tę drużynę w jednym meczu? Wykresy rozrzutu wykazują, że te zmienne są ze sobą liniowo związane. Podobnie, liniowo związane są: liczba wygranych meczy oraz liczba punktów zdobywanych przez przeciwnika. Te zmienne posiadają relację ujemną. Jeśli liczba wygranych meczy rośnie, średnia liczba punktów zdobytych przez przeciwnika maleje. Dzięki regresji liniowej można stworzyć model relacji pomiędzy tymi zmiennymi. Dobry model pozwoli przewidzieć ile meczy zostanie wygranych przez poszczególne drużyny.

Statystyki. W przypadku każdej zmiennej: liczba ważnych obserwacji, średnia i odchylenie standardowe. Dla każdego modelu: współczynniki regresji, macierz korelacji, korelacje semicząstkowe i cząstkowe, wielokrotne R , R^2 , skorygowane R^2 , zmiana w R^2 , standardowy błąd oszacowania, tabela analizy wariancji powtarzanych pomiarów, tabela analizy wariancji, wartości przewidywane oraz reszty. Ponadto dla każdego współczynnika regresji przedziały ufności 95%, macierz wariancji i kowariancji, czynnik nadmiaru wariancji, tolerancja, test Durбина-Watsona, miary odległości (Mahalanobisa, Cooka oraz wartości wpływu), DfBeta, DfFit, przedziały predykcji oraz informacje diagnostyczne obserwacji. Wykresy: wykresy rozproszenia, wykresy cząstkowe, histogramy oraz normalne wykresy prawdopodobieństwa.

Wymagania dotyczące danych w regresji liniowej

Dane. Zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Zmienne kategoryjne, jak na przykład religia, główny przedmiot studiów, miejsce zamieszkania muszą być zakodowane w zmiennych binarnych (sztucznych) lub innych zmiennych kontrastowych.

Założenia. Dla każdej wartości zmiennej niezależnej rozkład zmiennej zależnej musi być normalny. Wariancja rozkładu zmiennej zależnej powinna być stała dla wszystkich wartości zmiennej niezależnej. Relacja między zmienną zależną a każdą zmienną niezależną powinna być liniowa, a wszystkie obserwacje powinny być niezależne.

Wykonywanie analizy regresji liniowej

1. Wybierz z menu następującą opcję:

Analiza > Regresja > Liniowe...

2. W oknie dialogowym Regresja liniowa wybierz numeryczną zmienną zależną.

3. Wybierz co najmniej jedną numeryczną zmienną niezależną.

Opcjonalnie można wykonać następujące czynności:

- Pogrupuj zmienne niezależne w bloki oraz określ metody wprowadzania dla poszczególnych podzbiorów zmiennych.
- Wybierz zmienną filtrującą, w celu ograniczenia analizy do podzbioru obserwacji zawierających określone wartości tej zmiennej.
- Wybierz zmienną opisu obserwacji w celu identyfikacji punktów na wykresach.
- Wybierz zmienną numeryczną WNK Waga na potrzeby analizy metodą ważonych najmniejszych kwadratów.

WNK. Pozwala na uzyskanie modelu ważonych najmniejszych kwadratów. Wartości analizowanych zmiennych są tu ważne przez odwrotność ich wariancji. Oznacza to, że obserwacje o dużej wariancji będą miały mniejszy wpływ na wyniki analizy niż obserwacje powiązane z małą wariancją. Jeśli wartość zmiennej ważącej wynosi zero, jest ujemna lub jest brakującą wartością to dana obserwacja zostaje wyłączona z analizy.

Regresja liniowa: Metody wyboru zmiennych

Wybór metody pozwala na określenie w jaki sposób zmienne niezależne będą wprowadzane do analizy. Korzystając z różnych metod, dla jednego zbioru zmiennych można skonstruować wiele modeli regresji.

- *Wprowadzanie (regresja)*. Procedura doboru zmiennych, w której wszystkie zmienne z bloku są jednocześnie wprowadzane do analizy.
- *Stepwise*. W każdym kroku zmienna niezależna nie znajduje się w równaniu, które ma najmniejsze prawdopodobieństwo F, jeśli takie prawdopodobieństwo jest wystarczająco małe. Zmienne uwzględnione już w równaniu regresji zostają z niego usunięte, jeśli związane z nimi prawdopodobieństwo F staje się dostatecznie duże. Procedura kończy się, kiedy nie da się wykluczyć ani dołączyć żadnej zmiennej.
- *Usuń*. Procedura doboru zmiennych, w której wszystkie zmienne z bloku są jednocześnie usuwane z analizy.
- *Eliminacja wsteczna*. Procedura wyboru zmiennej, w której wszystkie zmienne są wprowadzane do równania, a następnie usuwane sekwencyjnie. Zmienna o najmniejszej korelacji cząstkowej ze zmienną zależną jest brana pod uwagę do usunięcia w pierwszej kolejności. Jeśli spełnia kryteria eliminacji, zostaje usunięta. Po usunięciu pierwszej zmiennej, kolejną braną pod uwagę do usunięcia jest ta zmienna pozostająca w równaniu, która ma najmniejszą korelację cząstkową ze zmienną zależną. Procedura kończy działanie, gdy w równaniu nie występują inne zmienne spełniające kryteria usunięcia.
- *Selekcja postępująca*. Procedura wyboru zmiennej krokowej, w której zmienne są sekwencyjnie wprowadzane do modelu. Jako pierwsza rozważana jest ta zmienna, która jest najsilniej skorelowana ze zmienną zależną. Jest ona wprowadzana do modelu tylko wtedy, gdy spełnia kryterium wprowadzenia. Po wprowadzeniu pierwszej zmiennej pod uwagę brana jest ta zmienna nie wprowadzona do równania, która ma największą wartość współczynnika korelacji cząstkowej ze zmienną zależną. Procedura kończy działanie, gdy nie ma już żadnych zmiennych spełniających kryterium wprowadzenia.

Wartości istotności wyników zależą od dopasowania pojedynczego modelu. Z tego względu, jeśli stosowana jest metoda krokowa (krokowa, eliminacji wstecznej lub selekcji postępującej) wartości istotności są zazwyczaj niepoprawne.

Niezależnie od wybranej metody wprowadzania wszystkie zmienne, aby były wprowadzone do równania, muszą spełniać kryteria tolerancji. Domyślnym poziomem tolerancji jest 0,0001. Zmienna nie jest wprowadzana także, kiedy jej wprowadzenie spowodowałoby spadek tolerancji innej zmiennej już uwzględnionej w modelu poniżej kryterium tolerancji.

Wszystkie wybrane zmienne niezależne są dodawane do jednego modelu regresji. Można jednak również określić różne metody wprowadzania dla różnych podzbiorów zmiennych. Na przykład jeden blok zmiennych można wprowadzić do modelu regresji przy użyciu selekcji krokowej, a drugi za pomocą selekcji postępującej. Aby do modelu regresji dodać drugi blok zmiennych, kliknij przycisk **Dalej**.

Regresja liniowa: Filtrowanie

Analizie poddawane są obserwacje zdefiniowane przez regułę filtrowania. Na przykład jeśli wybrane zostaną zmienna, **równe** oraz wartość 5, to analiza obejmie tylko te obserwacje, w których wybrana zmienna ma wartość równą 5. Dozwolone są także wartości łańcuchowe.

Regresja liniowa: Wykresy

Wykresy są pomocne przy sprawdzaniu założeń dotyczących normalności, liniowości i równości wariancji. Pomagają również w wykrywaniu wartości odstających, obserwacji nieprzeciętnych oraz obserwacji wpływowych. Po zapisaniu ich jako nowych zmiennych, wartości przewidywane, reszty oraz inne informacje diagnostyczne są dostępne w Edytorze danych, umożliwiając tworzenie wykresów zmiennych niezależnych. Dostępne są następujące rodzaje wykresów:

Wykres rozrzutu. Na wykresie można umieścić dowolne dwa z poniższych elementów: zmienna zależna, standaryzowane wartości przewidywane, reszty standaryzowane, reszty usuniętych, skorygowane wartości przewidywane, reszty studentyzowane lub studentyzowane reszty usunięte. Wykres reszt standaryzowanych i wartości oczekiwanych pozwala sprawdzić liniowość i równość wariancji.

Lista zmiennych źródłowych. Wyświetla listę zmiennych zależnych (DEPENDNT) oraz następujące wartości przewidywane i reszty: standaryzowane wartości przewidywane (*ZPRED), standaryzowane reszty (*ZRESID), reszty usuniętych (*DRESID), skorygowane wartości przewidywane (*ADJPRED), studentyzowane reszty (*SRESID), studentyzowane reszty usuniętych (*SDRESID).

Przedstaw wszystkie wykresy cząstkowe. Ta opcja umożliwia wyświetlenie wykresów rozproszenia reszt każdej zmiennej niezależnej oraz reszt zmiennej zależnej, gdy obie zmienne są oddzielnie poddawane regresji względem innych zmiennych niezależnych. Aby można było przedstawić wykres cząstkowy, w równaniu muszą się znaleźć co najmniej dwie zmienne niezależne.

Wykresy reszt standaryzowanych. Można otrzymać histogramy reszt standaryzowanych oraz normalne wykresy prawdopodobieństwa, co pozwala na porównanie rozkładu reszt standaryzowanych z rozkładem normalnym.

Jeśli zostanie wybrany jakiś wykres, dla standaryzowanych wartości przewidywanych oraz reszt standaryzowanych wyświetlane są statystyki podsumowujące (***ZPRED** oraz ***ZRESID**).

Regresja liniowa: Zapisywanie zmiennych wynikowych

Można zapisać wartości przewidywane, reszty oraz inne statystyki przydatne w przypadku informacji diagnostycznych. Każde zaznaczenie powoduje dodanie jednej lub więcej nowych zmiennych do aktywnego pliku danych.

Wartości przewidywane. Wartości przewidywane za pomocą modelu regresji dla każdej obserwacji.

- *Niestandaryzowane.* Wartość zmiennej zależnej przewidywana przez model.
- *Standaryzowane.* Przekształcenie każdej przewidywanej wartości w jej znormalizowaną formę. przez odjęcie średniej wartości przewidywanej i podzieleniu różnicy przez jej odchylenie standardowe. Standaryzowana wartość przewidywana ma średnią 0 i odchylenie standardowe 1.
- *Skorygowane.* Przewidywana wartość obserwacji wykluczonej z obliczeń współczynników regresji.
- *Predykcje dla błędu standardowego średnich.* Błędy standardowe wartości przewidywanych. Oszacowanie odchylenia standardowego średniej wartości zmiennej zależnej dla obserwacji, które mają takie same wartości zmiennych niezależnych.

Odległości. Miary identyfikujące obserwacje zawierające nietypowe kombinacje wartości zmiennych niezależnych i obserwacje mające duży wpływ na model regresji.

- *Mahalanobisa.* Miara wielkości liter w zmiennych niezależnych od średniej dla wszystkich obserwacji. Duże wartości wskaźnika Mahalanobisa oznaczają, że obserwacja zawiera skrajne wartości jednej albo większej liczby zmiennych niezależnych.
- *Cooka.* Miara stopnia, w jakim zmieniłyby się wskaźniki reszt dla wszystkich obserwacji przy wykluczaniu poszczególnych obserwacji z obliczeń współczynników regresji. Duże wartości odległości

Cooka wskazują na to, że usunięcie obserwacji z obliczeń statystyk regresyjnych zmienia istotnie wielkość tych współczynników.

- *Wartości wpływu*. Mierzy wpływ punktu na dopasowanie regresji. Wycentrowane wartości wpływu zawierają się w przedziale od 0 (brak wpływu na dopasowanie) do $(N-1)/N$.

Przedziały predykcji. Dolna i górna granica przedziałów predykcji dla pojedynczej obserwacji oraz dla średniej.

- *Średnia*. Dolna i górna granica (dwie zmienne) przedziału predykcji dla średniej wartości przewidywanej.
- *Indywidualne*. Dolna i górna wartość graniczna przedziału, w którym mieści się przewidywana wartość zmiennej zależnej dla jednej obserwacji.
- *Przedział ufności*. Wprowadź wartość z zakresu od 1 do 99.99, aby określić poziom ufności dla dwóch przedziałów predykcji. Przed wprowadzeniem tej wartości należy wybrać opcję Średnia lub Pojedyncze. Typowe wartości przedziału ufności to 90, 95 i 99.

Reszty. Empiryczna wartość zmiennej zależnej minus wartość przewidywana za pomocą modelu regresji.

- *Niestandaryzowane*. Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- *Standaryzowane*. Reszta podzielona przez oszacowanie jego odchylenia standardowego. Standaryzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Studentyzowane*. Reszta podzielona przez oszacowanie jego odchylenia standardowego, które różni się w zależności od przypadku, w zależności od odległości między wartościami poszczególnych obserwacji w zmiennych niezależnych od średnich zmiennych niezależnych. Czasami nazywane wewnętrznymi resztami studentyzowanymi.
- *Usunięte*. Rezydualna dla przypadku, gdy ta sprawa jest wykluczona z obliczenia współczynników regresji. Jest to różnica pomiędzy wartością zmiennej zależnej, a jej skorygowaną wartością przewidywaną.
- *Studentyzowane usuniętych*. Usunięta reszta dla sprawy podzielonej przez jej błąd standardowy. Różnica między studentyzowaną, usuniętą resztką a powiązaną ze studentyzowaną resztką wskazuje, ile różnicy eliminuje przypadek w oparciu o własną predykcję. Czasami określane jako reszty studentyzowane zewnętrznie.

Statystyki wpływu. Zmiana współczynników regresji ($DfBeta$) oraz wartości przewidywanych ($DfFit$), wynikająca z wyłączenia określonej obserwacji. Standaryzowane wartości $DfBeta$ i $DfFit$ są także dostępne w ilorazie kowariancji.

- *$DfBeta$* . Różnica w wartości beta jest zmianą współczynnika regresji, który wynika z wyłączenia danej obserwacji. Wartość tej różnicy jest obliczana dla wszystkich składników modelu, łącznie ze stałą (wyrazem wolnym).
- *Standaryzowana $DfBeta$* . Standaryzowana różnica w wartości beta. Jest to zmiana we współczynniku regresji spowodowana usunięciem danej obserwacji. Zwykle warto poddać analizie obserwacje, których wartość bezwzględna standaryzowanej $DfBeta$ jest większa niż 2 podzielone przez pierwiastek kwadratowy z N , gdzie N oznacza liczbę obserwacji. Wartość tej różnicy jest obliczana dla wszystkich składników modelu, łącznie ze stałą (wyrazem wolnym).
- *$DfFit$* . Różnica w wartości dopasowania zmianą w wartości przewidywanej, wynikająca z wyłączenia danej obserwacji z oszacowania równania regresji.
- *Standaryzowane $DfFit$* . Różnica standaryzowana w wartości dopasowania. Jest to zmiana wartości przewidywanej spowodowana usunięciem danej obserwacji. Zwykle warto poddać analizie obserwacje, dla których wartość bezwzględna standaryzowanego $DfFit$ jest większa niż podwojony pierwiastek kwadratowy z p/N , gdzie p oznacza liczbę parametrów w modelu, a N liczbę obserwacji.
- *Współczynnik kowariancji*. Stosunek wyznacznika macierzy kowariancji z konkretnym przypadkiem wykluczonym z obliczenia współczynników regresji do wyznacznika macierzy kowariancji ze wszystkimi przypadkami uwzględnionymi. Jeśli iloraz jest bliski 1, dana obserwacja nie zmienia istotnie macierzy kowariancji.

Statystyki współczynników. Zachowuje współczynniki regresji w zbiorze danych lub plik danych. Zbiory danych są dostępne do późniejszego użytku w tej samej sesji lecz nie są zapisywane jako pliki, jeśli nie zostaną wprost zapisane pod koniec sesji. Nazwy zbiorów danych muszą być zgodne z regułami nazewnictwa zmiennych. .

Eksportuj informacje o modelu do pliku XML. Parametr szacuje i (opcjonalnie) i kowariancje są eksportowane do określonego pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

Regresja liniowa: Statystyki

Dostępne są następujące statystyki:

Współczynniki regresji- oszacowania : wyświetla współczynnik regresji B , błąd standardowy B , standaryzowany współczynnik beta, t wartość dla B oraz dwustronny poziom istotności t . **Przedziały ufności** powoduje wyświetlenie przedziałów ufności o określonym poziomie ufności dla każdego współczynnika regresji lub macierz kowariancji. Użycie opcji **Macierz kowariancji** umożliwia wyświetlenie macierzy wariancji-kowariancji współczynników regresji, w której kowariancje znajdują się poza przekątną, a wariancje znajdują się na przekątnej. Wyświetlana jest także macierz korelacji.

Dopasowanie modelu- wyświetlane są zmienne, które są wprowadzane i usuwane z modelu, oraz następujące "statystyki dobroci dopasowania": wiele R , R^2 i skorygowane R^2 , błąd standardowy oszacowania oraz tabela analizy wariancji.

Zmiana R kwadrat- Zmiana w statystyce R^2 , która jest tworzona przez dodanie lub usunięcie zmiennej niezależnej. Duża zmiana R^2 powiązana z daną zmienną oznacza, że zmienna ta jest dobrym predyktorem zmiennej zależnej.

Opisowy- Przedstawia liczbę ważnych obserwacji, średnią i odchylenie standardowe dla każdej zmiennej w analizie. Wyświetlane są także macierz korelacji z jednostronnym poziomem istotności oraz liczba obserwacji dla każdej korelacji.

Korelacja częściowa. Korelacja między zmienną zależną a zmienną niezależną, gdy efekty liniowe pozostałych zmiennych niezależnych w modelu zostały usunięte z niezależnej zmiennej. Jest ona powiązana ze zmianą w wartości R -kwadrat po dodaniu do równania zmiennej. Korelacja ta jest czasem zwana korelacją semicząstkową.

Korelacja cząstkowa. Korelacja, która pozostaje między dwiema zmiennymi po usunięciu korelacji, która wynika z ich wzajemnego powiązania z innymi zmiennymi. Korelacja pomiędzy zmienną zależną i zmienną niezależną, gdy efekty liniowe pozostałych zmiennych niezależnych w modelu zostały usunięte dla obu zmiennych.

Diagnostyka współliniowości- Nieliniowość (lub wielokoliniowość) jest niepożądaną sytuacją, gdy jedna niezależna zmienna jest funkcją liniową innych zmiennych niezależnych. Umożliwia wyświetlenie wartości własnych skalowanej i niecentrowanej macierzy iloczynów wektorowych, współczynników warunkowych oraz proporcji dekompozycji wariancji wraz z czynnikami nadmiaru wariancji (VIF) oraz wartościami tolerancji dla poszczególnych zmiennych.

Kryteria wyboru- obejmuje kryterium informacyjne Akaike (AIC), kryterium predykcyjne Ameniya (PC), Mallows warunkowa średnia kwadratowa błąd kryterium predykcji (C_p) i Schwarz Bayesowskie kryterium (SBC). Statystyki są wyświetlane w tabeli Podsumowanie modelu.

Reszty- Można wybrać opcję **'PRESS Statistic'**, aby użyć jako statystyki walidacji krzyżowej w celu porównania różnych modeli. Spowoduje to również wyświetlenie testu **'Durban-Watson'** w celu uzyskania korelacji seryjnej reszt. Wybierz **'Diagnostyka przypadku'** dla obserwacji, które spełniają kryterium wyboru (wartości odstających powyżej n odchyłeń standardowych).

Regresja liniowa: Opcje

Dostępne są następujące opcje:

Kryteria metod krokowych. Opcje te mają zastosowanie w przypadku określenia metody wyboru zmiennej terminowej, wstecznej lub krokowej. Zmienne można wprowadzać lub usuwać z modelu, zależy to od znaczenia (prawdopodobieństwa) wartości F lub samej wartości F .

- **Zastosuj prawdopodobieństwo F .** Zmienna jest wprowadzana do modelu, jeśli poziom istotności jego wartości F jest mniejszy od wartości pozycji i jest usuwany, jeśli poziom istotności jest większy niż wartość usunięcia. Wartość wprowadzenia musi być mniejsza od wartości usunięcia i obie muszą być dodatnie. Chcąc wprowadzić więcej zmiennych do modelu, należy zwiększyć wartość wprowadzenia. Aby usunąć więcej zmiennych, należy zmniejszyć wartość usunięcia.
- **Użyj wartości F .** Zmienna jest wprowadzana do modelu, jeśli jego wartość F jest większa niż wartość pozycji i jest usuwana, jeśli wartość F jest mniejsza niż wartość usunięcia. Wartość wprowadzenia musi być większa od wartości usunięcia i obie muszą być dodatnie. Chcąc wprowadzić więcej zmiennych do modelu, należy obniżyć wartość wprowadzenia. Chcąc usunąć więcej zmiennych, należy zwiększyć wartość usunięcia.

Tolerancja. Wartością domyślną jest .0001. Tolerancja to proporcja wariancji zmiennej w równaniu, która nie jest uwzględniana przez inne zmienne niezależne w równaniu. Minimalna tolerancja dla każdej zmiennej w równaniu, jeśli analizowana zmienna została uwzględniona w analizie, to minimalna tolerancja zmiennej, która nie została uwzględniona w równaniu. Zmienne muszą przejść testy tolerancji i minimum tolerancji, aby wprowadzić i pozostać w równaniu regresji. Jeśli zmienna spełnia kryteria tolerancji, kwalifikuje się ona do włączenia w oparciu o obowiązujący metodę.

Uwzględnij stałą w równaniu. Domyślnie, model regresji uwzględnia pewną stałą. Usunięcie zaznaczenia tej opcji powoduje, że linia regresji będzie przechodzić przez środek układu współrzędnych, czego zazwyczaj się nie stosuje. Niektóre wyniki przy linii regresji przechodzącej przez środek układu współrzędnych nie dają się porównać z wynikami regresji uwzględniającej stałą. Na przykład wartość R^2 nie może być interpretowana w zwykły sposób.

Braki danych. Można wybrać jedną z następujących opcji:

- **Wyłączanie wszystkich obserwacji.** Analizy obejmują jedynie obserwacje zawierające poprawne wartości dla wszystkich zmiennych.
- **Wyłączanie obserwacji parami.** Do obliczania współczynników korelacji, na których opiera się analiza regresji, używane są obserwacje z pełnymi danymi dla pary korelowanych zmiennych. Stopnie swobody są obliczane na podstawie minimalnej liczby par N .
- **Zastępowanie średnią.** W obliczeniach wykorzystywane są wszystkie obserwacje, przy czym brakujące obserwacje zastępowane są średnią dla danej zmiennej.

Dodatkowe właściwości komendy REGRESSION

Język składni komend umożliwia również:

- Zapisanie macierzy korelacji lub odczytanie macierzy zamiast danych surowych w celu uzyskania analizy regresji (za pomocą opcji komendy MACIERZ).
- Określanie poziomów tolerancji (za pomocą opcji komendy CRITERIA).
- Uzyskanie wielu modeli dla tych samych lub różnych zmiennych zależnych (za pomocą opcji komend METHOD i DEPENDENT).
- Uzyskiwanie dodatkowych statystyk (za pomocą opcji komend DESCRIPTIVES i STATISTICS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Regresja porządkowa

Regresja porządkowa umożliwia modelowanie zależności skumulowanej odpowiedzi porządkowej na zestawie predyktorów, które mogą być czynnikami lub współzmiennymi. Projekt regresji porządkowej jest oparty na metodologii McCullagha (1980, 1998), a nazwa procedury w składni systemu to PLUM.

Standardowa analiza regresji liniowej polega na zminimalizowaniu sumy kwadratów różnic pomiędzy zmienną odpowiedzi (zależną) a ważoną kombinacją zmiennych predyktora (niezależnych). Oszacowane

współczynniki odzwierciedlają sposób, w jaki zmiany w predyktorach oddziałują na odpowiedź. Zakładana odpowiedź jest numeryczna, a zmiany poziomu odpowiedzi są równorzędne w całym zakresie odpowiedzi. Na przykład różnica wzrostu pomiędzy osobą o wzroście 150 cm i osobą o wzroście 140 cm wynosi 10 cm, co oznacza taką samą różnicę pomiędzy osobą o wzroście 210 cm i 200 cm. Relacje te nie zawsze są zachowane dla zmiennych porządkowych, w których dobór kategorii odpowiedzi i ich liczba mogą być całkiem dowolne.

Przykład. Regresji porządkowej można użyć do badania reakcji pacjenta na dawkowanie leku. Możliwe reakcje można sklasyfikować jako *brak*, *łagodna*, *umiarkowana* lub *ostra*. Różnica pomiędzy reakcją łagodną a umiarkowaną jest trudna lub wręcz niemożliwa do określenia i jest oparta na percepcji. Co więcej, różnica pomiędzy reakcją łagodną a umiarkowaną może być większa lub mniejsza niż różnica pomiędzy reakcją umiarkowaną a ostrą.

Statystyki i wykresy. Częstości obserwowane, oczekiwane i skumulowane, reszty Pearsona dla częstości i częstości skumulowanych, prawdopodobieństwa obserwowane i oczekiwane, obserwowane i oczekiwane skumulowane prawdopodobieństwa dla każdej kategorii reakcji według relacji współzmiennych, asymptotyczne macierze korelacji i macierze kowariancji dla oszacowań parametrów, współczynniki chi-kwadrat Pearsona i współczynniki chi-kwadrat ilorazu wiarygodności, statystyka dobroci dopasowania, przebieg iteracji, test założenia linii równoległych, oszacowania parametrów, błąd standardowy, przedziały ufności, a także statystyka R^2 Coxa i Snella, Nagelkerke'a i McFaddena.

Wymagania dotyczące danych w regresji porządkowej

Dane. Zakłada się, że zmienna zależna jest porządkowa i może być wartością liczbową lub łańcuchem. Uporządkowanie jest określone przez sortowanie wartości zmiennej zależnej w kolejności malejącej. Najniższa wartość definiuje pierwszą kategorię. Zakłada się, że czynniki są jakościowe. Współzmiennie muszą mieć postać liczbową. Należy zauważyć, że użycie więcej niż jednej ciągłej współzmiennej może spowodować uzyskanie bardzo dużej tabeli prawdopodobieństwa komórki.

Założenia. Dozwolona jest tylko jedna zmienna odpowiedzi i należy ją określić. Podobnie w przypadku każdego oddzielnego wzorca wartości w zmiennych niezależnych zakłada się, że odpowiedzi są niezależnymi zmiennymi wielomianowymi.

Procedury pokrewne. Nominalna regresja logistyczna wykorzystuje podobne modele dla nominalnych zmiennych zależnych.

Uzyskiwanie regresji porządkowej

1. Z menu wybierz:

Analiza > Regresja > Porządkowy...

2. Wybierz jedną zmienną zależną.

3. Kliknij **OK**.

Opcje regresji porządkowej

Okno dialogowe Opcje umożliwia dostosowanie parametrów używanych w iteracyjnym algorytmie estymacji. Wybierz poziom ufności dla oszacowań parametrów, a następnie funkcję wiążącą.

Iteracje. Można dostosować algorytm iteracyjny.

- **Maksymalna liczba iteracji.** Podaj nieujemną liczbę całkowitą. Jeśli zostanie podane 0, procedura zwróci wstępne oszacowania.
- **Maksimum kroków połowienia.** Podaj dodatnią liczbę całkowitą.
- **Zbieżność logarytmu wiarygodności.** Algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana funkcji logarytmu wiarygodności jest mniejsza od tej wartości. Kryterium nie jest stosowane, jeśli wartość wynosi 0.
- **Zbieżność parametru.** Algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana każdego oszacowania parametru jest mniejsza od tej wartości. Kryterium nie jest stosowane, jeśli wartość wynosi 0.

Przedział ufności. Podaj wartość większą lub równą 0 i mniejszą od 100.

Delta. Wartość dodawana do zerowych częstości w komórkach. Podaj liczbę nieujemną mniejszą od 1.

Tolerancja osobliwości. Jest używana do sprawdzania wysoce zależnych predyktorów. Wybierz wartość z listy opcji.

Funkcja łączenia. Funkcja łączenia to transformacja skumulowanych prawdopodobieństw, która umożliwia estymację modelu. Dostępnych jest pięć funkcji łączenia.

- **Logit.** $f(x)=\log(x/(1-x))$. Zwykle używane w przypadku kategorii równomiernie rozproszonych.
- **Komplementarny logarytmiczno-logarytmiczny.** $f(x)=\log(-\log(1-x))$. Zazwyczaj używane, gdy wyższe kategorie są bardziej prawdopodobne.
- **Ujemny logarytmiczno-logarytmiczny.** $f(x)=-\log(-\log(x))$. Zazwyczaj używane, gdy niższe kategorie są bardziej prawdopodobne.
- **Probit.** $f(x)=\Phi^{-1}(x)$. Zazwyczaj używane, gdy zmienna utajona ma rozkład normalny.
- **Cauchit (odwrócony Cauchy'ego).** $f(x)=\tan(\pi(x-0,5))$. Zazwyczaj używane, gdy zmienna utajona zawiera wiele wartości skrajnych.

Wyniki regresji porządkowej

Okno dialogowe Wynik umożliwia wygenerowanie tabel, które można wyświetlić w przeglądarce, a zmienne można zapisać w roboczym pliku danych.

Pokaż. Generuje tabele do:

- **Pokaż przebieg iteracji co każdy n kroków.** Oceny parametrów i logarytmu wiarygodności są wyświetlane z określoną częstością iteracji wyświetlania. Zawsze wyświetlane są pierwsza i ostatnia iteracja.
- **Statystyki dobroci dopasowania.** Statystyki chi-kwadrat Pearsona oraz ilorazu wiarygodności chi-kwadrat. Są one obliczane na podstawie klasyfikacji podanej w liście zmiennych.
- **Statystyki podsumowujące.** Statystyki R^2 Coxa i Snella, Nagelkerkego i McFaddena.
- **Oceny parametrów.** Oceny parametrów, błędy standardowe i przedziały ufności.
- **Asymptotyczna korelacja ocen parametrów.** Macierz korelacji oszacowania parametrów.
- **Asymptotyczna kowariancja ocen parametrów.** Macierz kowariancji oszacowania parametrów.
- **Informacje o komórkach.** Częstości obserwowane, oczekiwane i skumulowane, reszty Pearsona dla częstości i częstości skumulowanych, prawdopodobieństwa obserwowane i oczekiwane, a także obserwowane i oczekiwane skumulowane prawdopodobieństwa dla każdej kategorii reakcji według relacji współzmiennych. Należy zauważyć, że w przypadku modeli z wieloma relacjami współzmiennych (np. modeli z ciągłymi współzmiennymi) ta opcja może wygenerować bardzo dużą i nieporęczną tabelę.
- **Test równoległości linii.** Test hipotezy zakładającej, że parametry położenia są jednakowe dla wszystkich poziomów zmiennej zależnej. Opcja ta jest dostępna wyłącznie dla modelu uwzględniającego tylko położenie.

Zapisywane zmienne. Umożliwia zapisanie następujących zmiennych w roboczym pliku danych:

- **Estymowane prawdopodobieństwa reakcji.** Prawdopodobieństwa oszacowane przez model klasyfikacji relacji czynnik/współzmienna w kategoriach odpowiedzi. Występuje tyle prawdopodobieństw, ile kategorii odpowiedzi.
- **Przewidywana kategoria.** Kategoria odpowiedzi, która ma maksymalne oszacowane prawdopodobieństwo dla relacji czynnik/współzmienna.
- **Prawdopodobieństwo przewidywanej kategorii.** Oszacowane prawdopodobieństwo klasyfikacji relacji czynnik/współzmienna w przewidywanej kategorii. To prawdopodobieństwo jest także wartością maksymalną oszacowanych prawdopodobieństw relacji czynnik/współzmienna.
- **Prawdopodobieństwo rzeczywistej kategorii.** Oszacowane prawdopodobieństwo klasyfikacji relacji czynnik/współzmienna w kategorii rzeczywistej.

Pokaż logarytm wiarygodności. Kontroluje wyświetlanie logarytmu wiarygodności. **Uwzględnij stałą wielomianu** zapewnia pełną wartość wiarygodności. Aby porównać wyniki pomiędzy składnikami bez stałej, można ją wykluczyć.

Model położenia regresji porządkowej

Okno dialogowe Położenie umożliwia określenie modelu położenia dla analizy.

Określ model. Model efektów głównych zawiera efekty główne współzmiennych i czynników, ale nie zawiera efektów interakcji. Można utworzyć model niestandardowy w celu określenia podzbiorów interakcji czynników lub interakcji współzmiennych.

Czynniki/współzmiennie. Lista zawiera czynniki i współzmiennie.

Pozycja. Model zależy od wybranych efektów głównych oraz efektów interakcji.

Dla wybranych czynników i współzmiennych:

Interakcje

Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu. Jest to wartość domyślna.

Efekty główne

Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

Wszystkie 2. rzędu

Tworzy wszystkie możliwe interakcje drugiego rzędu wybranych zmiennych.

Wszystkie 3 rzędu

Tworzy wszystkie możliwe interakcje trzeciego rzędu wybranych zmiennych.

Wszystkie 4. rzędu

Tworzy wszystkie możliwe interakcje czwartego rzędu wybranych zmiennych.

Wszystkie 5 rzędu

Tworzy wszystkie możliwe interakcje piątego rzędu wybranych zmiennych.

Budowanie składników i składniki użytkownika

Buduj składniki

Tej opcji należy użyć, chcąc uwzględnić niezagnieżdżone składniki konkretnego typu (takie jak efekty główne) dla wszystkich kombinacji wybranego zestawu czynników i współzmiennych.

Buduj składniki użytkownika

Tej opcji należy użyć, chcąc uwzględnić składniki zagnieżdżone lub konstruować składnik jawnie, zmienna po zmiennej. Proces budowania składnika zagnieżdżonego obejmuje następujące kroki:

Model skali regresji porządkowej

Okno dialogowe Skala umożliwia określenie modelu skali dla analizy.

Czynniki/współzmiennie. Lista zawiera czynniki i współzmiennie.

Skala. Model zależy od wybranych efektów głównych oraz efektów interakcji.

Dla wybranych czynników i współzmiennych:

Interakcje

Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu. Jest to wartość domyślna.

Efekty główne

Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

Wszystkie 2. rzędu

Tworzy wszystkie możliwe interakcje drugiego rzędu wybranych zmiennych.

Wszystkie 3 rzędu

Tworzy wszystkie możliwe interakcje trzeciego rzędu wybranych zmiennych.

Wszystkie 4. rzędu

Tworzy wszystkie możliwe interakcje czwartego rzędu wybranych zmiennych.

Wszystkie 5 rzędu

Tworzy wszystkie możliwe interakcje piątego rzędu wybranych zmiennych.

Budowanie składników i składniki użytkownika

Buduj składniki

Tej opcji należy użyć, chcąc uwzględnić niezagnieżdżone składniki konkretnego typu (takie jak efekty główne) dla wszystkich kombinacji wybranego zestawu czynników i współzmiennych.

Buduj składniki użytkownika

Tej opcji należy użyć, chcąc uwzględnić składniki zagnieżdżone lub konstruować składnik jawnie, zmienna po zmiennej. Proces budowania składnika zagnieżdżonego obejmuje następujące kroki:

Dodatkowe właściwości komendy PLUM

Można dostosować regresję porządkową, wklejając wybrane elementy do okna edytora komend i edytować wynikającą składnię komendy PLUM. Język składni komend umożliwia również:

- Tworzenie dostosowanych testów hipotezy poprzez określanie hipotez zerowych jako liniowych kombinacji parametrów.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Regresja liniowa sieci elastycznej

Liniowa Sieć Elastyczna wykorzystuje klasę Python `sklearn.linear_model.ElasticNet` do szacowania stałych modeli regresji liniowej dla zmiennej zależnej w jednej lub większej liczbie zmiennych niezależnych. Regularizacja łączy w sobie kary L1 (Lasso) i L2 (Ridge). Rozszerzenie obejmuje opcjonalne tryby wyświetlania wykresów śledzenia dla różnych wartości alfa dla danego współczynnika L1, a także wybranie współczynnika L1 i wartości hiperparametru alfa w oparciu o walidację krzyżową. Po dopasowaniu pojedynczego modelu lub w celu sprawdzenia, czy walidacja krzyżowa jest używana do wyboru współczynnika kar i (lub) alfa, do oszacowania wydajności poza próbką można użyć partycji danych wstrzymanych.

Oprócz dopasowania modelu o określonych wartościach współczynnika kary L1 i parametru regulacji alfa, elastyczna sieć liniowa może wyświetlać wykres śledzenia wartości współczynnika dla zakresu wartości alfa dla danego współczynnika lub ułatwić wybór wartości parametrów hiperparametrów poprzez interwalidację k-krotnie w określonych siatkach wartości. Jeśli pojedynczy model jest dopasowany lub stosowany jest stosunek i/lub wybór alfa poprzez walidację krzyżową, to ostateczny model może być zastosowany do danych wstrzymanych, które są tworzone przez partycję danych wejściowych w celu uzyskania poprawnego oszacowania wydajności pozapróbowej modelu.

Uzyskiwanie analizy regresji metodą elastycznej sieci liniowej

1. Z menu wybierz:

Analiza > Regresja > Alternatywy liniowe OSS > siatka elastyczna

Okno dialogowe umożliwia określenie zmiennej, która przypisuje każdą obserwację w aktywnym zbiorze danych do próby szkoleniowej lub holdout.

2. Wybierz liczbową zmienną przewidywaną. Do uruchomienia analizy wymagana jest tylko jedna zmienna przewidywana.
3. Określ wartość liczbową zależną.
4. Określ co najmniej jedną zmienną kategoriową lub zmienną współzmienną liczbową.

Opcjonalnie **Partycja** umożliwia utworzenie holdout lub podzbioru testowego danych wejściowych w celu oszacowania wydajności poza próbką określonego modelu lub wybranego modelu. Wszystkie partycjonowanie wykonywane jest po listowym usunięciu wszystkich obserwacji z niepoprawnymi danymi dla dowolnej zmiennej użytej w procedurze. Należy pamiętać, że w przypadku walidacji krzyżowej dane treningowe są tworzone w języku Python. Dane wstrzymane, które są tworzone przez partycję, nie są używane w estymacji, niezależnie od tego, jaki tryb jest stosowany.

Partycję można zdefiniować, określając współczynnik obserwacji losowo przypisanych do każdej próbki (w sekcji **Partycje szkoleniowe i wstrzymane**) lub przez zmienną, która przypisuje każdą obserwację do ucznia lub próby holdout. Nie można określić zarówno szkolenia, jak i zmiennych. Jeśli partycja nie zostanie określona, zostanie utworzona próbka wstrzymana, która wynosi około 30% danych wejściowych.

Wartość **Szkolenia%** określa względną liczbę obserwacji w aktywnym zbiorze danych, które losowo przypisują do próby ucznia. Szkolenie domyślne to 70%.

Regresja elastyczna sieci liniowej: Opcje

Na karcie Opcje dostępne są opcje dotyczące:

Dominanta

Ten wybór udostępnia opcje służące do określania jednego z następujących trybów:

Dopasuj do określonego współczynnika L1 i alfa

Po wybraniu tego trybu pojedynczy model jest dopasowany do danych uczących przy użyciu określonych wartości współczynnika L1 i wartości regulacji alfa. Jest to ustawienie domyślne. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W obszarze **Wykres** można wybrać wykresy obserwowanych i/lub reszt względem wartości przewidywanych.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Wykres śladowy

Po wybraniu tego trybu, trzy wykresy dla danych uczących są wyświetlane jako funkcja alfa dla określonego zestawu wartości alfa:

- Wykres śledzenia współczynników regresji.
- Wykres R^2 .
- Wykres średniej kwadratowej błędów kwadratowego (MSE).

Chociaż partycja jest uhonorowana, nie są udostępniane żadne wyniki dla danych testowych wstrzymanych, ponieważ żaden ostatni model nie wynika z tego trybu.

Współczynnik L1 i/lub wybór alfa poprzez walidację krzyżową

Po wybraniu tego trybu zostanie przeprowadzone wyszukiwanie w siatce z krzyżowym sprawdzaniem poprawności w celu oceny modeli, a najlepszy współczynnik i wartości alfa są wybierane na podstawie najlepszej średniej R^2 dla krotności walidacji. Pole **Liczba krotności walidacji krzyżowej** może być używane do zmiany domyślnej wartości pięciu podziłów lub fałdów na potrzeby walidacji krzyżowej. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W sekcji **Wyświetlanie** można wybrać opcję wyświetlania podstawowych informacji o modelu tylko z wybraną wartością współczynnika i alfa (**Najlepszy**), podstawowe informacje o wszystkich modelach w porównaniu (**Porównaj modele**) lub pełne informacje o wszystkich podziałach lub składkach dla wszystkich modeli (**Porównaj modele i złożone**). Wartością domyślną jest **Najlepszy**.

W obszarze **Wykres** można wybrać wykresy średniej R^2 i/lub MSE przy użyciu krotności walidacji. Wykresy obserwowanych i/lub reszt względem wartości przewidywanych mogą być również wybrane.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Określ pojedyncze współczynniki L1

Po wybraniu tego trybu dla opcji **Dopasuj do określonego współczynnika L1 i alfa** lub **Wykres śledzenia** można określić pojedynczą wartość współczynnika kary L1. W przypadku wybrania dla trybu **L1 i/lub wyboru alfa w trybie sprawdzania krzyżowego** można określić wiele wartości.

Określ siatkę wartości alfa

Po wybraniu tego trybu dla trybu **L1 i/lub wyboru alfa przez sprawdzanie krzyżowe**, siatka unikalnych wartości alfa może być określona z wartości **Początek** (value1) do wartości **Koniec** (value2) z przyrostem **Wg** (value3). Jeśli zostanie podany, dozwolony jest tylko jeden poprawny zestaw [value1 T0 value2 BY value3]. Musi on spełniać następujące warunki: $0 \leq \text{value1} \leq \text{value2} \leq 1$. W przypadku, gdy produkt $\text{value1} = \text{value2}$ jest równoważny z określeniem pojedynczego value1, bez względu na value3.

Wykresy są wyświetlane przy użyciu określonej wielkości mierzonej dla poziomych osi X zmiennych wartości alfa.

Określ pojedyncze wartości alfa

Po wybraniu tego trybu dla trybu **Dopasuj do określonego współczynnika L1 i alfa** można określić pojedynczą wartość regulowania alfa. Po wybraniu dla opcji **Wykres śledzenia** lub **L1 i/lub wybór alfa w trybie sprawdzania krzyżowego** można określić wiele wartości.

Określ siatkę wartości alfa

Po wybraniu tego trybu dla trybu **Wykres śledzenia** lub **L1 i/lub wybór alfa przez sprawdzanie krzyżowe**, siatka unikalnych wartości alfa może być określona z wartości **Początek** (value1) do wartości **Koniec** (value2) z przyrostem **Wg** (value3). Jeśli zostanie podany, dozwolony jest tylko jeden poprawny zestaw [value1 T0 value2 BY value3]. Musi on spełniać następujące warunki: $0 \leq \text{value1} \leq \text{value2} \leq 1$. W przypadku, gdy produkt $\text{value1} = \text{value2}$ jest równoważny z określeniem pojedynczego value1, bez względu na value3.

Metryka Alfa dla zakresów wartości może mieć wartość **Liniowa** lub **Podstawowa 10 logarytmiczna** (10 jest podniesiona do potęgi określonych wartości).

Wykresy są wyświetlane przy użyciu określonej wielkości mierzonej dla poziomych osi X zmiennych wartości alfa.

Kryteria

Steruje analizami.

Uwzględnij wyraz wolny

To kryterium obejmuje wyraz wolny w dopasowanym modelu (modelach). Należy zauważyć, że procedura rozszerzenia nie jest w centrum ani nie standaryzuje zmiennej zależnej, a wyraz wolny nie jest karany w trakcie oszacowania.

Standaryzacja predyktorów

Standaryzuje wszystkie zmienne niezależne.

Liczba krotności walidacji krzyżowej

Liczba splitów lub fałdów do oceny krzyżowej modeli. Wartość musi być dodatnią liczbą całkowitą większą niż 1. Wartością domyślną jest 5.

Stan losowy Python

Wartość ustawienia random_state w języku Python używana podczas wykonywania wartościowania walidacji krzyżowej modeli. Umożliwia on replikację wyników, które obejmują pseudolosowe liczby. Wartość musi być liczbą całkowitą z zakresu od 0 do $2^{32}-1$. Domyślną wartością jest 0.

Limit czasu (minuty)

Liczba minut dozwolona dla obliczeń modelu, które mają zostać uruchomione. Jeśli zostanie podana wartość 0, licznik czasu zostanie wyłączony. Wartość domyślna to 5.

Wyświetlanie odpowiedzi

Ta opcja określa ilość danych wyjściowych do wyświetlenia dla trybu **L1 i/lub wyboru alfa w trybie walidacji krzyżowej**.

Najlepsza

Wyświetla tylko podstawowe wyniki dla wybranego najlepszego modelu. Jest to ustawienie domyślne.

Porównywanie modeli

Wyświetla podstawowe wyniki dla wszystkich ocenionych modeli.

Porównywanie modeli i krotności

Wyświetla pełne szczegółowe wyniki dla każdego podziału lub krotności dla każdego ocenianego modelu.

Wykres

Ta opcja określa wykresy wartości obserwowanych lub rezydualnych w porównaniu z wartościami przewidywanymi oraz ze sprawdzaniem krzyżowym, specyfikacją wykresów średniej średniej kwadratowej błędów (MSE) i/lub średnią R^2 w przekrojowych składach walidacyjnych a wartości alfa.

Średni błąd średniej walidacji krzyżowej (MSE) w porównaniu z alfa

Dla trybu **L1 i/lub wyboru alfa przez sprawdzanie krzyżowe** wyświetlany jest wykres liniowy średniej liczby krotności walidacji krzyżowej w porównaniu z alfa dla podanej lub wybranej najlepszej wartości współczynnika $L1$. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Średnia krzyżowa walidacja R^2 kwadrat a alfa

Dla trybu **L1 i/lub wyboru alfa przez sprawdzanie krzyżowe** wyświetlany jest wykres liniowy średniej liczby R^2 dla krotności krzyżowej walidacji a alfa dla podanej lub wybranej najlepszej wartości współczynnika $L1$. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Obserwowany a przewidywany

Wyświetla wykres rozrzutu obserwowanych i przewidywanych wartości dla określonego lub najlepszego modelu.

Reszty względem przewidywanego

Wyświetla wykres rozrzutu reszt w porównaniu z wartościami przewidywanymi dla określonego lub najlepszego modelu.

Zapisz

Określa zmienne, które mają zostać zapisane w aktywnym zestawie danych.

Wartości przewidywane

Zapisz przewidywane wartości z określonego lub najlepszego modelu w aktywnym zestawie danych. Można również określić wartość **Nazwa zmiennej niestandardowej**.

Reszty

Zapisz reszty z określonych lub najlepszych predykcji modelu do aktywnego zestawu danych. Można również określić wartość **Nazwa zmiennej niestandardowej**.

Regresja liniowa lasso

Liniowa Lasso używa klasy Python `sklearn.linear_model.Lasso` do oszacowania $L1$ modeli regresji liniowej dla zmiennej zależnej w jednej lub większej liczby zmiennych niezależnych i zawiera opcjonalne tryby do wyświetlania wykresów śledzenia oraz do wybierania wartości parametru hyperparameter alfa na podstawie walidacji krzyżowej. Po dopasowaniu pojedynczego modelu lub w celu sprawdzenia, czy do wyboru jest alpha, można użyć partycji danych holdout w celu oszacowania wydajności pojedynczego.

Oprócz dopasowania modelu o określonej wartości parametru regulacji alfa, lasso liniowa może wyświetlać wykres śledzenia wartości współczynników dla zakresu wartości alfa lub ułatwić wybór wartości nadparametrycznej poprzez interwalidację k -krotnie w określonych siatkach wartości. Jeśli pojedynczy model jest dopasowany lub wybierany jest wybór alfa poprzez walidację krzyżową, ostateczny model może być zastosowany do danych wstrzymanych, które są tworzone przez partycję danych wejściowych w celu uzyskania poprawnego oszacowania wydajności pozapróbowej modelu.

Uzyskiwanie analizy regresji liniowej Lasso Regresja

1. Z menu wybierz:

Analiza > Regresja > Alternatywy liniowe OLS > Lasso

Okno dialogowe umożliwia określenie zmiennej, która przypisuje każdą obserwację w aktywnym zbiorze danych do próby szkoleniowej lub holdout.

2. Wybierz liczbową zmienną przewidywaną. Do uruchomienia analizy wymagana jest tylko jedna zmienna przewidywana.

3. Określ wartość liczbową zależną.

4. Określ co najmniej jedną zmienną kategoriową lub zmienną współzmienną liczbową.

Opcjonalnie **Partycja** umożliwia utworzenie holdout lub podzbioru testowego danych wejściowych w celu oszacowania wydajności poza próbką określonego modelu lub wybranego modelu. Wszystkie partycjonowanie wykonywane jest po listowym usunięciu wszystkich obserwacji z niepoprawnymi danymi dla dowolnej zmiennej użytej w procedurze. Należy pamiętać, że w przypadku walidacji krzyżowej dane treningowe są tworzone w języku Python. Dane holdout utworzone przez partycję nie są używane w estymacji, niezależnie od tego, jaki tryb jest stosowany.

Partycję można zdefiniować, określając iloraz obserwacji losowo przypisanych do każdej próbki (w sekcji **Szkolenia i partycje Holdout**) lub przez zmienną, która przypisuje każdą obserwację do szkolenia lub próby holdout. Nie można określić zarówno szkolenia, jak i zmiennych. Jeśli partycja nie zostanie określona, zostanie utworzona próbka wstrzymana, która wynosi około 30% danych wejściowych.

Wartość **Szkolenia%** określa względną liczbę obserwacji w aktywnym zbiorze danych, które losowo przypisują do próby ucznia. Szkolenie domyślne to 70%.

Regresja liniowa lasso: Opcje

Na karcie Opcje dostępne są opcje dotyczące:

Dominanta

Ten wybór udostępnia opcje służące do określania jednego z następujących trybów:

Dopasuj do określonej wartości alfa

Po wybraniu tego trybu pojedynczy model jest dopasowany do danych uczących za pomocą tylko jednej wartości regulacyjnej alfa. Jest to ustawienie domyślne. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W obszarze **Wykres** można wybrać wykresy obserwowanych i/lub reszt względem wartości przewidywanych.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Wykres śladowy

Po wybraniu tego trybu, trzy wykresy dla danych uczących są wyświetlane jako funkcja alfa dla określonego zestawu wartości alfa:

- Wykres śledzenia współczynników regresji.
- Wykres R^2 .
- Wykres średniej kwadratowej błędów kwadratowego (MSE).

Chociaż partycja jest uhonorowana, nie są udostępniane żadne wyniki dla danych testowych wstrzymanych, ponieważ żaden ostatni model nie wynika z tego trybu.

Wybór alfa poprzez walidację krzyżową

Po wybraniu siatki wyszukiwania za pomocą walidacji krzyżowej w celu oceny modeli i wybrania najlepszego alfa na podstawie najlepszej średniej, R^2 nad składami sprawdzania poprawności. Pole **Liczba krotności walidacji krzyżowej** może być używane do zmiany domyślnej wartości pięciu

podziów lub fałdów na potrzeby walidacji krzyżowej. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W sekcji **Wyświetlanie** można wybrać opcję wyświetlania podstawowych informacji o modelu tylko z wybraną wartością alfa (**Najlepszy**), podstawowe informacje o wszystkich modelach w porównaniu (**Porównaj modele**) lub pełne informacje o wszystkich podziałach lub składkach dla wszystkich modeli (**Porównaj modele i złożone**). Wartością domyślną jest **Najlepszy**.

W obszarze **Wykres** można wybrać wykresy średniej R^2 i/lub MSE przy użyciu krotności walidacji. Wykresy obserwowanych i/lub reszt względem wartości przewidywanych mogą być również wybrane.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Określ pojedyncze wartości alfa

Po wybraniu trybu **Dopasuj do określonego alfa** można określić pojedynczą wartość regulacji alfanumerycznych. Po wybraniu opcji **Wykres śledzenia** lub **Wybór alfa przez walidację krzyżową** można określić wiele wartości.

Wartości

Określ co najmniej jedną dodatnią wartość regularyzacji alfa. Wiele wartości może być określonych pojedynczo lub jako zakresy. Domyślną wartością jest 1.

Określ siatkę wartości alfa

Po wybraniu trybu **Wykres śledzenia** lub **Wybór alfa w wyniku walidacji krzyżowej** siatka unikalnych wartości alfa może być określona na podstawie wartości **Początek** (value1) do wartości **Koniec** (value2) z przyrostem **Wg** (value3). Jeśli zostanie podany, dozwolony jest tylko jeden poprawny zestaw [value1 TO value2 BY value3]. Musi on spełniać następujące warunki: $0 \leq \text{value1} \leq \text{value2} \leq 1$. W przypadku, gdy produkt $\text{value1} = \text{value2}$ jest równoważny z określeniem pojedynczego value1, bez względu na value3.

Metryka Alfa dla zakresów wartości może mieć wartość **Liniowa** lub **Podstawowa 10 logarytmiczna** (10 jest podniesiona do potęgi określonych wartości).

Wykresy są wyświetlane przy użyciu określonej wielkości mierzonej dla poziomych osi X zmiennych wartości alfa.

Kryteria

Steruje analizami.

Uwzględnij wyraz wolny

Zawiera wyraz wolny w dopasowanym modelu (modelach). Należy zauważyć, że procedura rozszerzenia nie jest w centrum ani nie standaryzuje zmiennej zależnej, a wyraz wolny nie jest karany w trakcie oszacowania.

Standaryzacja predyktorów

Standaryzuje wszystkie zmienne niezależne.

Liczba krotności walidacji krzyżowej

Liczba splitów lub fałdów do oceny krzyżowej modeli. Wartość musi być dodatnią liczbą całkowitą większą niż 1. Wartością domyślną jest 5.

Stan losowy Python

Wartość ustawienia random_state w języku Python używana podczas wykonywania wartościowania krzyżowego modeli. Umożliwia replikację wyników, które obejmują pseudolosowe liczby. Musi to być liczba całkowita z zakresu od 0 do $2^{32}-1$. Domyślną wartością jest 0.

Limit czasu (minuty)

Liczba minut dozwolona dla obliczeń modelu, które mają zostać uruchomione. Jeśli zostanie podana wartość 0, licznik czasu zostanie wyłączone. Wartość domyślna to 5.

Wyświetlanie podpowiedzi

Określa ilość danych wyjściowych do wyświetlenia dla trybu **Wybór alfa w drodze walidacji krzyżowej**.

Najlepsza

Wyświetla tylko podstawowe wyniki dla wybranego najlepszego modelu. Jest to ustawienie domyślne.

Porównywanie modeli

Wyświetla podstawowe wyniki dla wszystkich ocenionych modeli.

Porównywanie modeli i krotności

Wyświetla pełne szczegółowe wyniki dla każdego podziału lub krotności dla każdego ocenianego modelu.

Wykres

Określa wykresy wartości obserwowanych lub rezydualnych w porównaniu z wartościami przewidywanymi oraz ze sprawdzaniem krzyżowym, specyfikacją wykresów średniej kwadratowej błędów (MSE) i/lub średnią R^2 w przekrojowych składach walidacyjnych w porównaniu z wartościami alfa.

Średni czas sprawdzania krzyżowego błędów kwadratowych (MSE) w porównaniu z alfa

Dla trybu **Wybór alfa przez sprawdzanie krzyżowe** wyświetlany jest wykres liniowy średniej liczby krotności walidacji krzyżowej (MSE) w porównaniu z alfa. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Średnia krzyżowa walidacja R^2 kwadratowa a alfa

Dla trybu **Wybór alfa przez sprawdzanie krzyżowe** wyświetlany jest wykres liniowy średniej liczby R^2 dla krotności walidacji krzyżowej w porównaniu z alfa. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Obserwowany a przewidywany

Wyświetla wykres rozrzutu obserwowanych i przewidywanych wartości dla określonego lub najlepszego modelu.

Reszty względem przewidywanego

Wyświetla wykres rozrzutu reszt w porównaniu z wartościami przewidywanymi dla określonego lub najlepszego modelu.

Zapisz

Określa zmienne, które mają zostać zapisane w aktywnym zestawie danych.

Wartości przewidywane

Zapisz przewidywane wartości z określonego lub najlepszego modelu w aktywnym zestawie danych. Można również określić wartość **Nazwa zmiennej niestandardowej**.

Reszty

Zapisz reszty z określonych lub najlepszych predykcji modelu do aktywnego zestawu danych. Można również określić wartość **Nazwa zmiennej niestandardowej**.

Regresja liniowa grzbietowa

Liniowa Grzbiet wykorzystuje klasę Python `sklearn.linear_model.Ridge` do oszacowania L2 lub kwadratu, regulowanych liniowych modeli regresji dla zmiennej zależnej w jednej lub większej liczby zmiennych niezależnych, a także zawiera opcjonalne tryby do wyświetlania wykresów śledzenia oraz do wybierania wartości parametru hiperparameter alfa na podstawie walidacji krzyżowej. Po dopasowaniu pojedynczego modelu lub w celu sprawdzenia, czy do wyboru jest alfa, można użyć partycji danych holdout w celu oszacowania wydajności pojedynczego.

Oprócz dopasowania modelu o określonej wartości parametru regulacji alfa, liniowa grzbiet może wyświetlać wykres śledzenia grzbietu wartości współczynników dla zakresu wartości alfa lub ułatwić wybór wartości nadparametrycznej poprzez interwalidację k-krotnie w określonych siatkach wartości. Jeśli pojedynczy model jest dopasowany lub wybierany jest wybór alfa poprzez walidację krzyżową, ostateczny model może być zastosowany do danych wstrzymanych, które są tworzone przez partycję danych wejściowych w celu uzyskania poprawnego oszacowania wydajności pozapróbowej modelu.

Uzyskiwanie analizy regresji metodą liniowej regresji grzbietowej

1. Z menu wybierz:

Analiza > Regresja > Liniowe alternatywy dla OLS > Grzbiet

Okno dialogowe umożliwia określenie zmiennej, która przypisuje każdą obserwację w aktywnym zbiorze danych do próby szkoleniowej lub holdout.

2. Wybierz liczbową zmienną przewidywaną. Do uruchomienia analizy wymagana jest tylko jedna zmienna przewidywana.

3. Określ wartość liczbową zależną.

4. Określ co najmniej jedną zmienną kategoriową lub zmienną współzmienną liczbową.

Opcjonalnie **Partycja** umożliwia utworzenie holdout lub podzbioru testowego danych wejściowych w celu oszacowania wydajności poza próbką określonego modelu lub wybranego modelu. Wszystkie partycjonowanie wykonywane jest po listowym usunięciu wszystkich obserwacji z niepoprawnymi danymi dla dowolnej zmiennej użytej w procedurze. Należy pamiętać, że w przypadku walidacji krzyżowej dane treningowe są tworzone w języku Python. Dane wstrzymane, które są tworzone przez partycję, nie są używane w estymacji, niezależnie od tego, jaki tryb jest stosowany.

Partycję można zdefiniować, określając współczynnik obserwacji losowo przypisanych do każdej próbki (w sekcji **Partycje szkoleniowe i wstrzymane**) lub przez zmienną, która przypisuje każdą obserwację do ucznia lub próby holdout. Nie można określić zarówno szkolenia, jak i zmiennych. Jeśli partycja nie zostanie określona, zostanie utworzona próbka wstrzymana, która wynosi około 30% danych wejściowych.

Wartość **Szkolenia%** określa względną liczbę obserwacji w aktywnym zbiorze danych, które losowo przypisują do próby ucznia. Szkolenie domyślne to 70%.

Regresja liniowa regresji grzbietowej: Opcje

Na karcie Opcje dostępne są opcje dotyczące:

Dominanta

Ten wybór udostępnia opcje służące do określania jednego z następujących trybów:

Dopasuj do określonej wartości alfa

Po wybraniu tej opcji pojedynczy model jest dopasowany do danych uczących, w których używana jest tylko jedna wartość regulacyjna alfa. Jest to ustawienie domyślne. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W obszarze **Wykres** można wybrać wykresy obserwowanych i/lub reszt względem wartości przewidywanych.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Wykres śladowy

Po wybraniu tej opcji trzy wykresy dla danych uczących są wyświetlane jako funkcja alfa dla określonego zestawu wartości alfa:

- Wykres śledzenia grzbietu współczynników regresji.
- Wykres R^2 .
- Wykres średniej kwadratowej błędów kwadratowego (MSE).

Chociaż partycja jest uhonorowana, nie są udostępniane żadne wyniki dla danych testowych wstrzymanych, ponieważ żaden ostatni model nie wynika z tego trybu.

Wybór alfa poprzez walidację krzyżową

Po wybraniu tej opcji zostanie przeprowadzone wyszukiwanie w siatce z krzyżowym sprawdzaniem poprawności w celu oceny modeli, a wybór najlepszej alfa jest wybierany na podstawie najlepszej

średniej R^2 nad składami sprawdzania poprawności. Pole **Liczba krotności walidacji krzyżowej** może być używane do zmiany domyślnej wartości pięciu podziłów lub fałdów na potrzeby walidacji krzyżowej. Jeśli określona jest partycja, model pojedynczy lub końcowy, który jest dopasowany, jest stosowany do danych testowych wstrzymanych w celu oszacowania wydajności poza próbką.

W sekcji **Wyświetlanie** można wybrać opcję wyświetlania podstawowych informacji o modelu tylko z wybraną wartością alfa (**Najlepszy**), podstawowe informacje o wszystkich modelach w porównaniu (**Porównaj modele**) lub pełne informacje o wszystkich podziałach lub składkach dla wszystkich modeli (**Porównaj modele i złożone**). Wartością domyślną jest **Najlepszy**.

W obszarze **Wykres** można wybrać wykresy średniej R^2 i/lub MSE przy użyciu krotności walidacji. Wykresy obserwowanych i/lub reszt względem wartości przewidywanych mogą być również wybrane.

W sekcji **Zapisz** można określić przewidywane wartości i reszty do zapisania.

Określ pojedyncze wartości alfa

Po wybraniu trybu **Dopasuj do określonego alfa** można określić pojedynczą wartość regulacji alfanumerycznych. Po wybraniu opcji **Wykres śledzenia** lub **Wybór alfa przez walidację krzyżową** można określić wiele wartości.

Wartości

Określ co najmniej jedną dodatnią wartość regularyzacji alfa. Można określić wiele wartości pojedynczo lub jako zakresy. Domyślną wartością jest 1.

Określ siatkę wartości alfa

Po wybraniu trybu **Wykres śledzenia** lub **Wybór alfa w wyniku walidacji krzyżowej** siatka unikalnych wartości alfa może być określona na podstawie wartości **Początek** (value1) do wartości **Koniec** (value2) z przyrostem **Wg** (value3). Jeśli zostanie podany, dozwolony jest tylko jeden poprawny zestaw [value1 TO value2 BY value3]. Musi on spełniać następujące warunki: $0 \leq \text{value1} \leq \text{value2} \leq 1$. W przypadku, gdy produkt $\text{value1} = \text{value2}$ jest równoważny z określeniem pojedynczego value1, bez względu na value3.

Metryka Alfa dla zakresów wartości może mieć wartość **Liniowa** lub **Podstawowa 10 logarytmiczna** (10 jest podniesiona do potęgi określonych wartości).

Wykresy są wyświetlane przy użyciu określonej wielkości mierzonej dla poziomych osi X zmiennych wartości alfa.

Kryteria

Steruje analizami.

Uwzględnij wyraz wolny

Kryterium to obejmuje wyraz wolny w jednym lub kilku modelach. Należy zauważyć, że procedura rozszerzenia nie jest w centrum ani nie standaryzuje zmiennej zależnej, a wyraz wolny nie jest karany w trakcie oszacowania.

Standaryzacja predyktorów

To kryterium standaryzuje wszystkie zmienne niezależne.

Liczba krotności walidacji krzyżowej

To kryterium umożliwia ustawienie liczby podziały lub fałd dla oceny walidacji krzyżowej modeli. Liczba musi być dodatnią liczbą całkowitą większą niż 1. Ustawienie domyślne to 5.

Stan losowy Python

Wartość ustawienia random_state w języku Python jest używana podczas wykonywania wartościowania krzyżowego modeli. Umożliwia replikację wyników, które obejmują pseudolosowe liczby. Musi to być liczba całkowita z zakresu od 0 do $2^{32}-1$. Domyślną wartością jest 0.

Limit czasu (minuty)

Liczba minut dozwolona dla obliczeń modelu, które mają zostać uruchomione. Jeśli zostanie podana wartość 0, licznik czasu zostanie wyłączonej. Wartość domyślna to 5.

Wyświetlanie podpowiedzi

Ta sekcja określa ilość danych wyjściowych do wyświetlenia dla trybu **Wybór alfa w drodze walidacji krzyżowej**.

Najlepsza

Wyświetlane są tylko podstawowe wyniki dla wybranego najlepszego modelu. Jest to ustawienie domyślne.

Porównywanie modeli

Wyświetlane są również podstawowe wyniki dla wszystkich ocenionych modeli.

Porównywanie modeli i krotności

Na koniec wyświetlane są pełne szczegółowe wyniki dla każdego podziału lub krotności dla każdego ocenianego modelu.

Wykres

Określa wykresy wartości obserwowanych lub rezydujących w porównaniu z wartościami przewidywanymi, ze sprawdzaniem krzyżowym, specyfikacją wykresów średniej kwadratowej (MSE) i/lub średnią R^2 w przekrojowych składkach walidacyjnych a wartości alfa.

Średni czas sprawdzania krzyżowego błędu kwadratowego (MSE) w porównaniu z alfa

W trybie **Wybór alfa przez walidację krzyżową** wyświetlany jest wykres liniowy średniej liczby żądań MSE względem walidacji krzyżowej w porównaniu z alfa. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Średnia krzyżowa walidacja R^2 kwadrat a alfa

W trybie **Wybór alfa przez walidację krzyżową** wyświetlany jest wykres liniowy średniej liczby R^2 dla krotności walidacji krzyżowej w porównaniu z alfa. Dla trybu **Wykres śledzenia** podobna działka jest automatycznie tworzona w oparciu o pełne dane uczących.

Obserwowany a przewidywany

Wyświetla wykres rozrzutu obserwowanych i przewidywanych wartości dla określonego lub najlepszego modelu.

Reszty względem przewidywanego

Wyświetlana jest również wykres rozrzutu reszt w porównaniu z wartościami przewidywanymi dla określonego lub najlepszego modelu.

Zapisz

Określa zmienne, które mają zostać zapisane w aktywnym zestawie danych.

Wartości przewidywane

Zapisz przewidywane wartości z określonego lub najlepszego modelu w aktywnym zestawie danych. W przeciwnym razie można podać wartość **Nazwa zmiennej niestandardowej**.

Reszty

Zapisz reszty z określonych lub najlepszych predykcji modelu do aktywnego zestawu danych. W przeciwnym razie można podać wartość **Nazwa zmiennej niestandardowej**.

Estymacja krzywej

Procedura estymacji krzywej pozwala obliczyć statystykę regresyjną dla estymacji krzywej oraz uzyskać odpowiednie wykresy z użyciem 11 różnych modeli regresji. Dla każdej zmiennej zależnej tworzony jest odrębny model. Wartości przewidywane, reszty oraz przedziały predykcji można również zapisywać jako nowe zmienne.

Przykład. Dostawca usług internetowej wykrywa procent zainfekowanej wirusem poczty e-mail w swojej sieci na przestrzeni czasu. Wykres rozrzutu pozwala stwierdzić istnienie zależności liniowej. Dysponując tą informacją, można dopasować do danych odpowiedni model liniowy lub sześcienny i sprawdzić prawdziwość założeń oraz dobroć dopasowania modelu.

Statystyki. W przypadku każdego modelu: współczynniki regresji, wielorakie R , R^2 skorygowane R^2 , błąd standardowy oszacowania, tabela analizy wariancji, wartości przewidywane, reszty oraz przedziały predykcji. Modele: liniowy, logarytmiczny, odwrotny, kwadratowy, sześcienny, potęgowy, złożony, krzywej S, logistyczny, wzrostu i wykładniczy.

Wymagania dotyczące danych przy estymacji krzywej

Dane. Zmienne zależne i niezależne powinny być zmiennymi ilościowymi. Jeżeli jako zmienna niezależna zamiast zmiennej z aktywnego zbioru danych wybrana zostanie opcja **Czas**, to procedura Estymacja krzywej wygeneruje zmienną czasu z jednolitymi przedziałami czasu między obserwacjami. Przy wybranej opcji **Czas** zmienna zależna powinna być miarą szeregu czasowego. Analiza szeregów czasowych wymaga określonej struktury pliku danych, w której każda obserwacja (wiersz) odpowiada zbiorowi obserwacji dokonanych w różnych momentach czasu, przy czym przedziały czasu między pomiarami są jednakowe.

Założenia. Dane warto wyświetlić w postaci graficznej, by ustalić typ relacji między zmiennymi zależnymi i niezależnymi (liniowa, wykładnicza itp.). Reszty w przypadku dobrego modelu powinny być równomiernie rozmieszczone i zgodne z rozkładem normalnym. Jeśli używany jest model liniowy, spełnione muszą być następujące założenia: dla każdej wartości zmiennej niezależnej rozkład zmiennej zależnej musi być normalny. Wariancja rozkładu zmiennej zależnej powinna być stała dla wszystkich wartości zmiennej niezależnej. Relacja między zmienną zależną a zmienną niezależną powinna być liniowa, a wszystkie obserwacje powinny być niezależne.

Wykonywanie estymacji krzywej

1. Wybierz z menu następującą opcję:

Analiza > Regresja > Estymacja krzywej...

2. Wybierz co najmniej jedną zmienną zależną. Dla każdej zmiennej zależnej tworzony jest odrębny model.

3. Wybierz zmienną niezależną (zmienną w aktywnym zbiorze danych lub opcję **Czas**).

4. Opcjonalnie można wykonać następujące działania:

- Wybierz zmienną umieszczaną w etykietach obserwacji na wykresach rozrzutu. Dla każdego punktu na wykresie rozrzutu można użyć narzędzia wyboru punktów, aby wyświetlić wartość zmiennej Opis obserwacji.
- Kliknij przycisk **Zapisz**, aby zapisać wartości przewidywane, reszty i przedziały predykcji jako nowe zmienne.

Dostępne są także następujące opcje:

- **Uwzględnij stałą w równaniu.** Powoduje oszacowanie składnika stałego w równaniu regresji. Stała jest uwzględniana domyślnie.
- **Graficzna prezentacja modeli.** Sporządza wykres wartości zmiennej zależnej i każdego wybranego modelu względem zmiennej niezależnej. Dla każdej zmiennej zależnej tworzony jest oddzielny wykres.
- **Wyświetl tabelę ANOVA.** Wyświetla tabelę podsumowującą analizę wariancji dla każdego wybranego modelu.

Modele estymacji krzywej

Można wybrać jeden lub kilka różnych modeli regresyjnej estymacji krzywej. Aby ustalić, który model będzie najodpowiedniejszy, należy sporządzić wykres danych. Jeśli zmienne są ułożone w sposób zbliżony do linii prostej, należy użyć modelu prostej regresji liniowej. Jeśli zmienne nie są związane liniowo, można spróbować je przekształcić. W przypadku gdy przekształcenie nie daje rezultatu, można zastosować bardziej złożony model. Należy przyjrzeć się wykresowi rozrzutu danych. Jeśli układ punktów przypomina znaną funkcję matematyczną, będzie to wskazówką przy doborze odpowiedniego modelu. Na przykład jeśli punkty ułożone są podobnie do krzywej wykładniczej, należy użyć modelu wykładniczego.

Liniowa. Model, którego równanie to $Y = b_0 + (b_1 * t)$. Wartości szeregu są modelowane jako liniowa funkcja czasu.

Logarytmiczna. Model opisany równaniem: $Y = b_0 + (b_1 * \ln(t))$.

Odwrótność. Model opisany równaniem: $Y = b_0 + (b_1 / t)$.

Kwadratowa. Model, którego równanie to $Y = b_0 + (b_1 * t) + (b_2 * t^{**2})$. Model kwadratowy może być użyty do modelowania szeregu, który „rozkręca się” lub do szeregu wygasającego.

Sześcienny. Model opisany równaniem: $Y=b_0+(b_1*t)+(b_2*t^{**2})+(b_3*t^{**3})$.

Zasilanie. Model opisany równaniem: $Y=b_0*(t^{**b_1})$ lub $\ln(Y)=\ln(b_0)+(b_1*\ln(t))$.

Złożone. Model o równaniu $Y = b_0 * (b_1^{**t})$ lub $\ln(Y) = \ln(b_0) + (\ln(b_1) * t)$.

Krzywa S. Model opisany równaniem: $Y=e^{**}(b_0+(b_1/t))$ lub $\ln(Y)=b_0+(b_1/t)$.

Logistyczna. Model, którego równanie to $Y = 1/(1/u + (b_0 * (b_1^{**t})))$ lub $\ln(1/y-1/u) = \ln(b_0) + (\ln(b_1) * t)$, gdzie u jest górną granicą wartości. Po wybraniu opcji Logistyczny należy określić wartość górnej granicy, która zostanie użyta w równaniu regresji. Wartość ta musi być liczbą dodatnią i musi być większa od największej wartości zmiennej zależnej.

Wzrost. Model opisany równaniem: $Y = e^{**}(b_0+(b_1*t))$ lub $\ln(Y) = b_0+(b_1*t)$.

Wykładnicza. Model opisany równaniem: $Y=b_0*(e^{**}(b_1* t))$ lub $\ln(Y)=\ln(b_0)+(b_1*t)$.

Estymacja krzywej: Zapisz

Zapisz zmienne. Każdy wybrany model pozwala zapisać wartości przewidywane, reszty (obserwowana wartość zmiennej zależnej pomniejszona o wartość przewidywaną przez model) oraz przedziały predykcji (górną i dolną granicę). Nowe nazwy zmiennych wraz z opisami są wyświetlane w tabeli w oknie wynikowym.

Prognozuj obserwacje. Jeśli jako zmienna niezależna zamiast zmiennej w aktywnym zbiorze danych wybrana zostanie opcja **Czas**, możliwe jest podanie okresu prognozy, którego długość wykracza poza granicę szeregu czasowego. Można wybrać jedną z poniższych alternatyw:

- **Do ostatniej na podstawie okresu estymacji.** Prognozuje wartości dla wszystkich obserwacji w zbiorze w oparciu o okres estymacji. Okres estymacji, wyświetlany w dolnej części okna dialogowego, jest definiowany za pomocą sekcji Zakres w oknie Wybierz obserwacje w menu Dane. Jeśli nie został zdefiniowany żaden okres estymacji, wszystkie obserwacje będą wykorzystane do prognozowania wartości.
- **Prognozuj do.** Powoduje prognozowanie wartości do określonej daty, godziny lub numeru obserwacji włącznie, bazując na obserwacjach w okresie estymacji. Możliwe jest też wykorzystanie tej opcji do prognozowania wartości przekraczających ostatnią obserwację w szeregu czasowym. Aktualnie zdefiniowane zamiennie danych określają, które pola tekstowe są dostępne do określenia końca czasu predykcji. Jeśli nie ma zdefiniowanych zmiennych typu data, to można podać numer ostatniej obserwacji.

Aby zdefiniować zmienne typu data, należy użyć opcji Definiuj datę i czas w menu Dane.

Regresja metodą cząstkowych najmniejszych kwadratów

Procedura regresji metodą cząstkowych najmniejszych kwadratów szacuje modele regresji metodą cząstkowych najmniejszych kwadratów (PLS, zwaną także „odwzorowaniem na strukturę utajoną”). PLS jest techniką prognostyczną, stanowiącą zamiennik regresji metodą zwykłych najmniejszych kwadratów (OLS), korelacji kanonicznej lub modelowania równań strukturalnych, i jest szczególnie użyteczna, gdy predyktory są wysoko skorelowane lub kiedy liczba przeliczników jest większa niż liczba obserwacji.

Metoda cząstkowych najmniejszych kwadratów łączy cechy analizy składowych głównych i regresji wielokrotnej. W pierwszej kolejności wyodrębnia zbiór czynników utajonych, które wyjaśniają jak największą kowariancję między zmiennymi niezależnymi i zależnymi. Następnie, w fazie regresji, prognozuje się wartości zmiennych zależnych, używając dekompozycji zmiennych niezależnych.

Tabele

Udział wariancji wyjaśnionej (według czynnika utajonego), wagi czynników utajonych, ładunki czynników utajonych, ważność zmiennej niezależnej w odwzorowaniu (VIP) oraz oszacowania parametrów regresji (według zmiennej zależnej) są generowane domyślnie.

Wykresy

Ważność zmiennych w odwzorowaniu (VIP), oceny czynnikowe, wagi czynnikowe dla pierwszych trzech czynników utajonych oraz odległość od modelu są generowane za pomocą karty Opcje.

Wymagania dotyczące danych

Poziom pomiaru

Zmienne zależne i niezależne (predyktory) mogą mieć charakter nominalny, porządkowy lub ilościowy. W procedurze przyjmuje się, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych, choć można tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego. Zmienne jakościowe (nominalne lub porządkowe) są traktowane przez procedurę równoważnie.

Kodowanie zmiennych jakościowych

Procedura tymczasowo przekodowuje zależne zmienne jakościowe, używając jednego z c kodowań na czas trwania procedury. Jeśli istnieje c kategorii zmiennej, zmienna jest przechowywana jako c wektorów, z pierwszą kategorią oznaczoną $(1, 0, \dots, 0)$, następną kategorią $(0, 1, 0, \dots, 0)$, ..., i kategorii końcowej $(0, 0, \dots, 0, 1)$. Zależne zmienne jakościowe są przedstawiane przy użyciu sztucznego kodowania, tzn. z pominięciem wskaźnika odpowiadającego kategorii odniesienia.

Wagi liczebności

Wartości wag są przed użyciem zaokrąglane do najbliższej liczby całkowitej. Obserwacje, w których brakuje wag lub które mają wagi mniejsze niż 0,5, nie są wykorzystywane w analizach.

Brakujące dane

Zdefiniowane i systemowe braki danych są traktowane jako wartości nieprawidłowe.

Przeskalowanie

Wszystkie zmienne modelu są centrowane i standaryzowane; dotyczy to także zmiennych wskaźnikowych reprezentujących zmienne kategoriaalne.

Uzyskiwanie regresji metodą cząstkowych najmniejszych kwadratów

Z menu wybierz:

Analiza > Regresja > Cząstkowe najmniejsze kwadraty...

1. Wybierz co najmniej jedną zmienną zależną.
2. Wybierz co najmniej jedną zmienną niezależną.

Opcjonalnie można wykonać następujące czynności:

- Określ kategorię odniesienia dla jakościowych (nominalnych lub porządkowych) zmiennych zależnych.
- Określ zmienną, która ma być używana jako unikalny identyfikator wyników obserwacji i zapisanych zbiorów danych.
- Określ górny limit liczby wyodrębnianych czynników utajonych.

Wymagania wstępne

Procedura regresji metodą cząstkowych najmniejszych kwadratów jest komendą rozszerzenia Python i wymaga funkcjonalności Python, która jest częścią produktu IBM SPSS Statistics. Wymagane są również ogólnodostępne biblioteki NumPy oraz SciPy Python.

Uwaga: Użytkownicy, którzy będą pracowali w trybie analizy rozproszonej (wymagana wersja IBM SPSS Statistics Server), muszą mieć zainstalowane biblioteki NumPy i SciPy na serwerze. Dodatkowe wsparcie można uzyskać u administratora systemu.

Użytkownicy systemów Windows i Mac

W przypadku systemów Windows i Mac biblioteki NumPy i SciPy muszą być zainstalowane dla osobnej wersji środowiska Python 3.10 niż ta, która została zainstalowana wraz z produktem IBM SPSS Statistics. Jeśli dodatkowa wersja środowiska Python 3.10 nie jest zainstalowana, można ją

pobrać ze strony <http://www.python.org>. Następnie należy zainstalować biblioteki NumPy oraz SciPy dla środowiska Python, wersja 3.10. Pliki instalacyjne są dostępne na stronie <http://www.scipy.org/Download>.

Aby aktywować biblioteki NumPy i SciPy, należy ustawić lokalizację środowiska Python na wersję środowiska Python 3.10, w której zainstalowano NumPy i SciPy. Lokalizację środowiska Python można ustawić na karcie Lokalizacje plików w oknie dialogowym Opcje (Edycja > Opcje).

Użytkownicy systemu Linux

Zalecamy pobranie źródła i samodzielne utworzenie bibliotek NumPy i SciPy. Źródło jest dostępne na stronie <http://www.scipy.org/Download>. Biblioteki NumPy i SciPy można zainstalować w wersji środowiska Python 3.10 zainstalowanej wraz z produktem IBM SPSS Statistics. Znajduje się on w katalogu Python w miejscu, w którym zainstalowano produkt IBM SPSS Statistics .

Jeśli biblioteki NumPy i SciPy mają zostać zainstalowane wraz z wersją środowiska Python 3.10 inną niż zainstalowana z produktem IBM SPSS Statistics, wówczas należy skonfigurować lokalizację Python, tak aby wskazywała wybraną wersję. Lokalizację środowiska Python można ustawić na karcie Lokalizacje plików w oknie dialogowym Opcje (**Edycja > Opcje**).

Serwer Windows i Unix

Biblioteki NumPy i SciPy muszą być zainstalowane na serwerze dla osobnej wersji środowiska Python 3.10 niż ta, która została zainstalowana wraz z produktem IBM SPSS Statistics. Jeśli na serwerze nie zainstalowano osobnej wersji środowiska Python 3.10, można ją pobrać ze strony <http://www.python.org>. Biblioteki NumPy i SciPy dla środowiska Python 3.10 są dostępne na stronie <http://www.scipy.org/Download>. Aby aktywować biblioteki NumPy i SciPy, lokalizację środowiska Python dla serwera należy ustawić na wersję środowiska Python 3.10, w której zainstalowano biblioteki NumPy i SciPy. Lokalizację środowiska Python można ustawić za pośrednictwem repozytorium konsola administracyjna programu IBM SPSS StatisticsK.

Model

Określenie efektów modelu. Model efektów głównych zawiera wszystkie efekty główne czynników i współzmiennych. Wybierz pozycję **Użytkownika**, aby określić interakcje. Należy określić wszystkie składniki modelu.

Czynniki i współzmiennie. Lista zawiera czynniki i współzmiennie.

Model. Model zależy od charakteru danych. Po wybraniu opcji **Użytkownika** można wybrać efekty główne oraz interakcje będące przedmiotem zainteresowania w czasie analizy.

Budowanie czynników

Dla wybranych czynników i współzmiennych:

Interakcja. Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu. Jest to wartość domyślna.

Efekty główne. Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

Wszystkie 2 rzędu. Tworzy wszystkie możliwe interakcje drugiego rzędu wybranych zmiennych.

Wszystkie 3 rzędu. Tworzy wszystkie możliwe interakcje trzeciego rzędu wybranych zmiennych.

Wszystkie 4 rzędu. Tworzy wszystkie możliwe interakcje czwartego rzędu wybranych zmiennych.

Wszystkie 5 rzędu. Tworzy wszystkie możliwe interakcje piątego rzędu wybranych zmiennych.

Opcje

Karta Opcje pozwala użytkownikowi zapisać i wykreślić oszacowania modelu dla poszczególnych przypadków, czynników utajonych i przeliczników.

Dla każdego typu danych podaj nazwę zbioru danych. Nazwy zbiorów danych muszą być niepowtarzalne. W razie podania nazwy istniejącego zbioru danych jego zawartość zostaje zastąpiona; w przeciwnym razie tworzony jest nowy zbiór danych.

- **Zapisz oszacowania dla poszczególnych obserwacji.** Zapisywanie następujących oszacowań modelu dotyczących danego przypadku: prognozowane wartości, reszty, odległość od modelu czynników utajonych oraz statystyki czynników utajonych. Zostaną również narysowane wyniki czynników ukrytych.
- **Zapisz oszacowania czynników ukrytych.** Zapisywanie ładunków i wag czynników utajonych. Pozwala także wykreślić wagi czynników utajonych.
- **Zapisz oszacowania dla zmiennych niezależnych.** Zapisywanie oszacowań parametrów regresji i ważności zmiennych dla odwzorowania (VIP). Zostaną również narysowane wartości VIP wg czynnika ukrytego.

Analiza najbliższego sąsiedztwa

Analiza najbliższego sąsiedztwa jest metodą klasyfikacji obserwacji na podstawie ich podobieństwa do innych obserwacji. W uczeniu maszynowym procedura ta została opracowana jako sposób na rozpoznawanie układów danych bez konieczności idealnego dopasowania ich do przechowywanych układów lub obserwacji. Podobne obserwacje znajdują się blisko siebie, a niepodobne – daleko. Zatem odległość między dwoma obserwacjami stanowi miarę ich niepodobieństwa.

Obserwacje znajdujące się blisko siebie nazywają się „sąsiedztwem”. Podczas prezentacji nowej (wstrzymanej) obserwacji obliczana jest odległość od każdej obserwacji modelu. Zostaje określona klasyfikacja najbardziej podobnych obserwacji najbliższego sąsiedztwa, a nowa obserwacja zostaje umieszczona w kategorii, która zawiera największą liczbę obserwacji najbliższego sąsiedztwa.

Można określić liczbę najbliższych elementów sąsiednich do analizowania; ta wartość to k .

Analiza najbliższego sąsiedztwa może być również użyta do obliczania docelowych wartości ilościowych. W tej sytuacji do uzyskania przewidywanej wartości dla nowej obserwacji stosowana jest docelowa wartość średniej lub mediany najbliższych sąsiadów.

Wymagania dotyczące danych do analizy najbliższego sąsiedztwa












Wartości docelowe i funkcje. Wartości docelowe i funkcje mogą być:

- *Nominalny.* Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rankingu (na przykład dział przedsiębiorstwa, w którym pracuje pracownik). Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy.* Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorie z jakimś nieodłącznym rangą (na przykład poziomy zadowolenia z usługi z bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.
- *scale.* Zmienna może być traktowana jako skala (ilościowa), gdy jej wartości reprezentują uporządkowane kategorie z miarodajnym pomiarem, tak aby porównania odległości między wartościami były odpowiednie. Przykładami zmiennych ilościowych mogą być: wiek w latach lub przychód w tysiącach złotych.

Zmienne nominalne i porządkowe są traktowane tak samo w analizie najbliższego sąsiedztwa. W procedurze przyjmuje się, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych, można jednak tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego.

Ikona obok każdej zmiennej na liście zmiennych określa poziom pomiaru oraz typ danych:

Tabela 1. Ikony poziomu pomiaru

	Liczbowy	Łańcuch	Data	Czas
Zmienna (ilościowa)		nie dotyczy		
Porządkowy				
Nominalna				

Kodowanie zmiennych jakościowych. Procedura tymczasowo przekodowuje predyktory jakościowe i zmienne zależne, używając jednego z c kodowań na czas trwania procedury. Jeśli istnieje c kategorii zmiennej, to zmienna jest przechowywana jako c wektorów, gdzie pierwszą kategorię oznacza $(1,0,\dots,0)$, następną $(0,1,0,\dots,0)$, ..., a ostatnią $(0,0,\dots,0,1)$.

Taki schemat kodowania zwiększa wymiarowość przestrzeni właściwości. Całkowita liczba wymiarów jest liczbą predyktorów ilościowych plus liczbą kategorii we wszystkich predyktorach jakościowych. Taki schemat kodowania może spowodować wolniejszą naukę. Jeżeli szkolenie najbliższego sąsiedztwa przebiega bardzo wolno, przed uruchomieniem procedury można próbować ograniczyć liczbę kategorii predyktorów jakościowych przez połączenie podobnych kategorii lub rezygnację z obserwacji, które posiadają wyjątkowo rzadkie kategorie.

Kodowanie jeden z c jest oparte na danych szkoleniowych, nawet jeśli zdefiniowano próbę wstrzymania (patrz "Podziały" na stronie 172). Dlatego jeśli próba wstrzymana zawiera obserwacje z kategoriami predyktorów, które nie są obecne w danych szkoleniowych, obserwacje te nie są oceniane. Jeżeli próba wstrzymana zawiera obserwacje z kategoriami zmiennych zależnych, które nie są obecne w danych szkoleniowych, obserwacje te są oceniane.

Przeskalowanie. Funkcje skali są domyślnie znormalizowane. Przeskalowanie jest wykonywane na podstawie danych szkoleniowych, nawet jeśli zdefiniowano próbę wstrzymania (patrz temat "Podziały" na stronie 172). Jeżeli zostanie określona zmienna definiująca podziały, ważne jest, aby funkcja miała podobną dystrybucję w próbie szkoleniowej i próbie wstrzymanej. Należy użyć na przykład procedury Eksploracja, aby zbadać dystrybucję w podziałach.

Wagi liczebności. Wagi częstotliwości są ignorowane przez tę procedurę.

Replikacja wyników. Procedura używa generatora liczb losowych przy losowym przydziale podziałów i walidacji krzyżowej. Jeżeli wyniki mają być powielone dokładnie, oprócz użycia tych samych ustawień procedury należy ustawić wartość startową generatora Mersenne Twister (patrz "Podziały" na stronie 172) lub użyć zmiennych, aby zdefiniować podziały i walidację krzyżową.

W celu uzyskania analizy najbliższego sąsiedztwa

Z menu wybierz:

Analiza > Klasyfikacja > Najbliższe sąsiedztwo...

1. Określ jedną lub kilka funkcji, które mogą być traktowane jako zmienne niezależne lub predyktory, jeśli istnieje wartość docelowa.

Wartość docelowa (opcjonalnie). Jeżeli nie określono wartości docelowej (zmiennej zależnej lub odpowiedzi), procedura znajduje k obserwacji w najbliższym sąsiedztwie i nie jest przeprowadzana klasyfikacja ani predykcja.

Normalizuj zmienne ilościowe. Znormalizowane funkcje mają ten sam zakres wartości, co może poprawić wydajność algorytmu estymacji. Używana jest normalizacja skorygowana $[2*(x-\min)/(\max-\min)]-1$. Skorygowane wartości znormalizowane zawierają się w zakresie od -1 do 1.

Identyfikator obserwacji kluczowych (opcjonalnie). Umożliwia to oznaczenie obserwacji o szczególnym znaczeniu. Przykładowo: prowadzący badanie chce określić, czy wyniki testów

w szkołach z jednego regionu (obserwacja kluczowa) są porównywalne z wynikami w podobnych regionach. Analiza najbliższego sąsiedztwa jest wykorzystywana do znalezienia regionów, które są najbardziej podobne pod względem podanych funkcji. Następnie prowadzący badanie porównuje wyniki testów z regionu centralnego z tymi z najbliższego sąsiedztwa.

Obserwacje kluczowe mogą być również używane w badaniach klinicznych, aby wybrać obserwacje kontrolne, podobne do obserwacji klinicznych. Obserwacje kluczowe są wyświetlane w tabeli k obserwacji najbliższego sąsiedztwa i odległości, na wykresie przestrzeni właściwości, wykresie elementów równorzędnych oraz mapie kwadratowej. Informacje dotyczące obserwacji kluczowych są zapisywane w plikach określonych na karcie Wyniki.

Obserwacje z wartością dodatnią określonej zmiennej są traktowane jako obserwacje kluczowe. Nieprawidłowe jest określenie zmiennej bez wartości dodatnich.

Etykieta obserwacji (opcjonalnie). Obserwacje są opisywane za pomocą wartości na wykresie przestrzeni właściwości, wykresie elementów równorzędnych oraz mapie kwadratowej.

Zmienne z nieznanym poziomem pomiaru

Alert poziomemu pomiaru wyświetla się, gdy poziom pomiaru dla jednej lub większej ilości zmiennych w zbiorze danych jest nieznan. Ponieważ poziom pomiaru wpływa na wyliczenie wyników dla tej procedury, wszystkie zmienne muszą mieć zdefiniowany poziom pomiaru.

Skanowanie danych. Odczytuje dane w aktywnym zbiorze danych i przypisuje domyślny poziom pomiaru do wszystkich zmiennych, które mają aktualnie nieznaną poziom pomiaru. Jeśli zbiór danych jest duży, może to zająć trochę czasu.

Przypisz ręcznie. Otwiera okno dialogowe, które zestawia wszystkie zmienne z nieznanym poziomem pomiaru. Można użyć tego okna dialogowego do przypisania poziomu pomiaru do tych zmiennych. Można również przypisać poziom pomiaru w Widoku zmiennych Edytora danych.

Ponieważ poziom pomiaru jest ważny dla tej procedury, nie można wejść do tego okna dialogowego w celu uruchomienia tej procedury, dopóki wszystkie zmienne nie będą miały zdefiniowanego poziomu pomiaru.

Sąsiedztwo

Liczba najbliższych sąsiadów (k). Określ liczbę obserwacji najbliższego sąsiedztwa. Należy pamiętać, że większa liczba obserwacji najbliższego sąsiedztwa nie zawsze oznacza dokładniejszy model.

Jeżeli wartość docelowa jest określona na karcie Zmienne, można alternatywnie określić zakres wartości i pozwolić procedurze na wybranie najlepszej liczby obserwacji najbliższego sąsiedztwa w tym zakresie. Metoda określenia liczby obserwacji najbliższego sąsiedztwa zależy od tego, czy zaznaczono wybór funkcji na karcie Funkcje.

- Jeżeli wybór funkcji jest włączony, wybór funkcji jest wykonywany dla każdej wartości k we wprowadzonym zakresie i zostaje wybrane k oraz towarzyszący zestaw funkcji z najniższym poziomem błędów (lub najniższym błędem sumy kwadratów, jeśli wartość docelowa jest ilościowa).
- Jeżeli wybór funkcji nie jest włączony, V -krotna walidacja krzyżowa jest używana w celu wybrania najlepszej liczby obserwacji najbliższego sąsiedztwa. Na karcie Podział znajdują się kontrolki przydziału krotności.

Obliczanie odległości. Jest to miara używana do określenia metryki odległości używanej do pomiaru podobieństwa obserwacji.

- **Metryka euklidesowa.** Odległość pomiędzy dwiema obserwacjami, x i y , jest pierwiastkiem kwadratowym sumy, we wszystkich wymiarach, różnic pomiędzy wartościami obserwacji podniesionymi do kwadratu.
- **Metryka miejska.** Odległość pomiędzy dwiema obserwacjami jest sumą, we wszystkich wymiarach, bezwzględnych różnic pomiędzy wartościami tych obserwacji. Miara ta jest również nazywana odległością manhattańską.

Opcjonalnie jeśli wartość docelowa jest określona na karcie Zmienne, można wybrać funkcje wagi według znormalizowanej ważności podczas obliczania odległości. Ważność właściwości predyktora jest obliczana

jako współczynnik poziomu błędu lub błąd sumy kwadratów modelu z predyktorem usuniętym z modelu do poziomu błędu lub błędu sumy kwadratów pełnego modelu. Znormalizowana ważność jest obliczana przez zmianę wag wartości ważności właściwości, tak aby dawały w sumie 1.

Predykcja dla ilościowej zmiennej zależnej. Jeżeli docelowa wartość ilościowa jest określona na karcie Zmienne, określa to, czy przewidywane wartości są obliczane na podstawie średniej lub mediany wartości obserwacji najbliższego sąsiedztwa.

Predyktory

Karta Funkcje umożliwia wykorzystanie i określenie opcji wyboru funkcji, jeśli wartość docelowa jest określona na karcie Zmienne. Domyślnie wszystkie funkcje są uwzględniane przy wyborze funkcji, ale można wybrać zestaw funkcji, które zostaną wymuszone w modelu.

Kryterium zatrzymywania. Na każdym etapie funkcja, której dodanie do modelu powoduje najmniejszy błąd (obliczony jako poziom błędu jakościowej zmiennej docelowej i błąd sumy kwadratów ilościowej zmiennej docelowej), jest uwzględniana w dodaniu do zestawu modelu. Selekcja postępująca jest kontynuowana do osiągnięcia określonego warunku.

- **Określona liczba funkcji.** Algorytm dodaje stałą liczbę funkcji oprócz tych, które są wymuszone w modelu. Podaj dodatnią liczbę całkowitą. Zmniejszenie wartości liczby do wyboru tworzy skromniejszy model, z ryzykiem pominięcia istotnych funkcji. Zwiększenie wartości liczby do wyboru spowoduje ujęcie wszystkich istotnych funkcji, z ryzykiem dodania funkcji, które w rzeczywistości zwiększają błąd modelu.
- **Minimalna zmiana bezwzględne współczynnika błędu.** Algorytm zostaje zatrzymany, gdy zmiana bezwzględne współczynnika błędu wskazuje, że model nie może być dalej udoskonalony przez dodanie dalszych funkcji. Określ liczbę dodatnią. Zmniejszenie wartości minimalnej zmiany spowoduje uwzględnienie większej liczby funkcji przy ryzyku zawarcia funkcji, które nie mają dużej wartości w danym modelu. Zwiększenie wartości minimalnej zmiany spowoduje wyłączenie większej ilości funkcji przy ryzyku utraty funkcji, które są istotne w modelu. Optymalna wartość minimalnej zmiany zależy od danych i ich zastosowania. Dziennik błędów wyboru funkcji wyników pomaga ocenić, które funkcje są najbardziej istotne. Więcej informacji można znaleźć w temacie [“Dziennik błędów wyboru funkcji”](#) na stronie 176.

Podziały

Karta Podziały umożliwia podzielenie zbioru danych na zestaw szkoleniowy i wstrzymany oraz, jeśli jest to możliwe, pozwala na przydzielenie obserwacji do walidacji krzyżowej

Podzbiór uczący i walidacyjny. Ta grupa określa metodę podziału aktywnego zbioru danych na próby uczące i wstrzymane. **Próba ucząca** składa się z rekordów danych używanych do szkolenia modelu najbliższego sąsiedztwa. Pewna wartość procentowa obserwacji w zbiorze danych musi być przypisana do próby uczącej w celu uzyskania modelu. **Próba wstrzymana** jest niezależnym zestawem rekordów danych używanych do oceny modelu końcowego. Błąd w próbie wstrzymanej daje „uczciwą” ocenę możliwości predykcyjnej modelu, ponieważ obserwacje wstrzymane nie zostały użyte do budowy modelu.

- **Przydziel losowo obserwacje do podzbiorów.** Określ udział procentowy obserwacji przydzielanych do próby uczącej. Reszta zostaje przydzielona do próby wstrzymanej.
- **Użyj zmiennej do podziału obserwacji na grupy.** Określ zmienną numeryczną, która przydziela każdą obserwację w aktywnym zbiorze danych do próby szkoleniowej lub wstrzymanej. Obserwacje z wartością dodatnią są przydzielane do próby szkoleniowej, a przypadki z wartością 0 lub mniejszą – do próby wstrzymanej. Obserwacje z systemowymi brakami danych zostają wykluczone z analizy. Wartości zdefiniowanych przez użytkownika braków danych zmiennej dzielącej zawsze są traktowane jako prawidłowe.

Krotności walidacji krzyżowej. V-krotność walidacji krzyżowej jest używana do określenia „najlepszej” liczby obserwacji najbliższego sąsiedztwa. Nie jest dostępna w połączeniu z wyborem funkcji z powodów wydajności.

Walidacja krzyżowa dzieli próbę na kilka podprób (krotności). Następnie generowane są modele najbliższego sąsiedztwa, wyłączając kolejno dane z każdej podpróby. Pierwszy model jest oparty na

wszystkich obserwacjach z wyjątkiem tych w pierwszej krotności próby; drugi model jest oparty na wszystkich obserwacjach z wyjątkiem drugiej krotności próby itd. Dla każdego modelu szacowany jest błąd z zastosowaniem modelu na podpróbie wyłączanej podczas generowania modelu. „Najlepsza” liczba obserwacji najbliższego sąsiedztwa to ta, która powoduje najniższy błąd we wszystkich krotnościach.

- **Losowo przydziel obserwacje do krotności.** Określ liczbę krotności, które powinny być użyte do walidacji krzyżowej. Procedura losowo przypisuje obserwacje do krotności, ponumerowanych od 1 do V , liczby krotności.
- **Użyj zmiennej do podziału obserwacji na grupy.** Określ zmienną numeryczną, która przydziela każdą obserwację w aktywnym zbiorze danych do krotności. Zmienna musi być numeryczna i musi przyjmować wartości od 1 do V . Jeśli brakuje jakichkolwiek wartości w tym zakresie, a w przypadku wszystkich podziałów, jeśli są one aktywne, spowoduje to błąd. .

Wartość startowa generatora Mersenne Twister. Ustawienie wartości startowej umożliwia powielenie analizy. Używanie tego elementu jest podobne do ustawiania generatora Mersenne Twister jako aktywnego generatora oraz określenia stałego punktu startowego w oknie dialogowym Generator liczb losowych. Przy czym istnieje tu istotna różnica polegająca na tym, że ustawienie wartości startowej w tym oknie dialogowym zachowa bieżący stan generatora liczb pseudolosowych i przywróci stan po skończeniu analizy. .

Zapisywanie

Nazwy zapisywanych zmiennych. Automatyczne generowanie nazw zapewnia zachowanie wszystkich wyników pracy. Nazwy użytkownika umożliwiają likwidację/zastąpienie wyników poprzednich uruchomień bez wcześniejszego usuwania zapisanych zmiennych w Edytorze danych.

Zmienne przeznaczone do zapisania

- **Przewidywana wartość lub kategoria.** Funkcja zapisuje przewidywaną wartość ilościowej zmiennej docelowej lub przewidywaną kategorię jakościowej zmiennej docelowej.
- **Przewidywane prawdopodobieństwo.** Funkcja zapisuje przewidywane prawdopodobieństwa jakościowej zmiennej docelowej. Osobna zmienna jest zapisywana dla wszystkich pierwszych n kategorii, gdzie n jest określone w kontrolce **Maksymalna liczba kategorii dla kategoryjnej zmiennej zależnej**.
- **Zmienne podziału szkoleniowego/wstrzymanego.** Jeżeli obserwacje są przypisywane losowo do próby szkoleniowej i wstrzymanej na karcie Podziały, powoduje to zapisanie wartości podziału (szkoleniowego lub wstrzymanego), do którego została przydzielona obserwacja.
- **Podział na grupy walidacji krzyżowej.** Jeżeli obserwacje są przypisywane losowo do krotności walidacji krzyżowej w zakładce Podziały, powoduje to zapisanie wartości krotności, do której została przydzielona obserwacja.

Wynik

Dane wyjściowe w przeglądarce

- **Podsumowanie przetwarzania przypadku.** Wyświetla tabelę podsumowania przetwarzania obserwacji, która podsumowuje liczbę obserwacji włączonych i wyłączonych z analizy, w sumie i według próby szkoleniowej i próby wstrzymanej.
- **Wykresy i tabele.** Wyświetla wyniki związane z modelem, w tym tabele i wykresy. Tabele w widoku modelu zawierają k obserwacji najbliższego sąsiedztwa oraz odległości od obserwacji kluczowych, klasyfikację kategoryjnych zmiennych odpowiedzi i podsumowanie błędów. Wyniki graficzne w widoku modelu zawierają dziennik błędów wyboru, wykres ważności właściwości, wykres przestrzeni właściwości, wykres elementów równorzędnych oraz mapę kwadratową. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok modelu ”](#) na stronie 174.

Zbiory

- **Eksportuj model do XML.** Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. . Ta opcja nie jest dostępna, jeśli zdefiniowano podzielone dane.

- **Eksportuj odległości między obserwacjami kluczowymi i k najbliższymi sąsiadami.** Dla każdego punktu centralnego, tworzona jest osobna zmienna dla każdej z k obserwacji w najbliższym sąsiedztwie (z próby szkoleniowej) obserwacji kluczowych oraz odpowiadających im k najbliższych odległości.

Opcje

Braki danych zdefiniowane przez użytkownika. Zmienne kategoryjne muszą posiadać prawidłowe wartości dla obserwacji, która ma zostać zawarta w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako prawidłowe wśród zmiennych kategoryjnych.

Systemowe braki danych i brakujące wartości zmiennych ilościowych są zawsze traktowane jako nieprawidłowe.

Widok modelu

Po wybraniu opcji **Wykresy i tabelle** z karty Wyniki procedura tworzy obiekt Modelu najbliższego sąsiedztwa w Edytorze raportów. Przez aktywację (dwukrotne kliknięcie) obiektu użytkownik uzyskuje interaktywny przegład modelu. Widok modelu zawiera okno z 2 panelami:

- Pierwszy panel wyświetla przegład modelu nazywany widokiem głównym.
- Drugi panel wyświetla jeden z dwóch rodzajów widoków:

Pomocniczy widok modelu przedstawia więcej informacji o modelu, ale nie koncentruje się na samym modelu.

Połączony widok jest widokiem przedstawiającym szczegółowe informacje o modelu, gdy użytkownik rozwinie część widoku głównego.

Domyślnie pierwszy panel przedstawia przestrzeń właściwości, a drugi – wykres ważności zmiennych. Jeżeli wykres ważności zmiennych nie jest dostępny (tzn. jeśli nie wybrano opcji **Funkcje wagi według ważności** na karcie Funkcje), zostaje wyświetlony pierwszy dostępny widok z listy rozwijanej Widok.

Jeżeli widok nie posiada dostępnych informacji, tekst w menu rozwijanym Widok zostaje wyłączony.

.

Przestrzeń właściwości

Wykres przestrzeni właściwości jest interaktywnym wykresem przestrzeni właściwości (lub podprzestrzeni, jeśli istnieją więcej niż 3 właściwości). Każda oś reprezentuje funkcję modelu, a lokalizacja punktów na wykresie przedstawia wartości tych funkcji dla obserwacji w podziale szkoleniowym i wstrzymanym.

Klucze. Oprócz wartości funkcji, punkty wykresu przekazują inne informacje.

- Kształt określa podział, do którego należy punkt; jest to podział szkoleniowy lub wstrzymany.
- Kolor/deseń punktu oznacza wartość docelowej tej obserwacji; wartości o wyraźnych kolorach są równe kategoriom jakościowych wartości docelowych, a cienie oznaczają zakres wartości docelowych wartości ilościowych. Wskazana wartość podziału szkoleniowego jest wartością obserwowaną; w przypadku podziału wstrzymanego jest to wartość przewidywana. Jeżeli nie określono wartości docelowej, ten klucz nie jest wyświetlany.
- Grubszy obrys oznacza, że obserwacja jest centralna. Obserwacje kluczowe są połączone z k obserwacjami najbliższego sąsiedztwa.

Kontrolki i interaktywność. Kilka elementów sterujących na wykresie umożliwia eksplorację przestrzeni właściwości.

- Możliwy jest wybór zestawu właściwości wyświetlanych na wykresie i zmiana właściwości przedstawianych w określonych wymiarach.
- „Obserwacje kluczowe” są punktami wybranymi na wykresie przestrzeni właściwości. Jeżeli zostanie określona zmienna obserwacji kluczowej, punkty reprezentujące obserwacje centralne

zostaną wstępnie zaznaczone. Każdy punkt może jednak zostać obserwacją kluczową, jeśli będzie wybrany przez użytkownika. Obowiązują „standardowe” kontrolki wyboru punktu; kliknięcie punktu powoduje zaznaczenie tego punktu i odznaczenie innych punktów; kliknięcie punktu z naciśniętym klawiszem Ctrl dodaje punkt do grupy wybranych punktów. Połączone widoki, takie jak np. Wykres elementów równorzędnych, zostaną automatycznie aktualizowane na podstawie obserwacji wybranych w przestrzeni właściwości.

- Możliwa jest zmiana liczby najbliższych sąsiadów (k) wyświetlanych dla obserwacji kluczowych.
- Najechanie kursorem nad punkt na wykresie powoduje wyświetlenie informacji z wartością opisu obserwacji lub numerem obserwacji, jeśli opisy obserwacji nie są zdefiniowane, oraz z obserwowanymi i przewidywanymi wartościami docelowymi.
- Przycisk „Resetuj” umożliwia przywrócenie przestrzeni właściwości do oryginalnego stanu.

Dodawanie i usuwanie pól/zmiennych

Można dodawać nowe pola/zmienne do przestrzeni właściwości lub usuwać te, które są w danym momencie wyświetlane.

Paleta zmiennych

Paleta zmiennych musi zostać wyświetlona zanim możliwe będzie dodanie i usuwanie zmiennych. Aby wyświetlić Paletę zmiennych, przeglądarka modelu musi się znajdować w trybie Edycji i należy wybrać obserwację w przestrzeni właściwości.

1. Aby przełączyć przeglądarkę modelu do trybu Edycji należy z menu wybrać:

Widok > Tryb edycji

2. Po wybraniu Trybu edycji kliknij dowolną obserwację w przestrzeni właściwości.
3. Aby wyświetlić Paletę zmiennych, z menu wybierz:

Widok > Palety > Zmienne

W Paletce zmiennych zestawione są wszystkie zmienne w przestrzeni właściwości. Ikony obok nazwy zmiennej oznaczają poziom pomiaru zmiennej.

4. Aby tymczasowo zmienić poziom pomiaru zmiennej, kliknij prawym przyciskiem myszy paletę zmiennych i wybierz opcję.

Strefy zmiennych

Zmienne dodaje się do „stref” w przestrzeni właściwości. Aby wyświetlić te strefy należy rozpocząć przeciąganie zmiennej z Palety zmiennych lub wybrać opcję **Pokaż strefy**.

Przestrzeń właściwości ma strefy dla osi x , y i z .

Przenoszenie zmiennych do stref

Poniżej przedstawiono kilka ogólnych zasad i wskazówek dotyczących przenoszenia zmiennych do stref:

- W celu przeniesienia zmiennej do strefy należy kliknąć i przeciągnąć zmienną z Palety zmiennych oraz opuścić ją w tej strefie. Jeśli wybierze się opcję **Pokaż strefy**, można również kliknąć prawym przyciskiem myszy strefę i wybrać zmienną, która ma być dodana do strefy.
- Jeśli przeciągnie się zmienną z Palety zmiennych do zmiennej, która jest już zajęta przez inną zmienną, stara zmienna jest zastępowana nową.
- Jeśli przeciągnie się zmienną z jednej strefy, do strefy, która jest już zajęta przez inną zmienną, zmienne zamienią się pozycjami.
- Kliknięcie znaku X w strefie usuwa zmienną z tej strefy.
- Jeśli w wizualizacji znajduje się wiele elementów graficznych, każdy element graficzny może mieć powiązane własne strefy zmiennych. Najpierw wybierz żądany element graficzny.

Ważność zmiennych

Zazwyczaj działania modelujące mają koncentrować się na zmiennych, które są najważniejsze, a opuszczane lub ignorowane mają być te zmienne, które są najmniej ważne. Wykres ważności zmiennych pomaga osiągnąć ten cel przez wskazanie względnej ważności każdej zmiennej przy szacowaniu modelu. Ponieważ wartości są względne, suma wartości wszystkich wyświetlanych zmiennych wynosi 1,0. Ważność zmiennej nie jest powiązana z dokładnością modelu. Jest powiązana z ważnością każdej zmiennej przy prognozach, a nie z tym, czy taka prognoza jest dokładna.

Elementy równorzędne

Wykres przedstawia obserwacje kluczowe oraz k obserwacji najbliższego sąsiedztwa dla każdej funkcji i wartości docelowej. Wykres jest dostępny, jeśli zaznaczono obserwację kluczową w przestrzeni właściwości.

Zachowanie łączenia. Wykres elementów równorzędnych jest połączony z przestrzenią właściwości na dwa sposoby.

- Obserwacje wybrane (centralne) w przestrzeni właściwości są wyświetlane na wykresie elementów równorzędnych razem z k obserwacjami w najbliższym sąsiedztwie.
- Wartość k wybrana w przestrzeni właściwości jest używana na wykresie elementów równorzędnych.

Odległości najbliższego sąsiedztwa

Tabela przedstawia k obserwacji najbliższego sąsiedztwa oraz odległości wyłącznie do obserwacji kluczowych. Wykres jest dostępny, jeśli określono identyfikator obserwacji kluczowej na karcie Zmienne. Przedstawia on wyłącznie obserwacje centralne określone przez tę zmienną.

Każdy wiersz:

- Kolumna **Obserwacja kluczowa** zawiera wartość zmiennej opisu obserwacji dla obserwacji centralnej. Jeżeli opis obserwacji nie jest zdefiniowany, kolumna zawiera numer obserwacji centralnej.
- i -ta kolumna grupy Najbliższe sąsiedztwo zawiera wartość zmiennej opisu obserwacji dla i -tej obserwacji najbliższego sąsiedztwa obserwacji kluczowej; jeśli opis obserwacji nie jest zdefiniowany, kolumna zawiera numer i -tej obserwacji najbliższego sąsiedztwa obserwacji centralnej.
- i -ta kolumna grupy Najbliższe sąsiedztwo zawiera odległość i -tej obserwacji najbliższego sąsiedztwa do obserwacji kluczowej

Mapa kwadratowa

Wykres przedstawia obserwacje kluczowe oraz k obserwacji najbliższego sąsiedztwa na wykresie rozrzutu (lub wykresie punktowym, w zależności od poziomu pomiaru wartości docelowej), z wartością docelową na osi y i funkcją ilościową na osi x , ograniczone funkcjami. Wykres jest dostępny, jeśli istnieje wartość docelowa i w przestrzeni właściwości zaznaczono obserwację kluczową.

- Linie referencyjne są rysowane dla zmiennych ilościowych, przy średnich zmiennej w podziale szkoleniowym.

Dziennik błędów wyboru funkcji

Punkty na wykresie przedstawiają błąd (współczynnik poziomy błędu lub błąd sumy kwadratów, w zależności od poziomu pomiaru wartości docelowej) na osi y dla modelu z funkcją przedstawioną na osi x (plus wszystkie funkcje na lewo od osi x). Wykres jest dostępny, jeśli istnieje wartość docelowa i działa wybór funkcji.

Dziennik błędów wyboru k

Punkty na wykresie przedstawiają błąd (współczynnik poziomy błędu lub błąd sumy kwadratów, w zależności od poziomu pomiaru wartości docelowej) na osi y dla modelu z liczbą obserwacji najbliższych

sąsiadów (k) przedstawioną na osi x . Wykres jest dostępny, jeśli istnieje wartość docelowa i działa wybór k .

Dziennik błędów wyboru k i funkcji

Są to wykresy wyboru funkcji (patrz “Dziennik błędów wyboru funkcji” na stronie 176), obok k . Ten wykres będzie dostępny, jeśli istnieje element docelowy i wybrano zarówno opcję k , jak i funkcję.

Tabela klasyfikacji

Ta tabela przedstawia klasyfikację krzyżową wartości obserwowanych i przewidywanych wartości docelowej, według podziału. Tabela jest dostępna, jeśli istnieje wartość docelowa i jest jakościowa.

- Wiersz (**Brakujące**) w podziale wstrzymanym zawiera obserwacje wstrzymane z brakującymi wartościami w wartości docelowej. Obserwacje te mają wpływ na próbę wstrzymaną: wartości Ogólnie procent, ale nie wartości Procent poprawny.

Podsumowanie błędów

Tabela jest dostępna, jeśli istnieje zmienna docelowa. W tabeli jest wyświetlany błąd powiązany z modelem; suma kwadratów dla docelowych wartości ilościowych oraz poziom błędu (100% - ogólnie procent poprawnie) dla docelowych wartości jakościowych.

Analiza dyskryminacyjna

Analiza dyskryminacyjna umożliwia budowanie modelu prognozowego przynależności do grup. Model składa się z funkcji dyskryminacyjnej (lub, w przypadku więcej niż dwóch grup, z zestawu funkcji dyskryminacyjnych) opartej na liniowych kombinacjach zmiennych predyktora, które zapewniają najlepsze rozróżnienie między grupami. Funkcje są generowane z próbki obserwacji, których przynależność do grupy jest znana. Funkcje mogą następnie zostać zastosowane do nowych obserwacji, gdzie znane są miary dla predyktorów, ale nie przynależność do grupy.

Uwaga: zmienna grupująca może mieć więcej niż dwie wartości. Jednakże kody dla zmiennej grupującej muszą być liczbami całkowitymi; należy też określić ich wartości minimalne i maksymalne. Obserwacje z wartościami znajdującymi się poza tymi granicami są wyłączone z analizy.

Przykład. Ludzie zamieszkujący kraje leżące w strefie o klimacie umiarkowanym zazwyczaj spożywają dziennie więcej kalorii niż ci, którzy zamieszkują kraje tropikalne. Oprócz tego w strefie umiarkowanej większa jest proporcja ludzi, którzy mieszkają w miastach. Osoba przeprowadzająca badanie chce połączyć te informacje w funkcję, aby stwierdzić, na ile można rozróżnić te dwie grupy krajów. Uważa ona, że wielkość populacji i czynniki ekonomiczne również mogą mieć znaczenie. Analiza dyskryminacyjna umożliwia ocenę współczynników liniowej funkcji dyskryminacyjnej, która wyglądem przypomina prawą stronę wielokrotnego równania regresji liniowej. I tak, funkcja wykorzystująca współczynniki a , b , c i d wygląda następująco:

$$D = a * climate + b * urban + c * population + d * gross domestic product per capita$$

Jeśli te zmienne są przydatne do rozróżnienia między dwiema strefami klimatycznymi, wartości D będą się różnić dla krajów umiarkowanych i tropikalnych. Jeśli użytkownik korzysta z metody krokowej wyboru zmiennej, uwzględnienie w tej funkcji wszystkich czterech zmiennych może okazać się zbędne.

Statystyki. Dla każdej zmiennej: średnie, odchylenia standardowe, ANOVA jednej zmiennej. Dla każdej analizy: M Boxa, macierz korelacji wewnątrzgrupowej, macierz kowariancji wewnątrzgrupowej, macierz kowariancji dla odrębnych grup, macierz kowariancji całkowitej. Dla każdej kanonicznej funkcji dyskryminacyjnej: wartość własna, procent wariancji, korelacja kanoniczna, lambda Wilksa, chi-kwadrat. Dla każdego kroku: prawdopodobieństwa a priori, współczynniki funkcji Fishera, niestandardyzowane współczynniki funkcji, lambda Wilksa dla każdej funkcji kanonicznej.

Wymagania dotyczące danych dla analizy dyskryminacyjnej

Dane. Zmienna grupująca musi mieć ograniczoną liczbę odrębnych kategorii, zakodowanych jako liczby całkowite. Zmienne niezależne, które są nominalne muszą być ponownie kodowane jako zmienne sztuczne lub kontrastowe.

Założenia. Obserwacje powinny być niezależne. Predyktory powinny mieć wielowymiarowy rozkład normalny, a macierze wariancji-kowariancji wewnątrzgrupowej powinny być równe we wszystkich grupach. Zakłada się, że przynależność do grupy jest wzajemnie wyłączna (oznacza to, że żadna obserwacja nie należy do więcej niż jednej grupy) i kolektywnie wyczerpująca (oznacza to, że wszystkie obserwacje należą do grup). Procedura jest najefektywniejsza, kiedy przynależność do grupy jest zmienną prawdziwie kategorialną. Jeśli przynależność do grupy opiera się na wartościach zmiennej ilościowej (na przykład wysoki a niski iloraz inteligencji), należy rozważyć możliwość wykorzystania regresji liniowej, tak aby skorzystać z bogatszych informacji oferowanych przez samą zmienną ilościową.

Wykonywanie analizy dyskryminacyjnej

1. Wybierz z menu następującą opcję:

Analiza > Klasyfikacja > Analiza dyskryminacyjna...

2. Wybierz zmienną grupującą o wartości będącej liczbą całkowitą i kliknij przycisk **Definiuj grupy**, aby określić kategorie, które Cię interesują.

3. Wybierz zmienną niezależną lub predyktor (jeśli wartości zmiennej grupującej nie są liczbami całkowitymi, za pomocą polecenia Automatyczne rekodowanie z menu Przekształcenia można utworzyć potrzebną wartość).

4. Wybierz metodę wprowadzania zmiennych niezależnych.

- **Wprowadź razem zmienne niezależne.** Jednocześnie wpisywane są wszystkie niezależne zmienne spełniające kryteria tolerancji.
- **Użyj metody krokowej.** Używa analizy krokowej do wprowadzania oraz usuwania zmiennej sterującej.

5. Opcjonalnie można wybrać obserwacje ze zmienną wyboru.

Analiza dyskryminacyjna: Definiuj zakres

Określ wartości minimum i maksimum zmiennej grupującej do analizy. Obserwacje, których wartości znajdują się poza tym zakresem, nie są wykorzystywane w analizie dyskryminacyjnej. Są one klasyfikowane do jednej z istniejących grup w oparciu o wyniki analizy. Wartości minimum i maksimum muszą być liczbami całkowitymi.

Analiza dyskryminacyjna: Wybierz obserwacje

Aby wybrać obserwacje do analizy:

1. W oknie dialogowym Analiza dyskryminacyjna, wybierz zmienną wyboru.

2. Kliknij **Wartość**, aby wpisać liczbę całkowitą jako wartość wyboru.

Do wyprowadzenia funkcji dyskryminacyjnych wykorzystywane są jedynie obserwacje o tej wartości dla zmiennej wyboru. Statystyki i wyniki klasyfikacji są generowane zarówno dla wybranych, jak i dla niewybranych obserwacji. W ten sposób uzyskujemy mechanizm klasyfikacji nowych obserwacji na podstawie już istniejących danych lub podzbioru danych na podzbiory szkoleniowe i testowe w celu przeprowadzenia sprawdzenia na wygenerowanym modelu.

Analiza dyskryminacyjna: Statystyki

Statystyki opisowe. Dostępne opcje to: średnia (w tym odchylenia standardowe), ANOVA dla każdej zmiennej i test *M* Boxa.

- *Średnie.* Wyświetla średnią ogólną i średnie w grupach oraz odchylenia standardowe dla zmiennych niezależnych.

- *ANOVA dla każdej zmiennej*. Dla każdej zmiennej niezależnej wykonuje test istotności różnic między średnimi grupowymi metodą jednoczynnikowej analizy wariancji.
- *M Boxa*. Test na równość macierzy kowariancji grupy. Przy odpowiednio dużych wielkościach prób nieistotna wartość p oznacza, że dowód nierówności macierzy jest niewystarczający. Test jest wrażliwy na odstępstwa od normalności rozkładu wielowymiarowego.

Współczynniki funkcji. Dostępne opcje to: współczynniki klasyfikacji Fishera i niestandardyzowane współczynniki.

- *Fishera*. Wyświetla współczynniki funkcji klasyfikacji Fishera, które mogą być używane bezpośrednio do klasyfikacji. Dla każdej grupy otrzymywany jest oddzielny zestaw współczynników funkcji klasyfikacji, a przypadek klasyfikuje się do tej grupy, dla której ma najwyższą ocenę dyskryminacyjną (wartość funkcji klasyfikacji).
- *Niestandardyzowane*. Wyświetla niestandardyzowane współczynniki funkcji dyskryminacyjnej.

Macierze. Dostępne macierze współczynników dla zmiennych niezależnych to: macierz korelacji wewnątrzgrupowej, macierz kowariancji wewnątrzgrupowej, macierz kowariancji dla odrębnych grup i macierz kowariancji całkowitej.

- *Korelacja wewnątrzgrupowa*. Wyświetla macierz sumarycznych (połączonych) korelacji wewnątrzgrupowych, uzyskiwaną przez uśrednienie macierzy kowariancji dla wszystkich grup przed obliczeniem korelacji.
- *Kowariancja wewnątrzgrupowa*. Wyświetla zbiorczą macierz kowariancji wewnątrzgrupowej, która może się różnić od łącznej macierzy kowariancji. Macierz jest uzyskiwana przez uśrednienie poszczególnych macierzy kowariancji dla wszystkich grup.
- *Kowariancja odrębnych grup*. Wyświetla osobne macierze kowariancji dla każdej grupy.
- *Kowariancja łączna*. Wyświetla macierz kowariancji obliczanych na podstawie wszystkich obserwacji, tak jakby pochodziły z jednej próby.

Analiza dyskryminacyjna: Użyj metody krokowej

Metoda. Wybierz statystykę, która ma być wykorzystywana do wprowadzania lub usuwania nowych zmiennych. Dostępne alternatywy to λ Wilksa, niewyjaśniona wariancja, odległość Mahalanobis, najmniejszy współczynnik F oraz V_{Rao} . W przypadku V_{Rao} można określić minimalny wzrost wartości V dla zmiennej, która ma zostać wprowadzona.

- *Lambda Wilksa*. Metoda wyboru zmiennych dla krokowej analizy dyskryminacyjnej, która wybiera zmienne do wprowadzenia do równania w oparciu o to, jak bardzo dolną λ Wilksa. W każdym kolejnym kroku procedury wprowadzona zostaje ta zmienna, która minimalizuje wartość tego współczynnika.
- *Niewyjaśniona wariancja*. W każdym kolejnym kroku analizy do modelu wprowadzana jest zmienna, która minimalizuje sumę niewyjaśnionej zmienności między grupami.
- *Odległość Mahalanobis*. Miara wielkości liter w zmiennych niezależnych od średniej dla wszystkich obserwacji. Duże wartości wskaźnika Mahalanobisa oznaczają, że obserwacja zawiera skrajne wartości jednej albo większej liczby zmiennych niezależnych.
- *Najmniejszy współczynnik F*. Metoda doboru zmiennych przy analizie metodą krokową, oparta na maksymalizacji ilorazu F , obliczanego na podstawie odległości Mahalanobisa pomiędzy grupami.
- *V Rao*. Miara różnic między środkami grupy. Znana jest także pod nazwą śladu Lawleya-Hotellinga. W każdym kolejnym kroku procedury wprowadzona zostaje ta zmienna, która powoduje największy wzrost wskaźnika V . Po wybraniu tej opcji wprowadź minimalną wartość, którą zmienna musi posiadać, aby została wprowadzona do analizy.

Kryteria. Dostępne alternatywy to **Użyj wartości F** i **Zastosuj prawdopodobieństwo F**. Podaj wartości dla wprowadzania i usuwania zmiennych.

- *Użyj wartości F*. Zmienna jest wprowadzana do modelu, jeśli jego wartość F jest większa niż wartość pozycji i jest usuwana, jeśli wartość F jest mniejsza niż wartość usunięcia. Wartość wprowadzenia musi być większa od wartości usunięcia i obie muszą być dodatnie. Chcąc wprowadzić więcej zmiennych

do modelu, należy obniżyć wartość wprowadzenia. Chcąc usunąć więcej zmiennych, należy zwiększyć wartość usunięcia.

- *Zastosuj prawdopodobieństwo F*. Zmienna jest wprowadzana do modelu, jeśli poziom istotności jego wartości F jest mniejszy od wartości pozycji i jest usuwany, jeśli poziom istotności jest większy niż wartość usunięcia. Wartość wprowadzenia musi być mniejsza od wartości usunięcia i obie muszą być dodatnie. Chcąc wprowadzić więcej zmiennych do modelu, należy zwiększyć wartość wprowadzenia. Aby usunąć więcej zmiennych, należy zmniejszyć wartość usunięcia.

Pokaż. Podsumowanie kroków wyświetla statystyki dla wszystkich zmiennych po każdym kroku. **F dla odległości parami** wyświetla macierz ilorazów F parami dla każdej pary grup.

Analiza dyskryminacyjna: Klasyfikacja

Prawdopodobieństwa a priori. Dzięki tej opcji można określić, czy współczynniki klasyfikacji są dostosowane do wiedzy a priori o przynależności do grup.

- **Dla wszystkich grup równe.** Dla wszystkich grup przyjmowane są równe prawdopodobieństwa wstępne. Nie ma to wpływu na współczynniki.
- **Oblicz na podstawie wielkości grup.** Wstępne prawdopodobieństwa przynależności do grup są określane na podstawie zaobserwowanych w próbkę rozmiarów grup. Na przykład jeśli 50% obserwacji włączonych do analizy należy do pierwszej grupy, 25% do drugiej i 25% do trzeciej, współczynniki klasyfikacji są dopasowane do zwiększonego prawdopodobieństwa przynależności do pierwszej grupy w stosunku do pozostałych dwóch.

Pokaż. Dostępne opcje wyświetlania to: Wyniki obserwacji, Tabela podsumowań i Klasyfikacja typu pozostaw-jedną-pozą.

- *Wyniki obserwacji.* Dla każdej wyświetlane są kody rzeczywistej grupy, przewidywanej grupy, prawdopodobieństw a posteriori i ocen dyskryminacyjnych.
- *Tabela podsumowania.* Liczba obserwacji poprawnie i nieprawidłowo przypisanych do każdej z grup w oparciu o analizę dyskryminacyjną. Czasem zwana „Macierzą nieporozumień”.
- *Klasyfikacja typu pozostaw-jedną-pozą.* Każda obserwacja w analizie jest klasyfikowana przez funkcje pochodzące ze wszystkich obserwacji innych niż ta sprawa. Znana również jako „metoda U”.

Zastąp brakujące wartości średnią. Zaznacz tę opcję, aby zastąpić średnią zmiennej niezależnej dla braku danych jedynie w fazie klasyfikacji.

Użyj macierzy kowariancji. Możesz klasyfikować obserwacje z wykorzystaniem macierzy kowariancji wewnątrzgrupowych lub macierzy kowariancji odrębnych dla grup.

- *Wewnątrzgrupowe.* Do klasyfikacji obserwacji wykorzystywana jest połączona macierz kowariancji wewnątrzgrupowych.
- *Oddzielne grupy.* Wykorzystuje do klasyfikacji macierze kowariancji dla poszczególnych grup. Ponieważ klasyfikacja oparta jest na funkcjach dyskryminacyjnych a nie na pierwotnych zmiennych, opcja ta nie zawsze jest równoważna dyskryminacji kwadratowej.

Wykresy. Dostępne opcje wykresów to: Połączone grupy, Odrębne dla grup i Mapa terytorialna.

- *Grupy złożone.* Tworzy wykres rozrzutu wartości dwóch pierwszych funkcji dyskryminacyjnych, obejmujący wszystkie grupy. Jeśli istnieje tylko jedna funkcja, wyświetlany jest histogram.
- *Oddzielne grupy.* Tworzy wykresy rozrzutu odrębne dla każdej z grup, z uwzględnieniem pierwszych dwu funkcji dyskryminacyjnych. Jeśli istnieje tylko jedna funkcja, wyświetlone zostaną histogramy.
- *Mapa terytorialna.* Oparty o wartości funkcji dyskryminacyjnej wykres granic, wykorzystany do klasyfikowania obserwacji do grup. Liczby odpowiadają grupom, do których zostały zaklasyfikowane poszczególne obserwacje. Średnie dla kolejnych grup są na wykresie oznaczone gwiazdkami, które znajdują się wewnątrz granic określonych dla tych grup. Mapa nie zostaje wyświetlona wtedy, gdy istnieje tylko jedna funkcja dyskryminacyjna.

Analiza dyskryminacyjna: Zapisz

Do aktywnego pliku danych można dodawać nowe zmienne. Dostępne opcje to: Przewidywana przynależność do grupy (pojedyncza zmienna), Oceny dyskryminacyjne (jedna zmienna dla każdej funkcji dyskryminacyjnej w rozwiązaniu) i Prawdopodobieństwo przynależności do grupy przy danych ocenach dyskryminacyjnych (jedna zmienna dla każdej grupy).

Można także eksportować informacje o modelu do określonego pliku w formacie XML. Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

Dodatkowe właściwości komendy DISCRIMINANT

Język składni komend umożliwia również:

- Przeprowadzanie wielu analiz dyskryminacyjnych (z jedną komendą) i kontrolowanie porządku, w którym wprowadzane są dane (za pomocą opcji komendy ANALYSIS).
- Określanie prawdopodobieństwa dla klasyfikacji (za pomocą opcji komendy PRIORS).
- Wyświetlanie wzorów rotowania i macierzy struktury (za pomocą opcji komendy ROTATE).
- Ograniczenie liczby wyodrębnionych funkcji dyskryminacyjnych (za pomocą opcji komendy FUNCTIONS).
- Ograniczenia klasyfikacji do obserwacji, które są wybrane (lub nie są wybrane) do analizy (za pomocą opcji komendy SELECT).
- Odczytywanie i analizę macierzy korelacji (za pomocą opcji komendy MATRIX).
- Pisanie macierzy korelacji do dalszej analizy (za pomocą opcji komendy MATRIX).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Analiza czynnikowa

Analiza czynnikowa służy identyfikacji zmiennych zwanych **czynnikami**, które wyjaśniają wzory korelacji występujące w ramach zbiorów obserwowanych zmiennych. Analiza czynnikowa jest często wykorzystywana w redukcji danych w celu identyfikacji niewielkiej liczby czynników, wyjaśniających większą część wariacji obserwowanej w dużej liczbie zmiennych. Analiza czynnikowa może być również wykorzystywana do ustalania hipotez dotyczących mechanizmów przyczynowo-skutkowych lub klasyfikowania zmiennych do dalszych analiz (na przykład do identyfikowania współliniowości przed rozpoczęciem analizy regresji liniowej).

Procedurę analizy czynnikowej cechuje wysoki stopień elastyczności:

- Korzystać można z siedmiu metod wyodrębniania czynników.
- Dostępnych jest pięć metod rotacji, w tym metoda prostej Oblimin i rotacja Promax dla rotacji nieortogonalnych.
- Dostępne są trzy metody wyliczania ocen czynnikowych, które to oceny można następnie zapisywać jako zmienne w celu dalszej analizy.

Przykład. Jakie postawy skłaniają ludzi do odpowiadania na pytania zamieszczone w sondażach dotyczących polityki w określony sposób? Analiza korelacji między poszczególnymi pozycjami sondażu wykazuje, że odpowiedzi na pytania z różnych podgrup w znacznym stopniu się pokrywają – wzajemną korelację wykazują na przykład odpowiedzi na pytania dotyczące podatków, obronności itd. Za pomocą analizy czynnikowej można określić liczbę czynników i w wielu przypadkach zidentyfikować, jakie znaczenie konceptualne mają poszczególne czynniki. Oprócz tego można wyliczyć oceny czynnikowe dla każdego respondenta, które mogą być następnie wykorzystywane do dalszych analiz. Można na przykład stworzyć model regresji logistycznej w celu przewidywania zachowania podczas głosowania w zależności od ocen czynnikowych.

Statystyki. W przypadku każdej zmiennej: liczba ważnych obserwacji, średnia i odchylenie standardowe. Dla każdej analizy czynnikowej: macierz korelacji zmiennych z poziomami istotności, wyznacznik i odwrotna; odtworzona macierz korelacji z macierzą przeciwobrazów; rozwiązanie wstępne (zasoby

zmienności wspólnej, wartości własne i procent wariancji wyjaśnionej); miara adekwatności doboru zmiennych Kaisera-Mayera-Olkina i test sferyczności Bartletta; rozwiązanie nierotowane z ładunkami czynnikowymi, zasoby zmienności wspólnej i wartości własne; rozwiązanie rotowane z rotowaną macierzą wzorów i macierzą transformacji. Dla rotacji ukośnych: rotowana macierz wzorów i macierz struktury czynników; macierz współczynników ocen czynnikowych oraz macierz kowariancji ocen czynnikowych. Wykresy: wykres osypiska dla wartości własnych i wykres ładunków czynnikowych dla pierwszych dwóch lub trzech czynników.

Wymagania dotyczące danych do analizy czynnikowej

Dane. Zmienne powinny być zmiennymi ilościowymi na poziomie *interwałowym* lub *ilorazowym*. Dane jakościowe (takie jak religia lub państwo pochodzenia) nie nadają się do analizy czynnikowej. Dane, dla których można wyliczyć współczynniki korelacji Pearsona, powinny być odpowiednie dla analizy czynnikowej.

Założenia. Dane powinny być skorelowane parami i wykazywać rozkład normalny dla każdej pary zmiennych, a obserwacje powinny być niezależne. Model analizy czynnikowej wymaga, aby zmienne były określane przy użyciu czynników wspólnych (oszacowanych przy użyciu modelu) i czynników unikatowych (nie pokrywających się dla obserwowanych zmiennych); wyliczone oszacowania oparte są na założeniu, że wszystkie unikatowe czynniki nie są skorelowane ani wzajemnie, ani z czynnikami wspólnymi.

Wykonywanie analizy czynnikowej

1. Wybierz z menu następującą opcję:

Analiza > Redukcja wymiarów > Czynniki...

2. Wybierz zmienne do analizy czynnikowej.

Wybór obserwacji do analizy czynnikowej

Aby wybrać obserwacje do analizy:

1. Wybierz zmienną filtrującą.

2. Kliknij **Wartość**, aby wpisać liczbę całkowitą jako wartość wyboru.

W analizie czynnikowej wykorzystywane są tylko obserwacje z tą wartością zmiennej filtrującej.

Analiza czynnikowa: Statystyki opisowe

Statystyki. **Statystyki opisowe jednowymiarowe** zawierają średnią, odchylenie standardowe i liczbę ważnych obserwacji dla każdej zmiennej. **Rozwiązanie wstępne** umożliwia wyświetlenie początkowych wartości zasobów zmienności wspólnej, wartości własnych oraz procentu wariancji wyjaśnionej.

Macierz korelacji. Dostępne opcje to: Współczynniki, Poziomy istotności, Wyznacznik, K-M-O i test sferyczności Bartletta, Odwrotna, Odtworzona i Przeciwbraz.

- **Test sferyczności KMO i Bartletta**. Miara odpowiedniości próbek Kaiser-Meyer-Olkin sprawdza, czy częściowe korelacje między zmiennymi są niewielkie. Test sferyczności Bartletta sprawdza, czy macierz korelacji jest macierzą jednostkową. Jeśli nią jest, model czynnikowy jest nieodpowiedni dla analizowanych zmiennych.
- **Reprodukowane.** Szacunkowa macierz korelacji z rozwiązania czynników. Program wyświetla również reszty (różnice pomiędzy oszacowanymi i obserwowanymi korelacjami).
- **Anty obraz.** Macierz korelacji antyobrazowej zawiera negatywy współczynników korelacji cząstkowych, a macierz kowariancji antyobrazowej zawiera negatywy dla kowariancji cząstkowych. Przy dobrym modelu czynników wartości większości elementów leżących poza przekątną będą niewielkie. Miara adekwatności doboru próby dla danej zmiennej jest wyświetlana na przekątnej macierzy korelacji przeciwbrazów.

Analiza czynnikowa: Wyodrębnianie

Metoda. Pozwala na określenie metody wyodrębniania czynników. Dostępne metody to: metoda głównych składowych, metoda nieważonych najmniejszych kwadratów, metoda uogólnionych najmniejszych kwadratów, metoda maksymalnej wiarygodności, metoda osi głównych, metoda alfa oraz metoda obrazu.

- *Analiza głównych składowych.* Metoda wyodrębniania czynników, używana do tworzenia nieskorelowanych kombinacji liniowych obserwowanych zmiennych. Pierwsza składowa tłumaczy najczęściej wariacji. Kolejne składowe wyjaśniają coraz mniejsze części wariacji i są pomiędzy sobą nieskorelowane. Analiza głównych składowych jest wykorzystywana jako metoda wyodrębniania czynników wstępnych. Może być używana wtedy, gdy macierz korelacji jest macierzą osobliwą.
- *Metoda nieważonych najmniejszych kwadratów.* Metoda wyodrębniania czynników, która minimalizuje sumy kwadratów różnic pomiędzy obserwowanymi i odtworzonymi macierzami korelacji, z pominięciem wartości leżących na głównych przekątnych).
- *Metoda uogólnionych najmniejszych kwadratów.* Metoda wyodrębniania czynników, która minimalizuje sumę kwadratów różnic między obserwowanymi i reprodukowanymi macierzami korelacji. Korelacje są ważone przez odwrotność ich wariacji swoistej tak, że zmienne o wysokiej swoistości uzyskują mniejsze wagi od zmiennych o niskiej swoistości.
- *Metoda największej wiarygodności.* Metoda wyodrębniania czynników, która generuje oszacowania parametrów, które najprawdopodobniej wygenerowały obserwowaną macierz korelacji, jeśli próbka pochodzi z rozkładu normalnego wielowariantu. Korelacje są ważone przez odwrotność wariacji swoistej zmiennych z zastosowaniem algorytmu iteracyjnego.
- *Metoda czynnika głównego.* Metoda wyodrębniania czynników z oryginalnej macierzy korelacji, z kwadratowym wielokrotnym współczynnikiem korelacji umieszczonym w przekątnej jako początkowe szacunki wartości współczynnika. Uzyskane w ten sposób ładunki czynnikowe są następnie podstawą do oszacowania nowych zasobów zmienności wspólnej, które zastępują stare oszacowania na głównej przekątnej macierzy korelacji. Iteracje kontynuowane są tak długo, dopóki zmiany w wartościach zasobów zmienności wspólnej, w następujących po sobie iteracjach, nie spełnią kryterium zbieżności.
- *Czynnikowa Alfa.* Metoda wyodrębniania czynników, która uwzględnia zmienne w analizie, które mają być próbką z poziomu wszechświata potencjalnych zmiennych. Ta metoda maksymalizuje wartość wiarygodności alfa czynników.
- *Metoda obrazu.* Metoda ekstrakcji czynnikowca opracowana przez Guttmana i oparta na teorii obrazu. Wariancja wspólna każdej zmiennej (nazywana tu cząstkowym obrazem) definiowana jest nie jako funkcja hipotetycznych czynników, ale jako jej liniowa regresja na pozostałych zmiennych.

Analiza. Pozwala na określenie macierzy korelacji lub macierzy kowariancji.

- **Macierz korelacji.** Użyteczna, gdy zmienne w analizie są mierzone w różnych skalach.
- **Macierz kowariancji.** Użyteczna, gdy chce się zastosować analizę czynnikową do wielu grup z różnymi wariacjami dla każdej zmiennej.

Wyodrębnianie. Można albo zachować wszystkie czynniki, których wartości własne przekraczają określoną wartość, albo zachować określoną liczbę czynników.

Pokaż. Pozwala na określanie, czy wyświetlane ma być nierotowane rozwiązanie czynnikowe i wykres osypiska wartości własnych.

- *Nierotowane rozwiązanie czynnikowe.* Wyświetla nierotowane ładunki czynnikowe (macierz modelowa czynników), zasoby zmienności wspólnej i wartości własne dla rozwiązania czynnikowego.
- *Wykres osypiska.* Wykres wariacji powiązanej z każdym z czynników. Stosowany do określenia liczby czynników, jakie należy pozostawić w modelu. Z reguły na wykresie tym widać wyraźną przerwę pomiędzy stromym nachyleniem ważnych czynników i stopniowo malejącym nachyleniem pozostałych (osypisko).

Maksimum iteracji dla uzyskania zbieżności. Pozwala na określanie maksymalnej liczby kroków realizowanych przez algorytm w celu oszacowania rozwiązania.

Analiza czynnikowa: Rotacja

Metoda. Umożliwia wybranie metody rotacji czynnika i określenie, czy ma zostać zastosowana normalizacja Kaisera. Dostępными metodami są: Varimax, prosta Oblimin, Quartimax, Equamax i Promax.

- *Metoda varimax.* Ortogonalna metoda rotacji, która minimalizuje liczbę zmiennych, które mają wysokie obciążenia dla każdego czynnika. Upraszcza w ten sposób interpretację czynników.
- *Metoda Oblimin prosta.* Metoda rotacji ukośnych (nieortogonalnych). Kiedy delta jest równa 0 (ustawienie domyślne) osie czynników są najbardziej ukośne. Im większą wartość ujemną przyjmie wskaźnik delta, tym mniej ukośne będą osie czynników. Aby zmienić domyślną wartość delty (równą 0) należy wprowadzić liczbę mniejszą od, lub równą, 0,8.
- *Metoda quartimax.* Metoda rotacji, która minimalizuje liczbę czynników potrzebnych do wyjaśnienia każdej zmiennej. Metoda ta upraszcza interpretację obserwowanych zmiennych.
- *Metoda equamax.* Metoda rotacji, która jest kombinacją metody varimax, która upraszcza czynniki, oraz metodę quartimax, która upraszcza zmienne. Technika ta minimalizuje liczbę zmiennych, które mają wysokie ładunki na poszczególnych czynnikach oraz liczbę czynników potrzebnych do wyjaśnienia poszczególnych zmiennych.
- *Rotacja Promax.* Rotacja ukośna, która umożliwia korelowanie czynników. Można ją wyliczyć szybciej niż rotację prostą Oblimin, dlatego jest ona użyteczna w przypadku dużych zbiorów danych.
- *Zastosuj normalizację Kaisera.* Domyślnie, umożliwia zastosowanie normalizacji Kaisera, gdy określono rotacji.

Pokaż. Pozwala na dołączenie wyniku rozwiązania rotowanego oraz wykresów ładunków dla pierwszych dwóch lub trzech czynników.

- *Rozwiązanie rotowane.* Aby uzyskać rotowane rozwiązanie, należy wybrać metodę rotacji. Dla rotacji ortogonalnych wyświetlana jest rotowana macierz modelowa czynników i macierz transformacji czynników. Dla rotacji ukośnych wyświetlana jest macierz modelowa czynników, macierz struktury czynników oraz macierz korelacji czynników.
- *Wykres ładunków czynnikowych.* Wykres trójwymiarowy obciążający wykres pierwszych trzech czynników. W przypadku rozwiązania dwuczynnikowego przedstawiony jest wykres dwuwymiarowy. Kiedy wyodrębniony został tylko jeden czynnik, wykres nie zostaje wyświetlony. Wykresy ukazują rotowane ładunki czynnikowe, o ile rotacja została zastosowana.

Maksimum iteracji dla uzyskania zbieżności. Pozwala na określenie maksymalnej liczby kroków realizowanych przez algorytm w celu przeprowadzenia rotacji.

Analiza czynnikowa: Oceny czynnikowe

Zapisz jako zmienne. Umożliwia utworzenie zmiennej wynikowej dla każdego czynnika w rozwiązaniu końcowym.

Metoda. Alternatywne metody obliczenia ocen czynnikowych to regresja, przybliżenie Bartletta i metoda Andersona-Rubina.

- *Metoda regresji.* Metoda szacowania współczynników oceny czynnika. Uzyskane wartości charakteryzują się średnią równą 0 i wariancją równą kwadratowi korelacji wielokrotnej pomiędzy szacunkowymi ocenami czynników i rzeczywistymi wartościami czynnikowymi. Otrzymane w ten sposób wartości mogą zostać skorelowane nawet wtedy, gdy czynniki są ortogonalne.
- *Wyniki Bartletta.* Metoda estymacji współczynników ocen czynnikowych. Średnia wartość generowanych wyników wynosi 0. Suma kwadratów unikalnych czynników w zakresie zmiennych jest minimalizowana.
- *Metoda Andersona-Rubina.* Metoda szacowania współczynników ocen czynnikowych; modyfikacja metody Bartletta, która zapewnia ortogonalność oszacowanych czynników. Otrzymane oceny mają średnią równą 0, odchylenie standardowe równe 1 i są nieskorelowane.

Wyświetl macierz współczynników ocen czynnikowych. Umożliwia wyświetlenie współczynników, przez które mnożone są zmienne w celu uzyskania ocen czynnikowych. Umożliwia również wyświetlenie korelacji między ocenami czynnikowymi.

Analiza czynnikowa: Opcje

Braki danych. Pozwala na określenie sposobu postępowania z brakami danych. Dostępne opcje to: wyłączenie wszystkich obserwacji z *brakami danych*, wyłączenie obserwacji *parami* i zastępowanie średnią.

Format wyświetlania współczynników. Pozwala na wybór opcji dotyczących macierzy wyjściowych. Współczynniki można sortować według wartości ładunków czynnikowych i ukrywać współczynniki o wartościach bezwzględnych mniejszych od określonej wartości.

Dodatkowe właściwości komendy FACTOR

Język składni komend umożliwia również:

- Określanie kryteriów zbieżności dla iteracji podczas wyodrębniania i rotacji.
- Określanie poszczególnych wykresów z rotowanym czynnikiem.
- Określanie ile ocen czynnikowych ma być zapisanych.
- Określanie przekątnych wartości dla metody wyznaczania czynników osi głównych.
- Pisanie macierzy korelacji lub macierzy ładunków czynnikowych do późniejszej analizy.
- Odczytywanie i analizę macierzy korelacji i macierzy ładunków czynnikowych.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Wybieranie procedury analizy skupień

Analiza skupień może być wykonywana przy wykorzystaniu procedury Dwustopniowe grupowanie, Hierarchiczna analiza skupień lub Analiza metodą K-średnich. W każdej procedurze funkcjonuje inny algorytm tworzenia skupień, który zawiera opcje niedostępne w innych.

Dwustopniowe grupowanie. W wielu zastosowaniach Dwustopniowe grupowanie jest najodpowiedniejszą metodą analizy skupień. Posiada ona następujące cechy unikalne:

- Automatyczne wybieranie najodpowiedniejszej liczby skupień, obok możliwości wybierania modeli skupień.
- Możliwość jednoczesnego tworzenia modeli skupień opartych na zmiennych jakościowych i ilościowych.
- Możliwość zapisania modelu skupień w zewnętrznym pliku XML, a następnie odczytania takiego pliku i aktualizacji modelu skupień przy wykorzystaniu nowszych danych.

Procedura Dwustopniowe grupowanie umożliwia ponadto analizowanie dużych plików danych.

Hierarchiczna analiza skupień. Procedura Hierarchiczna analiza skupień może być stosowana tylko do mniejszych plików danych (kilkaset grupowanych obiektów), lecz posiada następujące cechy unikalne:

- Możliwość grupowania obserwacji lub zmiennych.
- Możliwość obliczania zakresu możliwych rozwiązań i zapisywania przynależności do skupień dla każdego z takich rozwiązań.
- Kilka metod formowania skupień, transformacji zmiennych i pomiaru niepodobieństwa pomiędzy skupieniami.

O ile zmienne są tego samego typu, procedura Hierarchiczna analiza skupień umożliwia analizę zmiennych interwałowych (ilościowych), liczebności i zmiennych binarnych.

Analiza skupień metodą k-średnich. Procedura Analiza skupień metodą k-średnich może być stosowana tylko do danych ilościowych i wymaga określenia liczby skupień z góry, lecz posiada następujące cechy unikalne:

- Możliwość zapisania odległości od centrów skupień dla każdego obiektu.
- Możliwość odczytywania wstępnych centrów skupień z zewnętrznego pliku IBM SPSS Statistics i zapisywania końcowych centrów skupień w tym pliku.

Procedura Analiza skupień metodą k-średnich umożliwia ponadto analizowanie dużych plików danych.

Dwustopniowa analiza skupień

Procedura Dwustopniowe grupowanie jest narzędziem eksploracyjnym, mającym na celu ujawnienie występowania w zbiorze danych naturalnych zgrupowań (lub skupień), które nie są widoczne w inny sposób. Algorytm zastosowany w tej procedurze posiada kilka wyjątkowych cech, które odróżniają go od tradycyjnych metod grupowania:

- **Obsługa zmiennych jakościowych i ilościowych.** Przy założeniu niezależności zmiennych, do zmiennych jakościowych i ilościowych można zastosować połączony rozkład wielomianowo-normalny.
- **Automatyczny wybór liczby skupień.** Przez porównanie wartości kryterium wyboru modelu dla różnych rozwiązań grupowania procedura może automatycznie określić optymalną liczbę skupień.
- **Skalowalność.** Przez utworzenie drzewa cech skupień (CF) podsumowującego rekordy algorytm TwoStep umożliwia analizę dużych plików danych.

Przykład. Przedsiębiorstwa handlu detalicznego i produktów konsumpcyjnych regularnie stosują techniki grupowania do danych, które opisują zwyczaje nabywcze swoich klientów, płeć, wiek, dochody itp. Przedsiębiorstwa te dostosowują swoje strategie marketingowe i produktowe do każdej grupy konsumentów celem zwiększenia sprzedaży i pozyskiwania lojalności klientów wobec danej marki produktów.

Miara odległości. Wybrana tutaj opcja określa sposób wyliczenia podobieństwa dwóch skupień.

- **Logarytm wiarygodności.** Miara wiarygodności stosuje do zmiennych rozkład prawdopodobieństwa. Zakłada się, że zmienne ilościowe mają rozkład normalny, natomiast kategoryjne rozkład wielomianowy. Zakłada się, że wszystkie zmienne są niezależne.
- **Euklidesowa.** Odległość euklidesowa jest odległością „w linii prostej” pomiędzy dwoma skupieniami. Można jej użyć tylko wówczas, gdy wszystkie zmienne są zmiennymi ilościowymi.

Liczba skupień. Wybór tej opcji pozwala na określenie sposobu ustalenia liczby skupień.

- **Dobierz automatycznie.** Procedura automatycznie dobierze „optymalną” liczbę skupień przy zastosowaniu kryterium określonego w grupie opcji Kryterium grupowania. Opcjonalnie można wprowadzić dodatnią liczbę całkowitą, określającą maksymalną liczbę skupień, które procedura weźmie pod uwagę.
- **Ustalona liczba skupień.** Umożliwia uwzględnienie w rozwiązaniu stałej liczby skupień. Wprowadź dodatnią liczbę całkowitą.

Zlicz zmienne ilościowe. Ta grupa zawiera podsumowanie specyfikacji standaryzacyjnych zmiennych ilościowych, określonych w oknie dialogowym Opcje. Aby uzyskać dodatkowe informacje, patrz temat: [“Dwustopniowe grupowanie: Opcje” na stronie 187.](#)

Kryterium grupowania. Wybrana tutaj opcja określa sposób ustalenia liczby skupień przez algorytm automatycznego grupowania. Dostępne opcje to Bayesowskie Kryterium Informacyjne (BIC) i Kryterium informacyjne Akaike (AIC).

Wymagania dotyczące danych dla dwustopniowego grupowania

Dane. Procedura znajduje zastosowanie zarówno w przypadku zmiennych ilościowych, jak i zmiennych kategoryjnych. Obserwacje reprezentują obiekty do pogrupowania, natomiast zmienne reprezentują atrybuty, na podstawie których odbywa się grupowanie.

Kolejność obserwacji. Warto zauważyć, że drzewa cech skupień i ostateczne rozwiązanie mogą zależeć od kolejności obserwacji. Aby zminimalizować wpływ kolejności, należy losowo ustawić obserwacje. Aby zweryfikować stabilność danego rozwiązania może wystąpić konieczność uzyskania kilku różnych rozwiązań przy sortowaniu przy różnej, przypadkowej kolejności obserwacji. W sytuacjach, kiedy jest to trudne z uwagi na niezwykle duże rozmiary plików, wiele uruchomień za pomocą obserwacji sortowanych w porządku losowym może zostać zastąpione.

Założenia. Miara odległości wiarygodności zakłada, że zmienne w modelu skupień są niezależne. Ponadto zakłada się, że każda zmienna ilościowa posiada rozkład normalny (Gaussa), a każda

zmienna kategoryalna rozkład wielomianowy. Chociaż empiryczne testy wewnętrzne wykazały dosyć dużą odporność procedury na niespełnienie założeń odnośnie niezależności i rozkładu, warto wiedzieć, w jakim stopniu założenia takie zostały spełnione.

Za pomocą procedury Korelacje parami należy przetestować niezależność dwóch zmiennych ilościowych. Za pomocą procedury Tabele krzyżowe należy przetestować niezależność dwóch zmiennych kategoryalnych. Za pomocą procedury Średnie należy przetestować niezależność pomiędzy zmienną ilościową a kategoryalną. Za pomocą procedury Eksploracja należy przetestować normalność zmiennej ilościowej. Procedura Test chi-kwadrat służy do testowania, czy zmienna kategoryalna ma określony rozkład wielomianowy.

Wykonywanie dwustopniowego grupowania

1. Z menu wybierz:

Analiza > Klasyfikacja > Dwustopniowe grupowanie...

2. Wybierz co najmniej jedną zmienną jakościową lub ilościową.

Opcjonalnie można wykonać następujące kroki:

- Skorygować kryteria tworzenia skupień.
- Wybrać ustawienia obsługi szumu, alokacji pamięci, standaryzacji zmiennych i danych do modelu skupień.
- Zażądać wyników przeglądarki modelu.
- Zapisać wyniki działania modelu w pliku roboczym lub w zewnętrznym pliku XML.

Dwustopniowe grupowanie: Opcje

Traktowanie wartości odstających. W tej grupie można określić odmienny sposób traktowania wartości odstających podczas grupowania w przypadku wypełnienia drzewa cech skupień (CF). Drzewo CF jest pełne, jeśli nie może przyjąć więcej obserwacji w węzle liścia i nie ma możliwości podziału żadnego z takich węzłów.

- Jeśli drzewo CF wypełni się przy wybranej opcji obsługi szumu, po przeniesieniu obserwacji z liści rzadkich do liścia „szumu” zostanie ukształtowane na nowo. Liść jest uważany za rzadki, jeśli zawiera mniej obserwacji niż określony procent jego maksymalnej wielkości. Po ponownym ukształtowaniu drzewa CF zostaną w nim umieszczone obserwacje odstające, jeśli istnieje taka możliwość. Jeśli nie będzie to możliwe, zostaną odrzucone.
- Jeśli drzewo CF wypełni się, kiedy nie będzie wybrana opcja obsługi szumu, zostanie ukształtowane na nowo przy wykorzystaniu większego progu zmiany odległości. Po zakończeniu finalnego grupowania wartości, których nie można przypisać do żadnego skupienia, zostaną oznaczone jako odstające. Grupie odstającej nadawany jest numer identyfikacyjny -1 i nie jest ona uwzględniana w zliczaniu liczby grup.

Alokacja pamięci. W tej grupie można określić maksymalną wielkość pamięci w megabajtach (MB) przeznaczoną na potrzeby algorytmu grupowania. Jeśli procedura przekroczy taką wielkość maksymalną, dane nie mieszczące się w pamięci będą przechowywane na dysku. Należy podać liczbę większą lub równą 4.

- Maksymalną wartość, jaką można użyć w systemie, można uzyskać od administratora systemu.
- Jeśli ta wartość będzie zbyt niska, algorytm może nie znaleźć właściwej lub określonej liczby skupień.

Standaryzacja zmiennych. Algorytm grupowania pracuje na standaryzowanych zmiennych ilościowych. Wszelkie niestandaryzowane zmienne ilościowe powinny zostać pominięte jako zmienne na liście Przeznaczone do standaryzacji. Dla zaoszczędzenia czasu i mocy obliczeniowych warto wybrać zmienne ilościowe, które już zostały ustandaryzowane i występują na liście Założona standaryzacja.

Opcje zaawansowane

Kryteria kształtowania drzewa CF. Poniższe ustawienia algorytmu grupowania są specyficzne dla drzewa cech skupień (CF). Przy ich zmianie należy zachować ostrożność:

- **Początkowy próg zmiany odległości.** Jest to początkowy próg wykorzystywany przy kształtowaniu drzewa CF. Jeśli umieszczenie danej obserwacji w liściu drzewa CF powodowałoby takie zagęszczenie obserwacji, że odległość między nimi byłaby niższa od progu, liść nie jest dzielony. Jeśli zagęszczenie jest wyższe od progu, liść jest dzielony.
- **Maksimum gałęzi (na węzeł).** Maksymalna liczba węzłów podrzędnych, które węzeł liścia może posiadać.
- **Maksymalna głębokość drzewa.** Maksymalna liczba poziomów, które drzewo CF może posiadać.
- **Maksymalna liczba możliwych węzłów.** Wskazuje to na maksymalną liczbę węzłów drzewa CF, które mogą zostać wygenerowane przez daną procedurę, na podstawie funkcji $(b^{d+1} - 1) / (b - 1)$, gdzie b jest maksymalną liczbą gałęzi, a d jest maksymalną głębokością drzewa. Należy pamiętać, że zbyt duże drzewo CF może zajmować dużą część zasobów systemowych i mieć negatywny wpływ na wydajność procedury. Każdy węzeł wymaga minimum 16 bajtów.

Aktualizacja modelu grupowania. Ta grupa umożliwia import i aktualizację modelu skupień wygenerowanego w poprzedniej analizie. Plik danych wejściowych zawiera drzewo CF w formacie XML. Po zaimportowaniu model zostaje uaktualniony o dane z aktywnego pliku. Nazwy zmiennych należy wybrać w głównym oknie dialogowym w tej samej kolejności, w jakiej zostały określone w poprzedniej analizie. Plik XML pozostaje niezmienny, o ile dane nowego modelu nie zostaną zapisane pod tą samą nazwą pliku. Więcej informacji zawiera temat [“Dwustopniowe grupowanie: Wyniki”](#) na stronie 188.

Jeśli wybrana zostanie opcja aktualizacji modelu skupień, do wygenerowania drzewa CF zostaną zastosowane opcje określone dla pierwotnego modelu. Należą do nich takie opcje, jak miara odległości, obsługa szumu, alokacja pamięci i ustawienia kryteriów kształtowania drzewa CF, a wszelkie ustawienia dla takich opcji określone w oknach dialogowych zostaną zignorowane.

Uwaga: podczas aktualizacji modelu skupień procedura zakłada, że żadna z wybranych obserwacji pochodzących z aktywnego zbioru danych, nie była wykorzystywana przy tworzeniu pierwotnego modelu. Procedura zakłada również, że obserwacje wykorzystywane przy aktualizacji modelu pochodzą z tej samej populacji, co obserwacje wykorzystane do utworzenia pierwotnego modelu, tzn. zakłada się, że średnie i wariancje zmiennych ilościowych oraz poziomy zmiennych kategoryjnych są takie same w obu zestawach obserwacji. Jeśli „nowy” i „stary” zestaw obserwacji pochodzą z populacji heterogenicznych, dla uzyskania najlepszych rezultatów należy wykonać procedurę Dwustopniowe grupowanie na połączonych zestawach obserwacji.

Dwustopniowe grupowanie: Wyniki

Wynik. Ta grupa zawiera opcje wyświetlania tabel z wynikami grupowania.

- **Tabele przestawne.** Wyniki są wyświetlane w tabelach przestawnych.
- **Wykresy i tabele w przeglądarce modelu.** Wyniki są wyświetlane w przeglądarce modelu.
- **Zmienne ewaluacyjne** Oblicza dane grup dla zmiennych, które nie były użyte przy tworzeniu grupy. Pola ewaluacji mogą być wyświetlone wraz z cechami wejściowymi w przeglądarce modelu po wybraniu ich w podrzędnym oknie Wyświetlanie. Pola bez wartości są ignorowane.

Roboczy plik danych. Ta opcja pozwala na zapisanie zmiennych w aktywnym zbiorze danych.

- **Utwórz zmienną informującą o przynależności do skupień.** Zmienna taka zawiera numer identyfikacyjny skupienia dla każdej obserwacji. Jej nazwa to tsc_n , gdzie n jest dodatnią liczbą całkowitą określającą porządek operacji zapisu aktywnego zbioru danych przez tę procedurę w danej sesji.

Pliki XML. Finalny model skupień i drzewo CF stanowią dwa typy wyników, które można wyeksportować w formacie XML.

- **Eksportuj model finalny.** Finalny model skupień zostaje wyeksportowany do określonego pliku w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .
- **Eksportuj drzewo CF.** Ta opcja umożliwia zapisanie bieżącego stanu drzewa skupienia i zaktualizowanie go w późniejszym czasie przy użyciu nowszych danych.

Przeglądarka skupień

Modele skupień są zwykle wykorzystywane do znajdowania grup (lub skupień), lub podobnych rekordów bazujących na badanych zmiennych w sytuacjach wysokiego podobieństwa między elementami tej samej grupy oraz niskiego podobieństwa między elementami różnych grup. Wyniki można wykorzystać do identyfikacji powiązań, które nie są widoczne w inny sposób. Na przykład dzięki analizie skupień preferencji klientów, poziomu dochodów oraz nawyków nabywczych, możliwe jest zidentyfikowanie typów klientów, którzy z większym prawdopodobieństwem odpowiedzą na określoną kampanię marketingową.

Istnieją dwa podejścia do interpretowania wyników na ekranie skupień:

- Zbadaj skupienia w celu określenia charakterystyki niepowtarzalnego skupienia. *Czy jedno skupienie zawiera wszystkich pożyczkobiorców o wysokich dochodach? Czy to skupienie zawiera więcej rekordów niż inne skupienia?*
- Zbadaj zmienne w skupieniach w celu określenia sposobu rozłożenia wartości po skupieniach. *Czy wykształcenie danej osoby determinuje przynależność do skupienia? Czy wysoka ocena kredytowa powoduje rozróżnienie w zakresie przynależności do jednego lub drugiego skupienia?*

Korzystając z głównych widoków oraz różnych, połączonych widoków w Przeglądarce skupień można uzyskać wiedzę, która pomoże odpowiedzieć na te pytania.

Aby zobaczyć informacje dotyczące modelu skupień, aktywuj (dwukrotnie kliknij) obiekt Przeglądarka modelu w Przeglądarce.

Przeglądarka skupień

Przeglądarka skupień składa się z dwóch paneli, widoku głównego z lewej strony i powiązanego lub dodatkowego widoku z prawej strony. Istnieją dwa główne widoki:

- Podsumowanie modelu (domyślny). Aby uzyskać dodatkowe informacje, patrz temat: [“Widok podsumowania modelu”](#) na stronie 189.
- Grupy. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok skupień”](#) na stronie 190.

Istnieją cztery połączone/dodatkowe widoki:

- Ważność predyktora. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok ważności predyktora skupień”](#) na stronie 191.
- Rozmiary grup (domyślne). Aby uzyskać dodatkowe informacje, patrz temat: [“Widok rozmiarów skupień”](#) na stronie 192.
- Rozkład komórek. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok rozkładu komórek”](#) na stronie 192.
- Porównanie skupień. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok porównania skupień”](#) na stronie 192.

Widok podsumowania modelu

Widok Podsumowanie modelu przedstawia przegląd lub podsumowanie modelu skupień, w uwzględnieniu miary Silhouette spójności i odrębności, która jest zacieniowana w celu wskazania słabych, dostatecznych, lub dobrych wyników. Ten przegląd pozwala na szybkie sprawdzenie, czy jakość jest słaba, kiedy to można zdecydować o powrocie do węzła modelowania w celu korekty ustawień modelu skupień, aby uzyskać lepszy wynik.

Wynik słaby, dostateczny lub dobry bazuje na pracy Kaufmana oraz Rousseeuwa (1990), dotyczącej interpretacji struktur skupień. W widoku podsumowania modelu, dobry wynik równa się danym, które odzwierciedlają ocenę Kaufmana oraz Rousseeuwa jako raczej sensowny lub silny dowód struktury skupienia, dostateczny odzwierciedla ich ocenę słabego dowodu, a słaby odzwierciedla ich ocenę braku istotnego dowodu.

Miara silhouette uśrednia poprzez wszystkie rekordy $(B-A)/\max(A,B)$, gdzie A oznacza odległość rekordu od środka grup, a B oznacza odległość rekordu od najbliższego środka grup, do którego rekord ten nie należy. Współczynnik silhouette o wartości 1 oznacza, że wszystkie obserwacje znajdują się bezpośrednio

w centrach ich skupień. Wartość 1 oznacza, że wszystkie obserwacje znajdują się w środkach grup innych grup. Wartość 0 oznacza, że średnio obserwacje znajdują się w równej odległości od centrum ich własnego skupienia i od najbliższego, innego skupienia.

Podsumowanie obejmuje tabelę zawierającą następujące informacje:

- **Algorytm.** Używany algorytm grupowania, na przykład, "TwoStep" (dwustopniowy).
- **Zmienne wejściowe.** Liczba zmiennych, znanych również jako **wejścia** lub **predyktory**.
- **Grupy.** Liczba skupień w rozwiązaniu.

Widok skupień

Widok skupień zawiera siatkę skupień według zmiennej, która zawiera nazwy, rozmiary i profile poszczególnych skupień.

Kolumny w siatce zawierają następujące informacje:

- **Grupowanie.** Numery skupień utworzone przez algorytm.
- **Etykieta.** Dowlone etykiety zastosowane do każdego skupienia (pole to jest domyślnie puste). Dwukrotnie kliknij komórkę, aby wprowadzić etykietę opisującą zawartość skupienia; na przykład, „Nabywcy luksusowych samochodów”.
- **Opis.** Dowlony opis zawartości skupienia (pole to jest domyślnie puste). Dwukrotnie kliknij komórkę, aby wprowadzić opis skupienia; na przykład, "wiek ponad 55 l., profesjonaliści, zarabiający powyżej 100 000 USD".
- **Rozmiar.** Rozmiar każdego skupienia jako wartość procentowa całego przykładowego skupienia. Każda komórka rozmiaru w siatce przedstawia pasek pionowy, który pokazuje procent rozmiaru w ramach skupienia, procent rozmiaru w formacie liczbowym oraz liczebność obserwacji skupień.
- **Zmienne.** Pojedyncze wejścia lub predyktory, posortowane domyślnie według całkowitej istotności. Jeśli jakieś kolumny mają równe rozmiary, są one wyświetlane w kolejności rosnącej wg numerów skupień.

Całkowita ważność właściwości jest oznaczona kolorem cieniowania tła komórki; najistotniejsza właściwość jest najciemniejsza; najmniej istotna właściwość nie jest cieniowana. Przewodnik nad tabelą wskazuje ważność przypisaną do każdego koloru komórki właściwości.

Po najejchaniu myszką na komórkę wyświetla się pełna nazwa / etykieta właściwości i wartość istotności dla komórki. Możliwe jest wyświetlenie dalszych informacji, zależnie od widoku i typu zmiennej. W widoku Środki grup uwzględniana jest statystyka komórki oraz jej wartość, na przykład: „Średnia: 4.32”. Dla zmiennych jakościowych komórki wyświetlają nazwę najczęstszej (modalnej) kategorii i jej wartość procentową.

W Widoku skupień można wybrać różne sposoby wyświetlania informacji o skupieniach:

- Transponuj grupy i zmienne Aby uzyskać dodatkowe informacje, patrz temat: [“Transponowanie skupień i zmiennych” na stronie 190.](#)
- Sortowanie zmiennych. Aby uzyskać dodatkowe informacje, patrz temat: [“Sortowanie zmiennych” na stronie 191.](#)
- Sortowanie skupień. Aby uzyskać dodatkowe informacje, patrz temat: [“Sortowanie skupień” na stronie 191.](#)
- Wybór zawartości komórki. Aby uzyskać dodatkowe informacje, patrz temat: [“Zawartość komórki” na stronie 191.](#)

Transponowanie skupień i zmiennych

Domyślnie skupienia wyświetlają się jako kolumny, a funkcje wyświetlają się jako wiersze. Aby odwrócić ten sposób wyświetlania, kliknij przycisk **Transponuj grupy i zmienne** z lewej strony przycisków **Sortowanie zmiennych według**. Może się to okazać potrzebne w sytuacji, gdy wyświetlonych jest wiele skupień. W wyniku tego zmniejszy się zakres do przewijania w poziomie w celu obejrzenia danych.

Sortowanie zmiennych

Przyciski **Sortowanie zmiennych według** umożliwiają wybór sposobu wyświetlania komórek zmiennych:

- **Całkowita ważność.** Jest to domyślny porządek sortowania. Zmienne są posortowane w kolejności malejącej według całkowitej istotności, a porządek sortowania jest taki sam we wszystkich skupieniach. Jeśli jakieś zmienne mają powiązane wartości istotności, powiązane zmienne są zestawione rosnąco według nazw zmiennych.
- **Istotność wewnątrzgrupowa.** Zmienne są posortowane według ich istotności dla każdego skupienia. Jeśli jakieś zmienne mają powiązane wartości istotności, powiązane zmienne są zestawione rosnąco według nazw zmiennych. Jeśli wybierze się tę opcję, wówczas zwykle zmienia się porządek sortowania w skupieniach.
- **Name.** Zmienne są posortowane według nazwy w kolejności alfabetycznej.
- **Kolejność danych.** Zmienne są posortowane według ich kolejności w zbiorze danych.

Sortowanie skupień

Domyślnie skupienia są posortowane w porządku malejącym według rozmiaru. Przyciski **Sortowanie grup według** umożliwiają posortowanie skupień według nazw w kolejności alfabetycznej, lub jeśli utworzono niepowtarzalne etykiety, alfanumerycznej kolejności etykiet.

Zmienne posiadające taką samą etykietę są posortowane według nazwy skupienia. Jeśli skupienia są posortowane według etykiety i użytkownik dokona edycji etykiety skupienia, porządek sortowanie zostanie automatycznie zaktualizowany.

Zawartość komórki

Przyciski **Komórki** umożliwiają zmianę sposobu wyświetlania zawartości komórki dla zmiennych i zmiennych ewaluacyjnych.

- **Środki grup.** Domyślnie w komórkach wyświetlają się nazwy/etykiety zmiennych oraz tendencja centralna dla każdej kombinacji skupień/zmiennych. Dla zmiennych ciągłych wyświetla się średnia oraz tryb (najczęściej występująca kategoria) z procentem kategorii dla zmiennych jakościowych.
- **Rozkłady bezwzględne.** Pokazuje nazwy/etykiety zmiennych oraz rozkłady bezwzględne zmiennych w ramach każdego skupienia. Dla zmiennych jakościowych, ekran ten pokazuje wykresy słupkowe, na które są nałożone kategorie uporządkowane w kolejności rosnącej wartości danych. Dla zmiennych ilościowych, ekran ten pokazuje gładki wykres gęstości, który używa tych samych punktów końcowych i odstępów dla każdego skupienia.

Ten stały, czerwony ekran pokazuje rozkład skupień, podczas gdy bledszy ekran przedstawia całkowite dane.

- **Rozkłady względne.** Pokazują przyszłe nazwy/etykiety i rozkłady względne w komórkach. Zasadniczo ekrany te są podobne do tych, które są wyświetlane dla rozkładów bezwzględnych, z tym, że zamiast nich wyświetlane są rozkłady względne.

Ten stały, czerwony ekran wyświetla rozkład skupień, podczas gdy bledszy ekran przedstawia całkowite dane.

- **Widok podstawowy.** Tam, gdzie jest wiele skupień, zobaczenie wszystkich szczegółów może być trudne bez przewijania. Aby zmniejszyć ilość przewijania, należy wybrać ten widok, aby zmienić sposób wyświetlania na bardziej pomniejszoną wersję tabeli.

Widok ważności predyktora skupień

Widok ważności predyktora skupień pokazuje względną ważność każdej zmiennej w ocenie modelu.

Widok rozmiarów skupień

Widok rozmiarów skupień przedstawia wykres kołowy zawierający każde skupienie. W każdym kawałku przedstawiony jest rozmiar procentowy każdego skupienia; najedź myszą na każdy kawałek, aby wyświetlić liczbę w tym kawałku.

Pod wykresem znajduje się tabela zawierająca następujące informacje o rozmiarach:

- Rozmiar najmniejszego skupienia (liczebność i wartość procentowa całości).
- Rozmiar największego skupienia (liczebność i wartość procentowa całości).
- Proporcja rozmiaru największego skupienia do najmniejszego skupienia.

Widok rozkładu komórek

Widok rozkładu komórek przedstawia rozszerzony, bardziej szczegółowy wykres rozkładu danych dla dowolnej komórki zmiennej wybranej w tabeli w panelu głównym Grupy.

Widok porównania skupień

Widok porównania skupień składa się z układu w postaci siatki, ze zmiennymi w wierszach oraz wybranymi skupieniami w kolumnach. Widok ten pomaga lepiej zrozumieć czynniki składające się na skupienia; umożliwia on również przeglądanie różnic między skupieniami nie tylko w porównaniu z całkowitymi danymi, ale również w porównaniu z samymi skupieniami.

Aby wybrać skupienia do wyświetlenia kliknij na górną część kolumny skupienia w panelu głównym Grupy. Użyj opcji Ctrl+kliknięcie lub Shift+kliknięcie, aby zaznaczyć lub odznaczyć więcej niż jedno skupienie do porównania.

Uwaga: do wyświetlenia można wybrać do pięciu skupień.

Grupy są przedstawione w kolejności, w jakiej zostały wybrane, podczas gdy kolejność zmiennych jest określona przez opcję **Sortowanie zmiennych według**. Po wybraniu opcji **Istotność wewnątrzgrupowa**, zmienne są zawsze posortowane według całkowitej istotności.

Wykresy w tle przedstawiają całkowite rozkłady wszystkich zmiennych:

- Zmienne jakościowe są przedstawione jako wykresy punktowe, gdzie rozmiar punktu oznacza najczęstszą/modalną kategorię dla każdego skupienia (według zmiennej).
- Zmienne ilościowe są wyświetlane jako wykresy skrzynkowe, które przedstawiają całkowite mediany i rozstępy ćwiartkowe.

Na widoki w tle nałożone są wykresy skrzynkowe dla wybranych skupień:

- Dla zmiennych ilościowych, znaczniki w postaci kwadratowych punktów oraz linie poziome wskazują rozstęp mediany i rozstęp ćwiartkowy dla każdego skupienia.
- Każdemu skupieniu odpowiada inny kolor, pokazany u góry widoku.

Nawigacja w Przeglądarce skupień


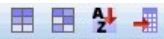

Przeglądarka skupień jest ekranem interaktywnym. Można wykonać następujące czynności:

- Wybierz pole lub klaster, aby wyświetlić więcej szczegółów.
- Porównać skupienia w celu wyboru interesujących nas elementów.
- Zmienić ekran.
- Transponuj osie.

Używanie pasków narzędzi

Informacjami pojawiającymi się w lewym i w prawym panelu można sterować przy pomocy opcji paska narzędzi. Można zmieniać orientację ekranu (z góry na dół, od lewej do prawej, od prawej do lewej) przy pomocy kontrolki paska narzędzi. Ponadto można również przywrócić przeglądarkę do ustawień domyślnych i otworzyć okno dialogowe w celu określenia treści widoku skupień na panelu głównym.

Opcje **Sortowanie zmiennych według**, **Sortowanie grup według**, **Komórki** oraz **Pokaż** są dostępne tylko po wybraniu widoku **Grupy** na panelu głównym. Aby uzyskać dodatkowe informacje, patrz temat: [“Widok skupień”](#) na stronie 190.

Tabela 2. Ikony na pasku narzędzi	
Ikona	Temat
	Patrz Transponuj grupy i zmienne
	Patrz Sortowanie zmiennych według
	Patrz Sortowanie skupień według
	Patrz Komórki

Kontrolowanie wyświetlania widoku skupień

Aby kontrolować, co się wyświetla w widoku skupień na panelu głównym, kliknij przycisk **Wyświetl**; otworzy się okno dialogowe Wyświetl.

Zmienne. Wybrane domyślnie. Aby ukryć wszystkie zmienne wejściowe, odznacz to pole wyboru.

Zmienne ewaluacyjne. Wybierz zmienne ewaluacyjne (zmienne nie służące do tworzenia modelu skupień, ale wysyłane do przeglądarki modelu w celu oceny skupień do wyświetlenia; domyślnie nie wyświetlają się żadne skupienia. *Uwaga* Zmienna ewaluacyjna musi być łańcuchem z więcej niż jedną wartością. To pole wyboru jest niedostępne, jeśli nie są dostępne żadne zmienne ewaluacyjne.

Opisy grup. Wybrane domyślnie. Aby ukryć wszystkie opisy skupień, odznacz to pole wyboru.

Rozmiary grup. Wybrane domyślnie. Aby ukryć wszystkie komórki rozmiarów skupień, odznacz to pole wyboru.

Maksymalna liczba kategorii. Podaj maksymalną liczbę kategorii, które mają się wyświetlać na wykresie zmiennych jakościowych; domyślna ilość to 20.

Filtrowanie rekordów

Aby dowiedzieć się więcej o obserwacjach w danym skupieniu lub w danej grupie skupień, można wybrać podzbiór rekordów do dalszej analizy na podstawie wybranych skupień.

- Wybierz skupienia w widoku skupień Przeglądarki skupień. Aby wybrać wiele skupień, naciśnij klawisz Ctrl i kliknij.
- Z menu wybierz:
Utwórz > Filtruj rekordy...
- Wprowadź nazwę zmiennej filtrującej. W przypadku rekordów z wybranych skupień zmienna ta przyjmie wartość 1. We wszystkich innych rekordach zmienna ta będzie miała wartość 0, zostanie więc wyłączona z dalszej analizy, chyba że zostanie zmieniony status filtra.
- Kliknij **OK**.

Hierarchiczna analiza skupień

Ta procedura umożliwia podjęcie próby identyfikacji względnie jednorodnych grup obserwacji (lub zmiennych) w oparciu o wybraną charakterystykę i z wykorzystaniem algorytmu, który rozpoczynając działanie w sytuacji, kiedy każda obserwacja (lub zmienna) jest skupiona oddzielnie, łączy skupienia aż do momentu, kiedy pozostanie tylko jedno. Możliwa jest analiza surowych zmiennych lub wybranie jednego z licznych przekształceń standaryzujących. Miary odległości lub podobieństwa są generowane przez procedurę Odległości. Na każdym etapie wyświetlane są statystyki, pomagające w wyborze najlepszego rozwiązania.

Przykład. Czy istnieją możliwe do zidentyfikowania grupy programów telewizyjnych, które mają podobną widownię w każdej z grup? Za pomocą hierarchicznej analizy skupień można skupić programy telewizyjne (obserwacje) w jednolite grupy w oparciu o charakterystyki widzów. Można to następnie wykorzystać do identyfikacji segmentów rynku w celach marketingowych. Można też skupić miasta (obserwacje) w jednorodnych grupach, tak aby możliwe było wybranie porównywalnych miast w celu przetestowania różnych strategii marketingowych.

Statystyki. Przegląd aglomeracji, macierz odległości (lub podobieństwa) i przynależność do skupień dla rozwiązania pojedynczego lub przedziału rozwiązań. Wykresy: dendrogramy i wykresy soplekowe.

Wymagania dotyczące danych dla hierarchicznej analizy skupień

Dane. Zmienne mogą być danymi ilościowymi, binarnymi lub liczebnościowymi. Istotne jest skalowanie zmiennych, ponieważ różnice w skalowaniu mogą mieć wpływ na rozwiązanie dla skupień. Jeżeli występują bardzo znaczne różnice w skalowaniu zmiennych (na przykład jedna zmienna jest mierzona w dolarach, a druga w latach), należy rozważyć możliwość ich standaryzacji (można to zrobić automatycznie za pomocą procedury Hierarchiczna analiza skupień).

Kolejność obserwacji. Jeśli w danych wejściowych istnieją wiązane odległości lub podobieństwa lub pojawiają się między uaktualnionymi skupieniami podczas łączenia, otrzymane rozwiązanie grupowania może zależeć od kolejności obserwacji w pliku. Aby zweryfikować stabilność danego rozwiązania może wystąpić konieczność uzyskania kilku różnych rozwiązań przy sortowaniu przy różnej, przypadkowej kolejności obserwacji.

Założenia. Wykorzystywane miary odległości lub podobieństwa powinny być odpowiednie dla rodzaju analizowanych danych (więcej informacji na temat wyboru miar odległości i podobieństwa znajduje się w opisie procedury Odległości). Analiza powinna także zawierać wszystkie odpowiednie zmienne. Rozwiązanie uzyskane w wyniku pominięcia istotnych zmiennych może być mylące. Ponieważ hierarchiczna analiza skupień jest metodą eksploracyjną, wyniki powinny być traktowane jako próbne do czasu potwierdzenia przez analizę niezależnej próby.

Wykonywanie hierarchicznej analizy skupień

1. Wybierz z menu następującą opcję:

Analiza > Klasyfikacja > Hierarchiczna analiza skupień...

2. Przy skupianiu obserwacji należy wybrać co najmniej jedną zmienną numeryczną. Przy skupianiu zmiennych należy wybrać co najmniej trzy zmienne numeryczne.

Opcjonalnie można wybrać zmienną identyfikacyjną do oznaczenia obserwacji etykietami.

Metoda hierarchicznej analizy skupień

Metoda aglomeracji. Dostępne są: średnia odległość między skupieniami, średnia odległość wewnątrz skupień, najbliższe sąsiedztwo, najdalsze sąsiedztwo, środek ciężkości, mediana i metoda Warda.

Miara. Umożliwia określenie miary odległości lub podobieństwa, która będzie wykorzystywana podczas skupiania. Należy wybrać typ danych i odpowiednią miarę odległości lub podobieństwa:

- **Interwałowe.** Dostępne są: odległość euklidesowa, kwadrat odległości euklidesowej, cosinus, korelacja Pearsona, odległość Czebyszewa, odległość miejska, odległość Minkowskiego i odległość użytkownika.
- **Counts.** Dostępne są: odległość chi-kwadrat i odległość phi-kwadrat.
- **Binarne.** Dostępne są: odległość euklidesowa, kwadrat odległości euklidesowej, różnica wielkości, różnica wzoru, wariancja, miara wariancyjna, rozproszenie, kształt, proste zgodności, phi korelacja 4-punktowa, lambda, *D* Anderberga, miara Dice'a, miara Hamanna, miara Jaccarda, miara Kulczyńskiego 1, miara Kulczyńskiego 2, miara Lance'a i Williamsa, miara Ochiai, miara Rogersa i Tanimoto, miara Russela i Rao, miara Sokala i Sneatha 1, miara Sokala i Sneatha 2, miara Sokala i Sneatha 3, miara Sokala i Sneatha 4, miara Sokala i Sneatha 5, *Y* Yule'a i *Q* Yule'a.

Przekształcanie wartości. Umożliwia standaryzację wartości danych albo dla obserwacji, albo wartości przed wyliczeniem odległości (nie dostępne dla danych binarnych). Dostępne metody standaryzacji to: wartości statystyki *z*, zakres od -1 do 1, zakres od 0 do 1, maksymalna wartość równa 1, średnia równa 1 i odchylenie standardowe równe 1.

Transformacja miar. Umożliwia przekształcenie wartości generowanych przez miarę odległości. Są one stosowane po wyliczeniu miary odległości. Dostępne alternatywny to: wartości bezwzględne, zmiana znaku i przeskalowanie na zakres 0-1.

Hierarchiczna analiza skupień: Statystyki

Przegląd aglomeracji. Powoduje wyświetlenie obserwacji lub skupień łączonych na każdym etapie, odległości między łączonymi obserwacjami lub skupieniami i ostatniego poziomu skupień na którym obserwacja (lub zmienna) została dołączona do skupienia.

Macierz odległości. Podaje odległości lub podobieństwa między elementami.

Przynależność do skupień. Powoduje wyświetlenie skupienia, do którego jest przypisana każda obserwacja na jednym lub kilku etapach łączenia skupień. Dostępne opcje to: rozwiązanie pojedyncze i zakres rozwiązań.

Hierarchiczna analiza skupień: Wykresy

Dendrogram. Powoduje wyświetlenie *dendrogramu*. Dendrogramy można wykorzystywać w celu oceny spójności uformowanych skupień. Mogą też one dostarczać informacji o liczbie skupień, które należy utrzymać.

Wykres sopelkowy. Powoduje wyświetlenie *wykresu sopelkowego*, zawierającego wszystkie skupienia lub określony zakres skupień. Wykresy sopelkowe zawierają informacje o sposobie łączenia obserwacji w skupienia przy każdej iteracji analizy. Opcja orientacji umożliwia wybór wykresu pionowego lub poziomego.

Hierarchiczna analiza skupień: Zapisz zmienne wynikowe

Przynależność do skupień. Umożliwia zapisanie przynależności do skupień dla pojedynczego rozwiązania lub zakresu rozwiązań. Zapisane zmienne można następnie wykorzystywać w późniejszych analizach do badania innych różnic między grupami.

Dodatkowe właściwości składni komendy CLUSTER

Procedura hierarchiczna skupień stosuje składnię komendy CLUSTER. Język składni komend umożliwia również:

- Stosowanie kilku metod grupowania w jednej analizie.
- Odczytywanie i analizę macierzy bliskości.
- Zapisywanie macierzy bliskości do późniejszej analizy.
- Określanie wartości potęgi i pierwiastka z niestandardowych (potęgowych) miar odległości.
- Określanie nazw zapisanych zmiennych.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Analiza skupień metodą k-średnich

Ta procedura umożliwia podjęcie próby identyfikacji względnie jednorodnych grup obserwacji w oparciu o wybraną charakterystykę i z wykorzystaniem algorytmu umożliwiającego obsługę dużej liczby obserwacji. Zastosowanie algorytmu wymaga jednak od użytkownika określenia liczby skupień. Jeśli użytkownik zna wstępne centra skupień, to może je określić. Można wybrać jedną lub dwie metody klasyfikacji obserwacji, które iteracyjnie aktualizują centra skupień lub tylko je klasyfikują. Można zapisać przynależność do skupień, informacje o odległości i ostateczne centra skupień. Opcjonalnie można określić zmienną, której wartości są wykorzystywane do nadawania etykiety wynikowi obserwacji. Można również przeprowadzić analizę statystyki F wariancji. Chociaż te statystyki są oportunistyczne (podjęta zostaje próba uformowania różniących się znacznie między sobą grup), to względny rozmiar statystyk dostarcza informacji na temat udziału każdej zmiennej w podziale grup.

Przykład. Czy istnieją możliwe do zidentyfikowania grupy programów telewizyjnych, które mają podobną widownię? Za pomocą analizy skupień metodą k -średnich można skupić programy telewizyjne (obserwacje) w k jednolitych grup w oparciu o charakterystyki widzów. Można to następnie wykorzystać do identyfikacji segmentów rynku w celach marketingowych. Można też skupić miasta (obserwacje) w jednorodnych grupach, tak aby możliwe było wybranie porównywalnych miast w celu przetestowania różnych strategii marketingowych.

Statystyki. Kompletne rozwiązanie: wstępne centra skupień, tabela ANOVA. Każda obserwacja: Informacje o skupieniach, odległość od centrum skupienia.

Wymagania dotyczące danych do analizy skupień metodą k -średnich

Dane. Zmienne powinny być ilościowe na poziomie interwałowym lub ilorazowym. Jeśli zmienne są liczebnościami lub mają charakter binarny, to należy użyć procedury hierarchicznej analizy skupień.

Kolejność obserwacji i wstępnych centrów skupień. Domyślny algorytm do wybierania wstępnych centrów skupień nie jest niezmienniczy wobec kolejności obserwacji. Opcja **Użyj średnich ruchomych** w oknie dialogowym Iteracja pozwala na potencjalne uzależnienie wynikającego rozwiązania od kolejności obserwacji, bez względu na to, jak wybrano wstępne centra skupień. Jeśli używana jest jedna z tych metod, może wystąpić konieczność uzyskania kilku różnych rozwiązań przy sortowaniu przy różnej, przypadkowej kolejności obserwacji w celu sprawdzenia stabilności danego rozwiązania. Określenie wstępnych centrów skupień i nie korzystanie z opcji **Użyj średnich ruchomych** pozwoli uniknąć kwestii związanych z kolejnością obserwacji. Ustawianie kolejności wstępnych centrów skupień może jednak mieć wpływ na rozwiązanie, jeśli istnieją związane odległości między obserwacjami a centrami skupień. Aby ocenić stabilność danego rozwiązania, można porównać wyniki z analizy z innymi permutacjami wstępnych wartości środka.

Założenia. Odległości są obliczane z wykorzystaniem prostej odległości euklidesowej. Jeśli ma być wykorzystana inna miara odległości lub podobieństwa, to należy użyć procedury hierarchicznej analizy skupień. Ważnym zagadnieniem jest skalowanie zmiennych. Jeśli zmienne użytkownika są mierzone według odmiennych skal (na przykład jedna zmienna wyrażana jest w dolarach, a inna w latach), to wyniki mogą być błędne. W takich przypadkach przed wykonaniem analizy skupień metodą k -średnich należy rozważyć standaryzację zmiennych (można tego dokonać w procedurze Statystyki opisowe). Procedura opiera się na założeniu, że została wybrana odpowiednia liczba skupień i zostały uwzględnione wszystkie istotne zmienne. Jeśli wybrano nieodpowiednią liczbę skupień lub pominięto ważne zmienne, to otrzymane wyniki mogą być błędne.

Wykonanie analizy skupień metodą k -średnich

1. Z menu wybierz:

Analiza > Klasyfikacja > Analiza skupień metodą k -średnich...

2. Wybierz zmienne, które mają być użyte w analizie skupień.

3. Określ liczbę skupień (liczba skupień musi wynosić co najmniej 2 i nie może przekraczać liczby obserwacji w pliku danych).

4. Wybierz metodę **Iteracja i klasyfikacja** lub metodę **Tylko klasyfikacja**.

5. Opcjonalnie wybierz zmienną identyfikacyjną do opisu obserwacji.

Efektywność analizy skupień metodą k -średnich

Komenda analizy skupień metodą k -średnich jest efektywna głównie dlatego, że nie wylicza odległości pomiędzy wszystkimi parami obserwacji, jak dzieje się to w przypadku wielu innych algorytmów wykorzystywanych w analizie skupień, włączając w to algorytmy wykorzystywane przez komendę hierarchicznej analizy skupień.

Dla uzyskania maksymalnej efektywności należy postąpić próbą obserwacji i zastosować metodę **Iteracja i klasyfikacja** w celu ustalenia centrów skupień. Zaznacz opcję **Zapisz ostatecznie jako**. Następnie przywróć cały plik danych i wybierz opcję **Tylko klasyfikacja** jako metodę i wybierz opcję **Wczytaj wstępne z** aby sklasyfikować całego pliku przy użyciu centów, które są oszacowane z próby. Można zapisywać i odczytywać plik i zbiór danych. Zbiory danych są dostępne do późniejszego użytku w tej samej sesji lecz nie są zapisywane jako pliki, jeśli nie zostaną wprost zapisane pod koniec sesji.

Nazwy zbiorów danych muszą być zgodne z regułami nazewnictwa zmiennych. Aby uzyskać dodatkowe informacje, patrz temat .

Analiza skupień metodą k-średnich: Iteracja

Uwaga: te opcje są dostępne tylko w przypadku wyboru metody **Iteracja i klasyfikacja** w oknie dialogowym Analiza skupień metodą k-średnich.

Maksymalna liczba powtórzeń. Ogranicza liczbę iteracji w algorytmie k-średnich. Iteracja jest zatrzymywana po podanej liczbie iteracji, nawet jeśli kryterium zbieżności nie jest spełnione. Dana liczba musi zawierać się pomiędzy 1 i 999.

Aby odtworzyć algorytm wykorzystywany przez komendę Quick Cluster w wersjach programu wcześniejszych niż wersja 5.0, należy ustalić **maksymalną liczbę iteracji** na 1.

Kryterium zbieżności. Określa moment zatrzymania iteracji. Kryterium to reprezentuje proporcję minimalnej odległości pomiędzy wstępnymi centrami skupień, tak więc jego wartość musi być większa niż 0 i nie większa niż 1. Na przykład jeśli wartość kryterium wynosi 0,02, iteracja zostaje zatrzymana, kiedy wykonana iteracja nie przesuwa żadnego z centrow skupień o odległość większą niż 2% najmniejszej odległości pomiędzy dowolnymi wstępnymi centrami skupień.

Użyj średnich ruchomych. Pozwala na aktualizowanie centrów skupień po przydzieleniu każdej z obserwacji. Jeśli opcja ta nie będzie zaznaczona, to nowe centra skupień będą obliczane po przydzieleniu wszystkich obserwacji.

Zapisywanie analizy skupień metodą k-średnich

Informacje o rozwiązaniu można zapisać jako nowe zmienne, które mogą być wykorzystane w kolejnych analizach:

Przynależność do skupień. Umożliwia utworzenie nowej zmiennej wskazującej ostateczną przynależność do skupień każdej obserwacji. Wartości nowej zmiennej wahają się od 1 do liczby skupień.

Odległość od centrum skupienia. Umożliwia utworzenie nowej zmiennej wskazującej odległość euklidesową pomiędzy każdą z obserwacji a jej centrum klasyfikacji.

Analiza skupień metodą k-średnich: Opcje

Statystyki. Użytkownik może wybrać następujące statystyki: wstępne centra skupień, tabela ANOVA oraz informacja o skupieniach dla każdego obiektu.

- *Wstępne centra skupień.* Pierwsze oszacowanie średnich średnich dla każdego z klastrów. Przy ustawieniach domyślnych program wybiera z danych taką liczbę odpowiednio oddalonych od siebie obserwacji, która jest równa liczbie skupień. Wstępne centra skupień zostają użyte w pierwszej, wstępnej klasyfikacji, a następnie korygowane są w procesie iteracyjnym.
- *Tabela ANOVA.* Wyświetla tabelę analizy wariancji, która zawiera testy F jednej zmiennej dla każdej zmiennej klastrowej. Testy F mają jedynie charakter opisowy i wynikające z nich prawdopodobieństwa nie powinny być interpretowane. Tabela ANOVA nie jest wyświetlana, gdy wszystkie obserwacje są przypisane do jednego skupienia.
- *Informacje o klastrze dla każdej obserwacji.* Służy do wyświetlania dla każdego przypadku końcowego przypisania klastra oraz odległości euklidesowej między przypadkiem a centrum skupienia używanym do sklasyfikowania sprawy. Ponadto wyświetla odległość euklidesową między końcowymi centrami skupień.

Braki danych. Dostępными opcjami są **Wyłączanie wszystkich obserwacji z brakami** i **Wyłączanie obserwacji parami**.

- **Wyłączanie wszystkich obserwacji z brakami.** Umożliwia wyłączenie z analizy obserwacji z brakami danych dla dowolnej zmiennej skupiającej.
- **Wyłączanie obserwacji parami.** Umożliwia przypisywanie obserwacji do skupień na podstawie odległości obliczonych ze wszystkich zmiennych bez braków danych.

Dodatkowe właściwości komendy QUICK CLUSTER

Procedura analizy skupień metodą k-średnich stosuje składnię komendy CLUSTER. Język składni komend umożliwia również:

- Wybieranie pierwszych obserwacji k jako centrów skupień, co pozwala na uniknięcie pobrania dodatkowych danych, które zazwyczaj potrzebne są do oszacowania ich.
- Określanie wstępnych centrów skupień bezpośrednio jako część składni komend.
- Określanie nazw zapisanych zmiennych.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Testy nieparametryczne

Testy nieparametryczne tworzą minimalne założenia dotyczące rozkładu danych. Testy dostępne w tych oknach dialogowych są pogrupowane na trzy szerokie kategorie oparte na sposobie organizacji danych:

- Test dla jednej próby analizuje jedną zmienną.
- Test dla prób zależnych porównuje co najmniej dwie zmienne w przypadku tego samego zestawu obserwacji.
- Test dla prób niezależnych analizuje jedną zmienną zgrupowaną wg kategorii innej zmiennej.

Testy nieparametryczne dla jednej próby

Testy nieparametryczne dla jednej próby identyfikują różnice w pojedynczych zmiennych przy pomocy jednego lub więcej testów nieparametrycznych. Testy nieparametryczne nie zakładają rozkładu normalnego danych.

Jaki jest cel? Cele pozwalają na szybkie określenie różnych, ale powszechnie używanych ustawień testów.

- **Automatyczne porównanie danych empirycznych z hipotetycznymi.** Ten cel przeprowadza Test dwumianowy dla zmiennych jakościowych, z wykorzystaniem tylko dwóch kategorii, test chi-kwadrat dla wszystkich innych zmiennych jakościowych i test Kolmogorowa-Smirnowa dla zmiennych ciągłych.
- **Test sekwencji na losowość.** Ten cel wykorzystuje Test serii w celu przetestowania sekwencji empirycznej wartości danych na losowość.
- **Analiza użytkownika.** Należy wybrać tę opcję chcąc ręcznie skorygować ustawienia testu w zakładce Ustawienia. Należy zwrócić uwagę, że to ustawienie jest zaznaczone automatycznie, gdy później dokona się zmian w zakładce Ustawienia w opcjach, które są niekompatybilne z obecnie wybranym celem.

Wykonywanie testów nieparametrycznych dla jednej próby

Z menu wybierz:

Analiza > Testy nieparametryczne > Jedna próba...

1. Kliknij opcję **Uruchom**.

Opcjonalnie można wykonać następujące czynności:

- Określ cel w zakładce Cel.
- Określ przypisania zmiennych w zakładce Zmienne.
- Określ ustawienia zaawansowane na karcie Ustawienia.

Zakładka Zmienne

Zakładka Zmienne określa zmienne do przetestowania.

Użyj wstępnie zdefiniowanych ról. Ta opcja wykorzystuje istniejące informacje i zmiennych. Wszystkie zmienne ilościowe ze wstępnie zdefiniowaną rolą jako Wejście lub Łącznie będą używane jako zmienne testowe. Wymagana jest co najmniej jedna zmienna testowa. .

Użyj niestandardowych przypisań. Opcja ta pozwala na nadpisanie ról zmiennych. Po wybraniu tej opcji, określ poniższe zmienne:

- **Testuj zmienne.** Wybierz co najmniej jedną zmienną.

Zakładka Ustawienia

Zakładka Ustawienia zawiera wiele różnych grup ustawień, które można zmieniać w celu precyzyjnego określenia sposobu przetwarzania danych użytkownika przez algorytm. Jeśli dokona się jakichkolwiek zmian w ustawieniach domyślnych, które są niekompatybilne z obecnie wybranym celem, zakładka Cel zostanie automatycznie zaktualizowana do zaznaczenia opcji **Analiza niestandardowa**.

Wybierz testy

Ustawienia te określają testy do wykonania w zmiennych podanych w zakładce Zmienne.

Automatycznie wybierz testy odpowiednie do danych. To ustawienie przeprowadza Test dwumianowy dla zmiennych jakościowych, z wykorzystaniem tylko dwóch kategorii, test chi-kwadrat dla wszystkich innych zmiennych jakościowych i test Kołmogorowa-Smirnowa dla zmiennych ciągłych.

Pozwól użytkownikowi wybrać testy. To ustawienie pozwala na wybór określonych testów do wykonania.

- **Porównanie prawdopodobieństw dla obserwowanych dychotomicznych z hipotetycznymi (test dwumianowy).** Test dwumianowy można wykonać dla wszystkich zmiennych. Wybór tej opcji powoduje przeprowadzenie testu dla jednej próby, który sprawdza, czy rozkład empiryczny zmiennej flagi (zmienna jakościowa z co najmniej dwoma kategoriami) jest taki sam, jakie są oczekiwania wobec podanego rozkładu dwumianowego. Ponadto można zażądać podania przedziałów ufności. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat [“Test dwumianowy: Opcje”](#) na stronie 199.
- **Porównanie prawdopodobieństw empirycznych z hipotetycznymi (test chi-kwadrat).** Test chi-kwadrat stosuje się do zmiennych nominalnych i zmiennych porządkowych. Opcja ta powoduje wykonanie testu dla jednej próby, który wylicza statystykę chi-kwadrat na podstawie różnic między empirycznymi i oczekiwanymi częstotliwościami kategorii zmiennej. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat [“Test chi-kwadrat: Opcje”](#) na stronie 200.
- **Testowanie rozkładu empirycznego względem hipotetycznego (Kołmogorow-Smirnow).** Test Kołmogorowa-Smirnowa stosuje się do zmiennych ciągłych i porządkowych. Opcja ta powoduje przeprowadzenie testu dla jednej próby pod kątem tego, czy funkcja skumulowanego rozkładu próby dla zmiennej jest jednorodna z jednostajnym, normalnym rozkładem Poisson, czy z rozkładem wykładniczym. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat [“Test Kołmogorowa-Smirnowa: Opcje”](#) na stronie 200.
- **Porównanie mediany z wartością hipotetyczną (test Wilcoxona znakowanych rang).** Test Wilcoxona znaków rangowanych stosuje się do zmiennych ciągłych i porządkowych. Opcja ta powoduje przeprowadzenie testu wartości mediany zmiennej dla jednej próby. Podaj jakąś liczbę jako hipotetyczną medianę.
- **Test sekwencji na losowość (test serii).** Test serii stosuje się dla wszystkich zmiennych. Powoduje on przeprowadzenie testu dla jednej próby pod kątem tego, czy sekwencja wartości podzielonych zmiennych jest losowa. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat [“Opcje testu serii”](#) na stronie 201.

Test dwumianowy: Opcje

Test dwumianowy jest przeznaczony dla zmiennych flagi (zmienne jakościowe posiadające tylko dwie kategorie), ale stosuje się go do wszystkich zmiennych przy pomocy reguł do definiowania "sukcesów".

Proporcja hipotetyczna. Określa oczekiwaną proporcję rekordów zdefiniowanych jako "sukcesy" lub p . Podaj wartość większą od 0 i mniejszą od 1. Wartością domyślną jest 0,5.

Oszacowanie przedziału ufności. Dostępne są poniższe metody do wyliczania przedziałów ufności dla danych binarnych:

- **Cloppera-Pearsona (dokładny).** Dokładny przedział bazujący na skumulowanym rozkładzie dwumianowym.
- **Jeffreya.** Przedział Bayesowski, bazujący na rozkładzie a posteriori p przy pomocy prawdopodobieństwa a priori Jeffreya.
- **Iloraz wiarygodności.** Przedział bazujący na funkcji prawdopodobieństwa dla p .

Definiuj sukces dla zmiennych jakościowych. Opcja ta określa, w jaki sposób definiuje się "sukces", wartości danych przetestowane względem hipotetycznej proporcji, dla zmiennych jakościowych.

- **Pierwsza kategoria znaleziona w danych** wykonuje test dwumianowy przy pomocy pierwszej wartości znalezionej próbie do zdefiniowania "sukcesu". Opcja ta ma zastosowanie tylko dla zmiennych nominalnych lub porządkowych posiadających tylko dwie wartości; wszystkie pozostałe zmienne jakościowe podane w zakładce Zmienne, w której wykorzystana jest ta opcja, nie będą testowane. Jest to wartość domyślna.
- **Określ wartości sukcesu** wykonuje test dwumianowy przy pomocy określonej listy wartości do zdefiniowania "sukcesu". Podaj listę wartości łańcuchowych lub liczbowych. Wartości na liście nie muszą występować w próbie.

Definiuj sukces dla zmiennych ilościowych. Opcja ta określa, w jaki sposób definiuje się "sukces", wartości danych przetestowane względem wartości testowej, dla zmiennych ciągłych. Sukces definiuje się jako wartości równe lub mniejsze niż punkt podziału.

- **Punkt środkowy próby** ustawia punkt podziału pośrodku, między wartością minimalną a wartością maksymalną.
- **Niestandardowy punkt podziału** pozwala użytkownikowi na określenie wartości dla punktu podziału.

Test chi-kwadrat: Opcje

Wszystkie kategorie z równym prawdopodobieństwem. Ta opcja tworzy takie same częstotliwości we wszystkich kategoriach w próbie. Jest to wartość domyślna.

Określone prawdopodobieństwo oczekiwane. Opcja ta pozwala na określenie nierównych częstotliwości dla określonej listy kategorii. Podaj listę wartości łańcuchowych lub liczbowych. Wartości na liście nie muszą występować w próbie. W kolumnie **Kategoria**, podaj wartości kategorii. W kolumnie **Częstość względna**, podaj wartość większą niż 0 dla każdej kategorii. Niestandardowe częstotliwości są traktowane jak ilorazy tak, aby podając na przykład częstotliwości 1, 2 i 3 było równe podaniu częstotliwości 10, 20 i 30, co w obu przypadkach oznacza, że według oczekiwań 1/6 rekordów będzie się zaliczała do pierwszej kategorii, 1/3 do drugiej, a 1/2 do trzeciej. Kiedy poda się niestandardowe, oczekiwane prawdopodobieństwa, wartości niestandardowej kategorii muszą obejmować wszystkie wartości zmiennych w danych; w przeciwnym razie dla tej zmiennej test nie zostanie przeprowadzony.

Test Kołmogorowa-Smirnowa: Opcje

To okno dialogowe określa rozkłady do przetestowania i parametry hipotetycznych rozkładów.

Gdy niektóre parametry rozkładu muszą być oszacowane z próby, test Kołmogorowa-Smirnowa nie ma już zastosowania. W tych przypadkach można użyć testu Lillieforsa do oszacowania wartości p z wykorzystaniem losowania Monte Carlo — by przetestować normalność rozkładu z nieznaną średnią i wariancją. Test Lillieforsa ma zastosowanie do trzech rozkładów ciągłych (**Normalny**, **Wykładniczy** **Jednostajny**). Należy zauważyć, że test nie ma zastosowania, jeśli rozkład bazowy jest dyskretny (**Poissona**). Jeśli nie zostaną określone odpowiednie parametry rozkładu, test jest zdefiniowany tylko dla wnioskowania dla jednej próby.

Normalny

W przypadku wybrania opcji **Użyj danych losowych** program wykorzystuje obserwowaną średnią i odchylenie standardowe i udostępnia opcje wyboru istniejących wyników testu **Test asymptotyczny** lub użycia testu **Test Lillieforsa na podstawie losowania Monte Carlo**. Opcja **Użytkownika** umożliwia określenie wartości.

Jednostajny

W przypadku opcji **Użyj danych losowych** program wykorzystuje obserwowane minimum i maksimum i test Lillieforsa na podstawie losowania Monte Carlo. Opcja **Użytkownika** umożliwia określenie wartości minimalnej i maksymalnej.

Wykładniczy

W przypadku opcji **Średnia z próby** program wykorzystuje obserwowaną średnią i test Lillieforsa na podstawie losowania Monte Carlo. Opcja **Użytkownika** umożliwia określenie obserwowanej wartości średniej.

Poissona

Opcja **Średnia** pozwala na określenie obserwowanej wartości średniej.

Opcje testu serii

Test serii jest przeznaczony dla zmiennych flagi (zmienne jakościowe posiadające tylko dwie kategorie), ale stosuje się go do wszystkich zmiennych przy pomocy reguł do definiowania grup.

Zdefiniuj grupy dla zmiennych jakościowych. Dostępne są następujące opcje:

- **Występują tylko 2 kategorie w próbie** przeprowadza test serii przy pomocy wartości znalezionych w próbie w celu zdefiniowania grup. Opcja ta ma zastosowanie tylko dla zmiennych nominalnych lub porządkowych posiadających tylko dwie wartości; wszystkie pozostałe zmienne jakościowe podane w zakładce Zmienne, w której wykorzystana jest ta opcja, nie będą testowane.
- **Rekoduj dane do 2 kategorii** przeprowadza test serii przy pomocy określonej listy wartości w celu zdefiniowania jednej z grup. Wszystkie pozostałe wartości w próbie definiują inną grupę. Nie wszystkie wartości na liście muszą być obecne w próbie, ale w każdej grupie musi się znajdować co najmniej jeden rekord.

Definiuj punkt podziału dla zmiennych ilościowych. Opcja ta określa sposób zdefiniowania grup dla zmiennych ciągłych. Pierwszą grupę definiuje się jako wartości równe lub mniejsze niż punkt podziału.

- **Mediana z próby** ustawia punkt podziału na medianie z próby.
- **Średnia z próby** ustawia punkt podziału na średniej z próby.
- **Niestandardowy** pozwala użytkownikowi na określenie wartości dla punktu podziału.

Opcje testu

Poziom istotności

Określa poziom istotności (alfa) dla wszystkich testów. Podaj wartość liczbową między 0 a 1. Wartość domyślna do 0,05.

Przedział ufności (%)

Określa on poziom ufności dla wszystkich utworzonych przedziałów ufności. Podaj wartość liczbową z zakresu od 0 do 100. Wartością domyślną jest 95.

Obserwacje wykluczone

Określa sposób ustalenia bazy obserwacji dla testów.

Wyłączanie obserwacji test po teście

Rekordy z brakami danych dla pola, które jest używane w konkretnym teście, są pomijane w tym teście. Gdy poda się wiele testów w analizie, każdy test jest oceniany oddzielnie.

Wyklucz wszystkie obserwacje z brakami

Rekordy z brakami danych dla dowolnej zmiennej, która jest nazwana na karcie Zmienne, są wyłączone z wszystkich analiz.

Losowanie Monte Carlo

Gdy niektóre parametry rozkładu muszą być oszacowane z próby, test Kołmogorowa-Smirnowa nie ma już zastosowania. W tych przypadkach można użyć testu Lillieforsa do oszacowania wartości p z wykorzystaniem losowania Monte Carlo — by przetestować normalność rozkładu z nieznaną średnią i wariancją. Test Lillieforsa ma zastosowanie do trzech rozkładów ciągłych (**Normalny**, **Wykładniczy** **Jednostajny**). Należy zauważyć, że test nie ma zastosowania, jeśli rozkład bazowy

jest dyskretny (**Poissona**). Jeśli nie zostaną określone odpowiednie parametry rozkładu, test jest zdefiniowany tylko dla wnioskowania dla jednej próby.

Wartość startowa generatora liczb losowych

Gdy ta opcja jest włączona, dostępna jest opcja ustawienia wartości **Wartość startowa generatora liczb losowych** używanej do losowania Monte Carlo. Wartość musi być pojedynczą liczbą całkowitą z zakresu od 1 do 2.147.483.647 (wartością domyślną jest 2.000.000). Domyślnie ustawienie to nie jest włączone, co powoduje utworzenie losowej wartości początkowej.

Liczba próbek

Ustawia liczbę replik w losowaniu Monte Carlo na potrzeby testu Lillieforsa. Wartość musi być jedną liczbą całkowitą z zakresu od 100 do największej liczby całkowitej. Wartością domyślną jest 10 000.

Poziom ufności symulacji (%)

Resetuje szacowany poziom przedziału ufności testu Kołmogorowa-Smirnowa. Wartość musi być jedną liczbą z zakresu od 0 do 100. Wartością domyślną jest 99.

Braki danych zdefiniowane przez użytkownika

Braki danych użytkownika dla zmiennych jakościowych. Zmienne jakościowe muszą posiadać prawidłowe wartości dla rekordu, który ma zostać zawarty w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako prawidłowe wśród zmiennych jakościowych. Systemowe braki danych i brakujące wartości zmiennych ciągłych są zawsze traktowane jako nieprawidłowe.

Dodatkowe właściwości komendy NPTESTS

Język składni komend umożliwia również:

- Określ test dla jednej próby, test dla prób niezależnych i test dla prób zależnych w pojedynczym przebiegu procedury.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Testy nieparametryczne dla prób niezależnych

Testy nieparametryczne dla prób niezależnych identyfikują różnice między dwoma lub więcej grupami przy pomocy jednego lub więcej testów nieparametrycznych. Testy nieparametryczne nie zakładają rozkładu normalnego danych.

Jaki jest cel? Cele pozwalają na szybkie określenie różnych, ale powszechnie używanych ustawień testów.

- **Automatyczne porównanie rozkładów w grupach.** Ten cel wykonuje test Manna-Whitneya dla danych z 2 grupami lub jednoczynnikową analizę wariancji Kruskal-Wallisa dla danych z k grup.
- **Porównanie median w grupach.** Ten cel wykorzystuje Test mediany do porównywania zaobserwowanych median w grupach.
- **Analiza użytkownika.** Należy wybrać tę opcję chcąc ręcznie skorygować ustawienia testu w zakładce Ustawienia. Należy zwrócić uwagę, że to ustawienie jest zaznaczone automatycznie, gdy później dokona się zmian w zakładce Ustawienia w opcjach, które są niekompatybilne z obecnie wybranym celem.

Uzyskiwanie testów nieparametrycznych dla jednej próby

Z menu wybierz:

Analiza > Testy nieparametryczne > Próby niezależne...

1. Kliknij opcję **Uruchom**.

Opcjonalnie można wykonać następujące czynności:

- Określ cel w zakładce Cel.

- Określ przypisania zmiennych w zakładce Zmienne.
- Określ ustawienia zaawansowane w zakładce Zaawansowane.

Zakładka Zmienne

Zakładka Zmienne określa, które zmienne mają zostać przetestowane oraz zmienne używane do definiowania grup.

Użyj wstępnie zdefiniowanych ról. Ta opcja wykorzystuje istniejące informacje i zmiennych. Wszystkie zmienne ciągłe i porządkowe z wstępnie zdefiniowaną rolą jako Celem lub Łącznie będą używane jako zmienne testowe. Jeśli występuje pojedyncza zmienna jakościowa ze wstępnie zdefiniowaną rolą jako Wejście, będzie ona używana jako zmienna grupowania. W przeciwnym razie żadna zmienna grupowania nie będzie używana domyślnie i użytkownik będzie musiał użyć przypisać zmiennych użytkownika. Wymagana jest co najmniej jedna zmienna testowa i jedna zmienna grupowania. .

Użyj niestandardowych przypisań. Opcja ta pozwala na nadpisanie ról zmiennych. Po wybraniu tej opcji, określ poniższe zmienne:

- **Testuj zmienne.** Wybierz co najmniej jedną zmienną ciągłą lub porządkową.
- **Grupy.** Wybierz zmienną jakościową.

Zakładka Ustawienia

Zakładka ustawienia zawiera wiele różnych grup ustawień, które można zmieniać w celu precyzyjnego określenia sposobu przetwarzania danych użytkownika przez algorytm. Jeśli dokona się jakichkolwiek zmian w ustawieniach domyślnych, które są niekompatybilne z obecnie wybranym celem, zakładka Cel zostanie automatycznie zaktualizowana do zaznaczenia opcji **Analiza niestandardowa**.

Wybierz testy

Ustawienia te określają testy do wykonania w zmiennych podanych w zakładce Zmienne.

Automatycznie wybierz testy odpowiednie do danych. To ustawienie wykonuje test Manna-Whitneya dla danych z 2 grupami, lub jednoczynnikową analizę wariancji Kruskal-Wallisa dla danych z k grup.

Pozwól użytkownikowi wybrać testy. To ustawienie pozwala na wybór określonych testów do wykonania.

- **Porównanie rozkładów w grupach.** Powoduje wykonanie testów dla prób niezależnych sprawdzających, czy próby pochodzą z tej samej populacji.

Test U Manna-Whitney'a (2 próby) używa rangi każdej obserwacji w celu sprawdzenia, czy grupy zostały pobrane z tej samej populacji. Pierwsza wartość definiuje pierwszą grupę w kolejności rosnącej zmiennych grupowania, a druga definiuje drugą grupę. Jeśli zmienna grupowania ma więcej niż dwie wartości, ten test nie jest wykonywany.

Test Kołmogorowa-Smirnowa (2 próby) jest czuły na wszelkie różnice w wartościach median, rozproszenia, skośności itd. między tymi dwoma rozkładami. Jeśli zmienna grupowania ma więcej niż dwie wartości, ten test nie jest wykonywany.

Test sekwencji na losowość (test Walda-Wolfowitza dla 2 prób) przeprowadza test serii z użyciem przynależności do grupy jako kryterium. Jeśli zmienna grupowania ma więcej niż dwie wartości, ten test nie jest wykonywany.

Jednoczynnikowa analiza wariancji Kruskala-Wallisa (k prób) jest rozszerzeniem testu Manna-Whitneya i nieparametrycznym odpowiednikiem jednoczynnikowej analizy wariancji. Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**.

Test na uporządkowane alternatywy (Jonckheere-Tespstra dla k prób) jest wydajniejszą alternatywą dla testu Kruskala-Wallisa, kiedy k prób ma naturalne uporządkowanie. Na przykład niech k populacji reprezentuje k rosnących temperatur. Testowana jest hipoteza, że różne temperatury powodują ten sam rozkład reakcji. Hipoteza alternatywna brzmi, że w miarę wzrostu temperatury wzrasta wielkość

reakcji. Test Jonckheere-Terpstra jest najbardziej odpowiedni, ponieważ w tym przypadku hipoteza alternatywna jest uporządkowana. **Od najmniejszych** określa alternatywną hipotezę, że parametr położenia pierwszej grupy jest mniejszy lub równy parametrowi drugiej grupy, który z kolei jest mniejszy lub równy parametrowi trzeciej grupy itd. **Od największych** określa alternatywną hipotezę, że parametr położenia pierwszej grupy jest większy lub równy parametrowi drugiej grupy, który z kolei jest większy lub równy parametrowi trzeciej grupy itd. W obu przypadkach w alternatywnej hipotezie zakłada się także, że żadne położenia nie są sobie równe. Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**.

- **Porównanie przedziałów w grupach.** Opcja ta wykonuje testy prób niezależnych pod kątem tego, czy próby mają ten sam zakres. **Test Mosesa skrajnych reakcji (2 próby)** testuje grupę kontrolną w porównaniu z grupą porównawczą. Pierwsza wartość definiuje grupę kontrolną w kolejności rosnącej zmiennych grupowania, a druga definiuje grupę porównawczą. Jeśli zmienna grupowania ma więcej niż dwie wartości, ten test nie jest wykonywany.
- **Porównanie median w grupach.** Opcja ta wykonuje testy prób niezależnych pod kątem tego, czy próby mają tę samą medianę. **Test mediany (k prób)** może używać mediany z łączonych prób (obliczanej ze wszystkich rekordów w zbiorze danych) lub niestandardowej wartości jako hipotetycznej mediany. Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**.
- **Porównanie przedziałów w grupach. Oszacowanie Hodgesa-Lehmana (2 próby)** tworzy oszacowanie niezależnych prób i przedział ufności dla różnicy w medianach obu grup. Jeśli zmienna grupowania ma więcej niż dwie wartości, ten test nie jest wykonywany.

Opcje testu

Poziom istotności. Określa poziom istotności (alfa) dla wszystkich testów. Podaj wartość liczbową między 0 a 1. Wartość domyślna do 0,05.

Przedział ufności (%). Określa poziom ufności dla wszystkich tworzonych przedziałów ufności. Podaj wartość liczbową z zakresu od 0 do 100. Wartością domyślną jest 95.

Wykluczone obserwacje. Określa sposób ustalenia bazy obserwacji dla testów. **Wyłączenie wszystkich obserwacji z brakami** oznacza, że rekordy z brakującymi wartościami dla jakiegokolwiek zmiennej nazwanej w jakiegokolwiek opcji komendy są wyłączone z wszelkich analiz. **Wyłączenie obserwacji test po teście** oznacza, że rekordy z brakującymi wartościami dla zmiennej, która jest używana do określonego testu, są pomijane w tym teście. Gdy poda się wiele testów w analizie, każdy test jest oceniany oddzielnie.

Braki danych zdefiniowane przez użytkownika

Braki danych użytkownika dla zmiennych jakościowych. Zmienne jakościowe muszą posiadać prawidłowe wartości dla rekordu, który ma zostać zawarty w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako prawidłowe wśród zmiennych jakościowych. Systemowe braki danych i brakujące wartości zmiennych ciągłych są zawsze traktowane jako nieprawidłowe.

Dodatkowe właściwości komendy NPTESTS

Język składni komend umożliwia również:

- Określ test dla jednej próby, test dla prób niezależnych i test dla prób zależnych w pojedynczym przebiegu procedury.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Testy nieparametryczne dla prób zależnych

Identyfikuje różnice między co najmniej dwoma zmiennymi zależnymi za pomocą co najmniej jednego testu nieparametrycznego. Testy nieparametryczne nie zakładają rozkładu normalnego danych.

Wymagania dotyczące danych. Każdy rekord odpowiada danemu obiektowi, w przypadku którego co najmniej dwa powiązane pomiary są przechowywane w oddzielnych zmiennych w zestawie danych.

Przykładowo: badanie dotyczące efektywności planów dietetycznych można przeanalizować za pomocą testów nieparametrycznych dla prób zależnych, jeśli waga każdego z obiektów jest mierzona w równych odstępach czasu i przechowywana w zmiennych takich jak *Waga przed dietą*, *Waga w trakcie diety* i *Waga po diecie*. Te zmienne są „powiązane ze sobą”.

Jaki jest cel? Cele pozwalają na szybkie określenie różnych, ale powszechnie używanych ustawień testów.

- **Automatyczne porównanie danych empirycznych z hipotetycznymi.** W przypadku tego celu test McNemara jest stosowany do danych jakościowych, gdy określone są dwie zmienne, Q Cochrańa do danych jakościowych, gdy określone są co najmniej dwie zmienne, test znaków rangowanych dopasowanej pary Wilcoxa, gdy określone są 2 zmienne i dwuczynnikowa analiza wariancji Friedmana (k prób) do danych ilościowych, gdy określone są więcej niż 2 zmienne.
- **Analiza użytkownika.** Należy wybrać tę opcję chcąc ręcznie skorygować ustawienia testu w zakładce Ustawienia. Należy zwrócić uwagę, że to ustawienie jest zaznaczone automatycznie, gdy później dokona się zmian w zakładce Ustawienia w opcjach, które są niekompatybilne z obecnie wybranym celem.

Gdy określone są zmienne o różnych poziomach pomiaru, najpierw są one dzielone wg poziomu pomiaru, a następnie do każdej grupy jest stosowany odpowiedni test. Przykładowo: użytkownik jako cel wybiera **Automatyczne porównanie danych empirycznych z hipotetycznymi** i określa 3 zmienne ciągłe oraz 2 zmienne nominalne, następnie test Friedmana jest stosowany do zmiennych ciągłych, a test McNemara do zmiennych nominalnych.

Uzyskiwanie testów nieparametrycznych dla jednej próby

Z menu wybierz:

Analiza > Testy nieparametryczne > Próby zależne...

1. Kliknij opcję **Uruchom**.

Opcjonalnie można wykonać następujące czynności:

- Określ cel w zakładce Cel.
- Określ przypisania zmiennych w zakładce Zmienne.
- Określ ustawienia zaawansowane w zakładce Zaawansowane.

Zakładka Zmienne

Zakładka Zmienne określa zmienne do przetestowania.

Użyj wstępnie zdefiniowanych ról. Ta opcja wykorzystuje istniejące informacje i zmiennych. Wszystkie zmienne z wstępnie zdefiniowaną rolą jako Cel lub Łącznie będą używane jako zmienne testowe. Wymagane są co najmniej dwie zmienne. .

Użyj niestandardowych przypisań. Opcja ta pozwala na nadpisanie ról zmiennych. Po wybraniu tej opcji, określ poniższe zmienne:

- **Testuj zmienne.** Wybierz co najmniej dwie zmienne. Każda zmienna odpowiada jednej próbie zależnej.

Zakładka Ustawienia

Zakładka Ustawienia zawiera wiele różnych grup ustawień, które można zmieniać w celu precyzyjnego określenia sposobu przetwarzania danych przez procedurę. W przypadku wprowadzenia zmian w ustawieniach domyślnych, które są niezgodne z innymi celami, zakładka Cel zostanie automatycznie zaktualizowana do zaznaczenia opcji **Analiza niestandardowa**.

Wybierz testy

Ustawienia te określają testy do wykonania w zmiennych podanych w zakładce Zmienne.

Automatycznie wybierz testy odpowiednie do danych. W przypadku tego ustawienia test McNemara jest stosowany do danych jakościowych, gdy określone są dwie zmienne, Q Cochrańa do danych

jakościowych, gdy określone są co najmniej dwie zmienne, test znaków rangowanych dopasowanej pary Wilcoxona, gdy określone są 2 zmienne i dwuczynnikowa analiza wariancji Friedmana (k prób) do danych ilościowych, gdy określone są więcej niż 2 zmienne.

Pozwól użytkownikowi wybrać testy. To ustawienie pozwala na wybór określonych testów do wykonania.

- **Test zmiany w danych binarnych. Test McNemara (2 próby)** można zastosować do zmiennych jakościowych. Powoduje to utworzenie testu dla prób sprawdzającego, czy kombinacje wartości między dwoma zmiennymi flagi (zmienne jakościowe tylko z dwoma wartościami) są równie prawdopodobne. Jeśli na zakładce Zmienne określono więcej niż dwie zmienne, test nie jest wykonywany. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat "[Test McNemara: Definiuj sukces](#)" na stronie 206. **Q Cochran (k prób)** można zastosować do zmiennych jakościowych. Powoduje to utworzenie testu dla prób sprawdzającego, czy kombinacje wartości między k zmiennymi flagi (zmienne jakościowe tylko z dwoma wartościami) są równie prawdopodobne. Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**. W celu uzyskania szczegółowych informacji na temat ustawień testu patrz temat "[Q Cochran: Definiuj sukces](#)" na stronie 207.
- **Test zmian w danych wielomianowych. Test jednorodności brzegowej (2 próby)** powoduje utworzenie testu dla prób zależnych sprawdzającego, czy kombinacje wartości między dwoma sparowanymi zmiennymi porządkowymi są równie prawdopodobne. Test jednorodności brzegowej zazwyczaj używany w przypadku powtarzanych pomiarów. Jest rozszerzeniem testu McNemara, przeznaczonego dla zmiennych dychotomicznych, na przypadek zmiennych wielokategoryjnych. Jeśli na zakładce Zmienne określono więcej niż dwie zmienne, test nie jest wykonywany.
- **Porównanie różnicy median z wartością hipotetyczną.** Każdy z tych testów powoduje utworzenie testu dla prób zależnych sprawdzającego, czy różnica median między dwoma zmiennymi jest różna od 0. Test stosuje się do zmiennych ciągłych i porządkowych. Jeśli na zakładce Zmienne określono więcej niż dwie zmienne, te testy nie są wykonywane.
- **Oszacowanie przedziału ufności.** Powoduje utworzenie oszacowania prób zależnych oraz przedziału ufności dla różnicy mediany między dwoma sparowanymi zmiennymi. Test stosuje się do zmiennych ciągłych i porządkowych. Jeśli na zakładce Zmienne określono więcej niż dwie zmienne, test nie jest wykonywany.
- **Podział związków. Kendall's coefficient of concordance (k samples)** produces a measure of agreement among judges or raters, where each record is one judge's rating of several items (fields). Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**.
- **Porównaj rozkłady. Dwuczynnikowa analiza wariancji Friedmana (k prób)** powoduje utworzenie testu dla prób zależnych sprawdzającego, czy k prób zależnych pochodzi z tej samej populacji. Opcjonalnie można zażądać wiele porównań prób k , wielokrotne porównania **wszystkie parami** lub porównania **metodą krokową zstępującą**.

Test McNemara: Definiuj sukces

Test McNemara jest przeznaczony dla zmiennych flagi (zmienne jakościowe posiadające tylko dwie kategorie), ale stosuje się go do wszystkich zmiennych jakościowych przy pomocy reguł do definiowania „sukcesów”.

Definiuj sukces dla zmiennych jakościowych. Ta opcja określa, w jaki sposób definiuje się „sukces” w przypadku zmiennych jakościowych.

- **Pierwsza kategoria znaleziona w danych** wykonuje test przy pomocy pierwszej wartości znalezionej próbie do zdefiniowania „sukcesu”. Opcja ta ma zastosowanie tylko dla zmiennych nominalnych lub porządkowych posiadających tylko dwie wartości; wszystkie pozostałe zmienne jakościowe podane w zakładce Zmienne, w której wykorzystana jest ta opcja, nie będą testowane. Jest to wartość domyślna.
- **Określ wartości sukcesu** wykonuje test przy pomocy określonej listy wartości do zdefiniowania „sukcesu”. Podaj listę wartości łańcuchowych lub liczbowych. Wartości na liście nie muszą występować w próbie.

Q Cochran: Definiuj sukces

Test Q Cochrana jest przeznaczony dla zmiennych flagi (zmiennie jakościowe zawierające tylko dwie kategorie), ale stosuje się go do wszystkich zmiennych jakościowych przy pomocy reguł do definiowania „sukcesów”.

Definiuj sukces dla zmiennych jakościowych. Ta opcja określa, w jaki sposób definiuje się „sukces” w przypadku zmiennych jakościowych.

- **Pierwsza kategoria znaleziona w danych** wykonuje test przy pomocy pierwszej wartości znalezionej próbie do zdefiniowania „sukcesu”. Opcja ta ma zastosowanie tylko dla zmiennych nominalnych lub porządkowych posiadających tylko dwie wartości; wszystkie pozostałe zmiennie jakościowe podane w zakładce Zmienne, w której wykorzystana jest ta opcja, nie będą testowane. Jest to wartość domyślna.
- **Określ wartości sukcesu** wykonuje test przy pomocy określonej listy wartości do zdefiniowania „sukcesu”. Podaj listę wartości łańcuchowych lub liczbowych. Wartości na liście nie muszą występować w próbie.

Opcje testu

Poziom istotności. Określa poziom istotności (alfa) dla wszystkich testów. Podaj wartość liczbową między 0 a 1. Wartość domyślna do 0,05.

Przedział ufności (%). Określa poziom ufności dla wszystkich tworzonych przedziałów ufności. Podaj wartość liczbową z zakresu od 0 do 100. Wartością domyślną jest 95.

Wykluczone obserwacje. Określa sposób ustalenia bazy obserwacji dla testów.

- **Wyłączanie wszystkich obserwacji z brakami** oznacza, że rekordy z brakującymi wartościami dla jakiegokolwiek zmiennej nazwanej w jakiegokolwiek opcji komendy są wyłączone z wszelkich analiz.
- **Wyłączanie obserwacji test po teście** oznacza, że rekordy z brakującymi wartościami dla zmiennej, która jest używana do określonego testu, są pomijane w tym teście. Gdy poda się wiele testów w analizie, każdy test jest oceniany oddzielnie.

Braki danych zdefiniowane przez użytkownika

Braki danych użytkownika dla zmiennych jakościowych. Zmiennie jakościowe muszą posiadać prawidłowe wartości dla rekordu, który ma zostać zawarty w analizie. Te elementy pozwalają zdecydować, czy wartości braków danych zdefiniowanych przez użytkownika są traktowane jako prawidłowe wśród zmiennych jakościowych. Systemowe braki danych i brakujące wartości zmiennych ciągłych są zawsze traktowane jako nieprawidłowe.

Dodatkowe właściwości komendy NPTESTS

Język składni komend umożliwia również:

- Określ test dla jednej próby, test dla prób niezależnych i test dla prób zależnych w pojedynczym przebiegu procedury.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Dodatkowe właściwości komendy NPTESTS

Język składni komend umożliwia również:

- Określ test dla jednej próby, test dla prób niezależnych i test dla prób zależnych w pojedynczym przebiegu procedury.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Wykresy tradycyjne

Istnieje wiele "tradycyjnych" wykresów, które również wykonują testy nieparametryczne. Te wykresy obsługują funkcje zapewniane przez opcję Testy dokładne.

Test chi-kwadrat. Umożliwia podział zmiennej na kategorie i wyliczenie statystyki chi-kwadrat w oparciu o różnice między częstościami obserwowanymi i oczekiwanymi.

Test dwumianowy. Umożliwia porównanie częstości obserwowanych we wszystkich kategoriach zmiennej dychotomicznej z oczekiwanymi częstościami z rozkładu dwumianowego.

Test serii. Umożliwia sprawdzenie, czy dwie wartości zmiennej występują w przypadkowym porządku.

Test dla jednej próby Kołmogorowa-Smirnowa. Umożliwia porównanie obserwowanej funkcji skumulowanego rozkładu dla zmiennej z określonym teoretycznym rozkładem normalnym, jednostajnym, wykładniczym lub Poissona.

Testy dla dwóch prób niezależnych. Umożliwiają porównanie dwóch grup obserwacji jednej zmiennej. Dostępne są: test *U* Manna-Whitneya, test dla dwóch prób Kołmogorowa-Smirnowa, test skrajnych reakcji Moseesa i test serii Walda-Wolfowitza.

Testy dla dwóch prób zależnych. Umożliwia porównanie rozkładów dwóch zmiennych. Dostępne są: test znaków rangowanych Wilcozona, test znaków i test McNemara.

Testy dla kilku prób niezależnych. Umożliwiają porównanie co najmniej dwóch grup obserwacji jednej zmiennej. Dostępne są: test Kruskala-Wallisa, test medianowy i test Jonckheere-Terpstra.

Testy dla kilku prób zależnych. Umożliwia porównanie rozkładu co najmniej dwóch zmiennych. Dostępne są: test Friedmana, *W* Kendalla i *Q* Cochra.

Dla wszystkich powyższych testów dostępne są: kwartyle i średnia, odchylenie standardowe, minimum, maksimum i liczba obserwacji bez braków.

Test chi-kwadrat

Procedura testu chi-kwadrat umożliwia podzielenie zmiennej na kategorie i wyliczenie statystyki chi-kwadrat. Ten test dobrego dopasowania umożliwia porównanie obserwowanych i oczekiwanych częstości w każdej kategorii, aby sprawdzić czy wszystkie kategorie zawierają wartości w tych samych proporcjach, albo czy każda z kategorii zawiera wartości w proporcjach określonych przez użytkownika.

Przykłady. Test chi-kwadrat umożliwia stwierdzenie, czy paczka żelków zawiera równą ilość niebieskich, brązowych, zielonych, pomarańczowych, czerwonych i żółtych cukierków. Można też sprawdzić, czy paczka żelków zawiera 5% niebieskich, 30% brązowych, 10% zielonych, 20% pomarańczowych, 15% czerwonych i 15% żółtych cukierków.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum i kwartyle. Wartość liczbowa i procentowa obserwacji z brakami i bez braków, liczba obserwacji obserwowanych i oczekiwanych dla każdej kategorii, reszty i statystyka chi-kwadrat.

Wymagania dotyczące danych dla testu chi-kwadrat

Dane. Należy używać uporządkowanych lub nieuporządkowanych kategoryalnych zmiennych numerycznych (porządkowe lub nominalne poziomy pomiaru). Aby skonwertować zmienne łańcuchowe do postaci zmiennych numerycznych, należy skorzystać z procedury automatycznego rekodowania w menu Przekształcenia.

Założenia. Testy nieparametryczne nie wymagają założeń dotyczących kształtu rozkładu. Zakłada się, że dane pochodzą z losowej próby. Oczekiwane częstości dla każdej kategorii powinny wynosić co najmniej 1. Nie więcej niż 20% kategorii powinno mieć oczekiwane częstości na poziomie niższym niż 5.

Wykonywanie testu chi-kwadrat

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > Chi-kwadrat...

2. Wybierz co najmniej jedną zmienną testowaną. Każda zmienna daje oddzielny test.

3. Opcjonalnie można kliknąć przycisk **Opcje**, aby uzyskać dostęp do statystyk opisowych, kwartyli i określić sposób traktowania brakujących danych.

Oczekiwany zakres i wartości oczekiwane dla testu chi-kwadrat

Oczekiwany zakres. Każda odrębna wartość zmiennej jest domyślnie definiowana jako kategoria. Aby ustanowić kategorie w określonym przedziale, należy wybrać opcję **Użyj określonego przedziału** i wprowadzić wartości całkowite dla dolnej i górnej granicy. Kategorie są ustanawiane dla każdej wartości całkowitej zawierającej się w przedziale zamkniętym, a obserwacje z wartościami znajdującymi się poza granicami są wyłączone. Na przykład jeśli określono wartość dolnej granicy jako 1, a górnej jako 4, to dla testu chi-kwadrat zostaną wykorzystane jedynie wartości całkowite od 1 do 4.

Wartości oczekiwane. Domyślnie, wartości oczekiwane wszystkich kategorii są równe. Oczekiwane proporcje kategorii mogą być określane przez użytkownika. Należy wybrać opcję **Wartości**, wprowadzić wartość większą od 0 dla każdej kategorii testowanej zmiennej i kliknąć przycisk **Dodaj**. Każda nowa wartość dodana do listy pojawia się na jej końcu. Ważny jest porządek wartości, który odpowiada rosnącemu porządkowi wartości kategorii testowanej zmiennej. Pierwsza wartość na liście odpowiada najniższej wartości grupy zmiennej testowanej, zaś ostatnia – wartości najwyższej. Elementy na liście wartości są sumowane, a następnie każda wartość jest dzielona przez tę sumę w celu wyliczenia proporcji obserwacji oczekiwanych w odpowiedniej kategorii. Na przykład lista wartości, na której znajdują się liczby 3, 4, 5, 4 określa oczekiwane proporcje jako 3/16, 4/16, 5/16 i 4/16.

Test chi-kwadrat: Opcje

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby obserwacji bez braków.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (test chi-kwadrat)

Język składni komend umożliwia również:

- Określenie różnych wartości minimum i maksimum lub oczekiwanych częstości dla różnych zmiennych (za pomocą opcji komendy CHISQUARE).
- Przetestowanie tej samej zmiennej względem różnych oczekiwanych częstości lub wykorzystanie różnych przedziałów (za pomocą opcji komendy EXPECTED).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

test dwumianowy

Procedura testu dwumianowego służy do porównywania częstości obserwowanych dwóch kategorii zmiennej dychotomicznej z częstościami oczekiwanymi z rozkładu dwumianowego o określonym parametrze prawdopodobieństwa. Domyślny parametr prawdopodobieństwa dla obu grup wynosi 0,5. Aby zmienić wartości prawdopodobieństwa, można wprowadzić testowaną proporcję dla pierwszej grupy. Wartość prawdopodobieństwa dla drugiej grupy będzie równa 1 minus wartość prawdopodobieństwa dla pierwszej grupy.

Przykład. Kiedy rzucasz monetą, prawdopodobieństwo orła jest równe 1/2. W oparciu o tę hipotezę rzucamy monetą 40 razy, a wyniki są rejestrowane (orły lub reszki). Gdyby w 3/4 rzutów wypadł orzeł, to na podstawie testu dwumianowego można by stwierdzić, że obserwowany poziom istotności jest niski

(0,0027). Wyniki te świadczą o tym, że nie jest wiarygodne, aby prawdopodobieństwo wyrzucenia orła wynosiło $1/2$ – moneta jest prawdopodobnie zniekształcona.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwantyle.

Wymagania dotyczące danych przy teście dwumianowym

Dane. Testowane zmienne powinny mieć charakter numeryczny i być dychotomiczne. Aby skonwertować zmienne łańcuchowe do postaci zmiennych numerycznych, należy skorzystać z procedury automatycznego rekodowania w menu Przekształcenia. **Zmienna dychotomiczna** to zmienna, która może przyjmować jedynie dwie wartości: *tak* lub *nie*, *prawda* lub *fałsz*, 0 lub 1 itd. Pierwsza napotkana wartość w zbiorze danych definiuje pierwszą grupę, a następną wartość definiuje drugą grupę. Jeśli zmienne nie są dychotomiczne, należy określić punkt podziału. Punkt podziału powoduje przypisanie obserwacji o wartościach mniejszych lub równych od niego do pierwszej grupy, a pozostałych obserwacji do drugiej grupy.

Założenia. Testy nieparametryczne nie wymagają założeń dotyczących kształtu rozkładu. Zakłada się, że dane pochodzą z losowej próby.

Wykonywanie testu dwumianowego

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > Dwumianowy...

2. Wybierz co najmniej jedną testowaną zmienną numeryczną.

3. Opcjonalnie można kliknąć przycisk **Opcje**, aby uzyskać dostęp do statystyk opisowych, kwantyli i określić sposób traktowania brakujących danych.

Test dwumianowy: Opcje

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby obserwacji bez braków.
- **Kwantyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych którejkolwiek testowanej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (test dwumianowy)

Język składni komend umożliwia również:

- Wybór określonych grup (i wyłączenie innych), kiedy zmienna ma więcej niż dwie kategorie (za pomocą opcji komendy BINOMIAL).
- Określenie różnych punktów podziału lub wartości prawdopodobieństwa dla różnych zmiennych (za pomocą opcji komendy BINOMIAL).
- Przetestowanie tej samej zmiennej względem różnych punktów podziału (za pomocą opcji komendy EXPECTED).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Test serii

Procedura Test serii umożliwia przeprowadzenie testu sprawdzającego, czy kolejność wystąpień dwóch wartości zmiennej ma charakter losowy. Seria jest ciągiem kolejno następujących po sobie identycznych

lub podobnych wartości zmiennej. Zbyt duża lub zbyt mała liczba serii w próbie pozwala sądzić, że próba nie ma charakteru losowego.

Przykłady. Załóżmy, że wśród 20 osób przeprowadzono ankietę z zapytaniem, czy nabyłyby jakiś produkt. Zakładana losowość próby byłaby poważnie kwestionowana, gdyby wszystkie 20 osób było tej samej płci. Do ustalenia, czy dana próba dobrana była w sposób losowy, może być wykorzystany test serii.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwartyle.

Wymagania dotyczące danych przy teście serii

Dane. Zmienne muszą mieć postać numeryczną. Aby skonwertować zmienne łańcuchowe do postaci zmiennych numerycznych, należy skorzystać z procedury automatycznego rekodowania w menu Przekształcenia.

Założenia. Testy nieparametryczne nie wymagają założeń dotyczących kształtu rozkładu. Należy używać prób o ciągłych rozkładach prawdopodobieństwa.

Wykonywanie testu serii

1. Wybierz z menu następującą opcję:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > Serii...

2. Wybierz co najmniej jedną testowaną zmienną numeryczną.
3. Opcjonalnie można kliknąć przycisk **Opcje**, aby uzyskać dostęp do statystyk opisowych, kwartyli i określić sposób traktowania brakujących danych.

Punkt podziału w teście serii

Punkt podziału. Aby podzielić wybrane zmienne na dwie grupy, należy określić punkt podziału. Jako punkt podziału wykorzystać można rzeczywistą wartość średniej, dominantę lub określoną wartość. Obserwacje o wartościach mniejszych od wartości punktu podziału są przydzielane do jednej grupy, a obserwacje o wartościach większych lub równych – do drugiej grupy. Dla każdego wybranego punktu podziału przeprowadzany jest jeden test.

Opcje testu serii

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby obserwacji bez braków.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (test serii)

Język składni komend umożliwia również:

- Określenie różnych punktów podziału dla różnych zmiennych (za pomocą opcji komendy RUNS).
- Przetestowanie tej samej zmiennej względem różnych punktów podziału użytkownika (za pomocą opcji komendy RUNS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Test Kołmogorowa-Smirnowa dla jednej próby

Procedura testu Kołmogorowa-Smirnowa dla jednej próby umożliwia porównanie obserwowanej funkcji skumulowanego rozkładu dla zmiennej z określonym teoretycznym rozkładem normalnym, jednostajnym, Poissona lub wykładniczym. Wartość Z testu Kołmogorowa-Smirnowa jest wyliczana z największej różnicy (w wartościach bezwzględnych) pomiędzy obserwowanymi a teoretycznymi funkcjami skumulowanego rozkładu. Test ten weryfikuje dobroć dopasowania, a więc sprawdza, czy dane obserwacje mogą pochodzić z określonego rozkładu.

Począwszy od wersji 27.0 test Lillieforsa można wykorzystać do oszacowania wartości p przy zastosowaniu losowania Monte Carlo — przeprowadzając test względem rozkładu normalnego z oszacowanymi parametrami (ta funkcja była wcześniej możliwa tylko w ramach procedury Eksploracja).

Przykład

W wielu testach parametrycznych wymagane są zmienne o rozkładzie normalnym. Test Kołmogorowa-Smirnowa dla jednej próby pozwala sprawdzić, czy dana zmienna (na przykład *dochód*) ma rozkład normalny.

Statystyki

Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji, kwartyle, test Lillieforsa i symulacja Monte Carlo.

Wymagania dotyczące danych przy teście Kołmogorowa-Smirnowa dla jednej próby

Dane

Należy używać zmiennych ilościowych (miary poziomów przedziałów lub proporcji).

Założenia

W teście Kołmogorowa-Smirnowa przyjmuje się, że parametry rozkładu testowego są z góry określone. W tej procedurze wykonywane jest oszacowanie parametrów z próby. Dla rozkładu normalnego parametrami są średnia z próby i odchylenie standardowe próby, w rozkładzie jednostajnym zakres wyznaczają wartość minimalna i maksymalna próby, parametrem rozkładu Poissona i rozkładu wykładniczego jest średnia z próby. Siła tego testu wykrywania odstępstw od hipotetycznego rozkładu może być znacznie ograniczona.

Gdy niektóre parametry rozkładu muszą być oszacowane z próby, test Kołmogorowa-Smirnowa nie ma już zastosowania. W tych przypadkach można użyć testu Lillieforsa do oszacowania wartości p z wykorzystaniem losowania Monte Carlo — by przetestować normalność rozkładu z nieznaną średnią i wariancją. Test Lillieforsa ma zastosowanie do trzech rozkładów ciągłych (**Normalny**, **Wykładniczy** i **Jednostajny**). Należy zauważyć, że test nie ma zastosowania, jeśli rozkład bazowy jest dyskretny (**Poissona**). Jeśli nie zostaną określone odpowiednie parametry rozkładu, test jest zdefiniowany tylko dla wnioskowania dla jednej próby.

Wykonywanie testu Kołmogorowa-Smirnowa dla jednej próby

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > K-S dla jednej próby...

2. Wybierz co najmniej jedną testowaną zmienną numeryczną. Każda zmienna daje oddzielny test.

3. Opcjonalnie wybierz metodę rozkładu dla testu:

Normalne

Gdy ta opcja jest wybrana, można określić, czy parametry rozkładu są szacowane na podstawie danych z próby (ustawienie domyślne), czy ustawień użytkownika. Gdy wybrana jest opcja **Użyj danych z próby**, używane są zarówno istniejące wyniki asymptotyczne, jak i korekta istotności Lillieforsa na podstawie losowania Monte Carlo. Gdy wybrana jest opcja **Użytkownika**, podaj wartości **Średnia** i **Odchylenie standardowe**.

Jednostajne

Gdy ta opcja jest wybrana, można określić, czy parametry rozkładu są szacowane na podstawie danych z próby (ustawienie domyślne), czy ustawień użytkownika. Gdy wybrana jest opcja **Użyj**

danych z próby, używany jest test Lillieforsa. Gdy wybrana jest opcja **Użytkownika**, podaj wartości **Min.** i **Maks.**

Poissona

Gdy ta opcja jest wybrana, określ wartość parametru **Średnia**.

Wykładnicze

Gdy ta opcja jest wybrana, można określić, czy parametry rozkładu są szacowane na podstawie średniej z próby (ustawienie domyślne), czy ustawień użytkownika. Gdy wybrana jest opcja **Użyj danych z próby**, używany jest test Lillieforsa. Jeśli wybrana jest opcja **Użytkownika**, należy podać wartość parametru **Średnia**.

4. Opcjonalnie kliknij opcję **Symulacja**, aby określić parametry symulacji Monte Carlo, kliknij przycisk **Dokładne**, aby określić dokładne parametry testu, lub kliknij przycisk **Opcje**, aby wybrać opcje statystyk opisowych, kwartyli i postępowania z brakami danych.

Test Kołmogorowa-Smirnowa dla jednej próby: Symulacja

Gdy niektóre parametry rozkładu muszą być oszacowane z próby, test Kołmogorowa-Smirnowa nie ma już zastosowania. W tych przypadkach można użyć testu Lillieforsa do oszacowania wartości p z wykorzystaniem losowania Monte Carlo — by przetestować normalność rozkładu z nieznaną średnią i wariancją. Test Lillieforsa ma zastosowanie do trzech rozkładów ciągłych (**Normalny**, **Wykładniczy** i **Jednostajny**). Należy zauważyć, że test nie ma zastosowania, jeśli rozkład bazowy jest dyskretny (**Poissona**). Jeśli nie zostaną określone odpowiednie parametry rozkładu, test jest zdefiniowany tylko dla wnioskowania dla jednej próby.

Parametry symulacji Monte Carlo

Poziom ufności

To ustawienie opcjonalne resetuje poziom przedziału ufności, który jest szacowany przez test Kołmogorowa-Smirnowa w przypadku korzystania z symulacji Monte Carlo. Wartość musi należeć do przedziału od 0 do 100. Ustawieniem domyślnym jest 99.

Liczba próbek

To ustawienie opcjonalne resetuje liczbę replik używanych w teście Lillieforsa na potrzeby losowania Monte Carlo. Wartość musi być jedną liczbą całkowitą z zakresu od 10000 do największej liczby prób. Wartością domyślną jest 10000.

Pomijaj wyniki Monte Carlo dla rozkładu normalnego

To ustawienie opcjonalne powoduje pominięcie losowania Monte Carlo dla wyników o rozkładzie normalnym. Domyślnie to ustawienie nie jest wybrane (co oznacza, że prezentowane są zarówno istniejące wyniki asymptotyczne, jak i wyniki testu Lillieforsa, które są oparte na losowaniu Monte Carlo).

Test Kołmogorowa-Smirnowa dla jednej próby: Opcje

Statystyki

Można wybrać jedną lub obie ze statystyk podsumowujących.

Statystyka opisowa

Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby obserwacji bez braków.

Kwartyle

Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych

Kontrola sposobu postępowania z brakami danych.

Wyklucz obserwacje test po teście

W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.

Wyklucz wszystkie obserwacje z brakami

Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (test Kołmogorowa-Smirnowa dla jednej próby)

Składnia komend umożliwia również określenie parametrów rozkładu testowego (za pomocą opcji komendy K-S).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Test dla dwóch prób niezależnych

Procedura testów dla dwóch prób niezależnych umożliwia porównywanie dwóch grup obserwacji jednej zmiennej.

Przykład. Wynaleziono nowy aparat ortodontyczny, który ma być wygodniejszy, bardziej estetyczny i ma umożliwiać szybsze postępy w prostowaniu zębów. Aby sprawdzić, czy aparaty nowego typu muszą być noszone tak długo, jak aparaty starego typu, wybrano losowo 10 dzieci, które miały nosić aparaty starego typu, i kolejnych 10, które miały nosić aparaty nowego typu. Za pomocą testu *U* Manna-Whitneya można stwierdzić, że dzieci posiadające aparaty nowego typu przeciętnie nie musiały ich nosić tak długo, jak te, które miały aparaty starego typu.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwartyle. Testy: *U* Manna-Whitney'a, test skrajnych reakcji Mosesa, test *Z* Kołmogorowa-Smirnowa, test serii Walda-Wolfowitza.

Wymagania dotyczące danych przy testach dla dwóch prób niezależnych

Dane. Należy używać zmiennych numerycznych, które można uporządkować.

Założenia. Należy korzystać z niezależnych, przypadkowych prób. Test Mann-Whitney *U* testuje równość dwóch rozkładów. Aby go użyć do testowania różnic lokacji pomiędzy dwoma rozkładami należy założyć, że rozkłady mają ten sam kształt.

Wykonywanie testów dla dwóch prób niezależnych

1. Wybierz z menu następującą opcję:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > Dwie próby niezależne...

2. Wybierz co najmniej jedną zmienną numeryczną.
3. Wybierz zmienną grupującą i kliknij przycisk **Definiuj grupy**, aby podzielić dane na dwie grupy lub próby.

Typy testów dla dwóch prób niezależnych

Typ testu. Dostępne są cztery testy, sprawdzające, czy dwie niezależne próby (grupy) pochodzą z tej samej populacji.

Test *U* Manna-Whitneya jest najpopularniejszym z testów dla dwóch prób niezależnych. Jest on odpowiednikiem testu sumy rang Wilcozona i testu dla dwóch grup Kruskala-Wallis. Test Manna-Whitneya służy do sprawdzania, czy dwie próbkowane populacje są tożsame pod względem położenia. Obserwacje z obu grup są łączone i nadawane są im rangi. W przypadku wiązań przypisuje się rangę średnią. Liczba wiązań powinna być niewielka w porównaniu z ogólną liczbą obserwacji. Jeżeli populacje są identyczne pod względem położenia, to rangi powinny być losowo wymieszane między dwoma próbami. Test wylicza liczbę razy, kiedy ocena grupy 1 poprzedza ocenę grupy 2 i liczbę razy, kiedy ocena grupy 2 poprzedza ocenę grupy 1. Statystyka *U* Mann-Whitneya jest mniejsza z tych dwóch liczb. Wyświetlona jest także statystka sumy rangi Wilcoxon *W*. *W* jest sumą rangi dla grupy o mniejszej średniej rangi, chyba że grupy mają taką samą średnią rangę, w którym to przypadku jest to suma rangi z grupy, która została nazwana jako ostatnia w oknie dialogowym Definiuj grupy dla dwóch prób niezależnych.

Test *Z* Kołmogorowa-Smirnowa i test serii Walda-Wolfowitza są testami o charakterze ogólniejszym, które wykrywają różnice zarówno w położeniach, jak i w kształtach rozkładów. Test Kołmogorowa-Smirnowa jest oparty na maksymalnej różnicy bezwzględnej między zaobserwowanymi funkcjami skumulowanego rozkładu dla obu prób. Kiedy różnica jest znacząca, dwa rozkłady uważa się za różne.

Test serii Walda-Wolfowitza łączy obserwacje z obu grup i nadaje im rangi. Jeśli dwie próbki pochodzą z tej samej populacji, te dwie grupy powinny być losowo rozrzucone w całym rankingu.

Test skrajnych reakcji Mosea opiera się na założeniu, że zmienna eksperymentalna wpłynie na niektóre podmioty w jednym kierunku, a na inne w kierunku przeciwnym. Test sprawdza odpowiedzi skrajne, porównane z grupą kontrolną. Ten test skupia się na rozpiętości grupy kontrolnej i stanowi metodę pomiaru wpływu wartości skrajnych w grupie eksperymentalnej na rozpiętość w połączeniu z grupą kontrolną. Grupę kontrolną definiuje wartość grupy 1 w oknie dialogowym Testy dla dwóch prób niezależnych: Definiuj grupy. Obserwacje z obu grup są łączone i nadawane im są rangi. Zakres grupy kontrolnej jest obliczany jako różnica między rangami największych i najmniejszych wartości w grupie kontrolnej plus 1. Ponieważ przypadkowe wartości odstające mogą łatwo zniekształcić zakres rozpiętości, 5% obserwacji kontrolnych jest automatycznie obcinanych z każdego końca.

Testy dla dwóch prób niezależnych: Definiuj grupy

Aby podzielić plik na dwie grupy lub próbki, należy wprowadzić wartość całkowitą dla grupy 1 i inną wartość dla grupy 2. Obserwacje z innymi wartościami są wykluczane z analizy.

Testy dla dwóch prób niezależnych: Opcje

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby niebrakujących obserwacji.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (testy dla dwóch prób niezależnych)

Język składni komend umożliwia również określenie liczby obserwacji, które mają zostać przycięte w teście Mosea (za pomocą opcji komendy MOSES).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Testy dla dwóch prób zależnych

Procedura Testy dla dwóch prób zależnych służy do porównywania rozkładu dwóch zmiennych.

Przykład. Czy zazwyczaj rodziny sprzedające swoje domy otrzymują za nie żadaną cenę? Stosując test znaków rangowanych Wilcozona dla danych dla 10 domów, można dowiedzieć się, że siedem rodzin otrzymuje cenę niższą od żądanej, jedna rodzina otrzymuje cenę wyższą od żądanej, a dwie rodziny otrzymują żadaną cenę.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwartyle. Testy: znaków rangowanych Wilcozona, znaków, McNemara. Jeśli zainstalowana jest opcja Exact Test (dostępna tylko w systemach operacyjnych Windows), dostępny jest także test jednorodności brzegowej.

Wymagania dotyczące danych przy testach dla dwóch prób zależnych

Dane. Należy używać zmiennych numerycznych, które można uporządkować.

Założenia. Dla dwóch zmiennych nie zakłada się konkretnych rozkładów, zakłada się jednak, że rozkład populacji różnic w próbach zależnych jest symetryczny.

Wykonywanie testów dla dwóch prób zależnych

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > Dwie próby zależne...

2. Wybierz co najmniej jedną parę zmiennych.

Typy testów dla dwóch prób zależnych

Testy w tej sekcji umożliwiają porównanie rozkładu dwóch zmiennych zależnych. Wybór odpowiedniego testu zależy od typu danych.

Jeżeli dane mają charakter ciągły, należy użyć testu znaków lub testu znaków rangowanych Wilcoxon. **Test znaków** oblicza różnicę między dwoma zmiennymi dla wszystkich obserwacji i sklasyfikuje różnice jako dodatnie, ujemne lub związane. Jeżeli rozkład dwóch zmiennych jest podobny, to liczba różnic dodatnich i ujemnych nie będzie się znacząco różnić. **Test znaków rangowanych Wilcoxon** umożliwia wzięcie pod uwagę zarówno informacji dotyczących znaku różnic, jak i wielkości różnic między parami. Ponieważ test znaków rangowanych Wilcoxon zawiera więcej informacji o danych, jest on bardziej wszechstronny od testu znaków.

Jeżeli dane mają charakter binarny, należy użyć **testu McNemara**. Test ten jest zazwyczaj używany w przypadku powtarzanych pomiarów, kiedy reakcje każdego podmiotu uzyskiwane są dwukrotnie – raz przed i raz po wystąpieniu określonego zdarzenia. Test McNemara umożliwia stwierdzenie, czy początkowy wskaźnik odpowiedzi (przed zdarzeniem) jest równy końcowemu wskaźnikowi odpowiedzi (po zdarzeniu). Test ten jest użyteczny w przypadku wykrywania zmian w odpowiedziach spowodowanych eksperymentalną interwencją w projektach typu „przed i po”.

Jeżeli dane mają charakter jakościowy, należy użyć **testu jednorodności brzegowej**. Jest rozszerzeniem testu McNemara, przeznaczonego dla zmiennych dychotomicznych, na przypadek zmiennych wielokategorialnych. Służy on do sprawdzania zmian w odpowiedziach (z wykorzystaniem rozkładu chi-kwadrat) i jest użyteczny przy wykrywaniu zmian w odpowiedziach spowodowanych eksperymentalną interwencją w projektach typu „przed i po”. Test jednorodności brzegowej jest dostępny jedynie po zainstalowaniu modułu testów dokładnych.

Opcje testów dla dwóch prób zależnych

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby niebrakujących obserwacji.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (Dwie próby zależne)

Język składni komend umożliwia również testowanie zmiennej z każdą zmienną na liście.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

testy dla kilku prób niezależnych

Procedura testów dla kilku prób niezależnych porównuje dwie lub więcej grup obserwacji pod względem jednej zmiennej.

Przykład. Czy istnieje różnica pomiędzy średnim czasem świecenia 100-watowych żarówek trzech różnych marek? Z jednoczynnikowej analizy wariancji Kruskala-Wallisa można wywnioskować, że trzy marki żarówek różnią się średnim czasem działania.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwartyle. Testy: H Kruskala-Wallisa, mediany.

Wymagania dotyczące danych dla testów dla kilku prób niezależnych

Dane. Należy używać zmiennych numerycznych, które można uporządkować.

Założenia. Należy korzystać z niezależnych, przypadkowych prób. Test H Kruskala-Wallisa wymaga założenia o podobieństwie kształtu testowanych prób.

Wykonywanie testów dla kilku prób niezależnych

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > K prób niezależnych...

2. Wybierz co najmniej jedną zmienną numeryczną.

3. Zaznacz zmienną grupującą, następnie kliknij przycisk **Definiuj zakres**, aby określić minimum i maksimum wartości całkowitych dla zmiennej grupującej.

Testy dla kilku prób zależnych: Typ testu

Dostępne są trzy testy w celu określenia, czy kilka niezależnych prób pochodzi z tej samej populacji. Test H Kruskala-Wallisa, test mediany oraz test Jonckheere-Terpstra — każdy z nich umożliwia sprawdzenie, czy kilka niezależnych prób pochodzi z tej samej populacji.

Test H Kruskala-Wallisa, rozszerzenie testu U Manna-Whitneya, jest nieparametrycznym odpowiednikiem jednoczynnikowej analizy wariancji i wykrywa różnice w położeniu rozkładu. **Test mediany**, który jest testem bardziej ogólnym, ale nie tak wszechstronnym, wykrywa różnice rozkładowe położenia i kształtu. Test H Kruskala-Wallisa oraz test mediany zakładają brak uporządkowania *a priori* k populacji, z których pobrano próby.

W przypadku gdy występuje naturalne uporządkowanie *a priori* (malejące lub rosnące) k populacji test **Jonckheere-Terpstra** jest bardziej odpowiedni. Na przykład niech k populacji reprezentuje k rosnących temperatur. Testowana jest hipoteza, że różne temperatury powodują ten sam rozkład reakcji. Hipoteza alternatywna brzmi, że w miarę wzrostu temperatury wzrasta wielkość reakcji. Test Jonckheere-Terpstra jest najbardziej odpowiedni, ponieważ w tym przypadku hipoteza alternatywna jest uporządkowana. Test Jonckheere-Terpstry jest dostępny tylko wtedy, jeśli zainstalowano moduł dodatku Exact Tests.

Testy dla kilku prób niezależnych: Definiuj zakres

Aby zdefiniować zakres, należy wprowadzić wartości całkowite **minimalne** oraz **maksymalne** odpowiadające najniższemu i najwyższemu kategoriom zmiennej grupującej. Obserwacje z wartościami znajdującymi się poza tymi granicami nie są uwzględniane. Na przykład jeśli wartość minimalna zostanie określona jako 1, a wartość maksymalna jako 3, to użyte zostaną jedynie wartości całkowite od 1 do 3. Wartość minimalna musi być mniejsza od wartości maksymalnej oraz należy określić obie wartości.

Testy dla kilku prób niezależnych: Opcje

Statystyki. Można wybrać jedną lub obie ze statystyk podsumowujących.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby niebrakujących obserwacji.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu percentylowi.

Braki danych. Kontrola sposobu postępowania z brakami danych.

- **Wyłączanie obserwacji test po teście.** W przypadku określenia kilku testów, każdy z nich jest oceniany oddzielnie pod względem braków danych.
- **Wyłączanie wszystkich obserwacji z brakami.** Obserwacje z brakami danych dla dowolnej zmiennej są wyłączone ze wszystkich analiz.

Dodatkowe właściwości komendy NPAR TESTS (K prób niezależnych)

Język składni komend także określić dla testu mediany wartość inną niż zaobserwowana mediana (za pomocą komendy MEDIAN).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

testy dla kilku prób zależnych

Testy dla kilku prób zależnych porównują rozkłady dwóch lub więcej zmiennych.

Przykład. Czy w oczach opinii publicznej zawody, takie jak lekarz, prawnik, policjant i nauczyciel, cieszą się różnym stopniem prestiżu? Poproszono dziesięć osób o uporządkowanie tych czterech zawodów od najmniej do najbardziej prestiżowego. Test Friedmana wykazał, że w opinii badanych wymienionym zawodom przypisano różny stopień prestiżu.

Statystyki. Średnia, odchylenie standardowe, minimum, maksimum, liczba niebrakujących obserwacji i kwartyle. Testy: Friedmana, *W* Kendalla oraz *Q* Cochra.

Wymagania dotyczące danych dla testów kilku prób zależnych

Dane. Należy używać zmiennych numerycznych, które można uporządkować.

Założenia. Testy nieparametryczne nie wymagają założeń dotyczących kształtu rozkładu. Należy używać zależnych prób losowych.

Wykonywanie testów dla kilku prób zależnych

1. Z menu wybierz:

Analiza > Testy nieparametryczne > Wykresy tradycyjne > K prób zależnych...

2. Wybierz co najmniej dwie zmienne numeryczne.

Testy dla kilku prób zależnych: Typ testu

Dostępne są trzy typy testu porównującego rozkłady kilku zmiennych zależnych.

Test Friedmana jest nieparametrycznym odpowiednikiem modelu pomiarów powtarzanych dla jednej próby lub dwuczynnikowej analizy wariancji z jedną obserwacją w komórce. Test Friedmana testuje hipotezę zerową zakładającą, że *k* powiązanych zmiennych pochodzi z tej samej populacji. Dla każdej obserwacji zmienne *k* są klasyfikowane w kolejności od 1 do *k*. Statystyka testu jest oparta na tych rangach.

Test *W* Kendalla stanowi normalizację statystyki Friedmana. Test *W* Kendalla interpretuje się jako współczynnik pomiaru zgodności pomiędzy oceniającymi. Każda obserwacja jest sędzią lub oceniającym, a każda zmienna jest pozycją lub osobą ocenianą. Dla każdej zmiennej obliczana jest suma rang. *W* Kendalla przyjmuje wartości od 0 (całkowity brak zgodności) do 1 (całkowita zgodność).

Test *Q* Cochra działa na podobnej zasadzie co test Friedmana, z tą różnicą, że ma zastosowanie, gdy wszystkie odpowiedzi są binarne. Test ten stanowi rozszerzenie testu McNemara dla *k* prób zależnych. Test *Q* Cochra weryfikuje hipotezę, że kilka zależnych zmiennych dychotomicznych ma tę samą średnią. Pomiar zmiennych dotyczy tego samego obiektu lub obiektów powiązanych.

Testy dla kilku prób zależnych: Statystyki

Można wybrać statystyki.

- **Opisowe.** Umożliwia wyświetlenie średniej, odchylenia standardowego, minimum, maksimum i liczby niebrakujących obserwacji.
- **Kwartyle.** Umożliwia wyświetlenie wartości odpowiadających dwudziestemu piątemu, pięćdziesiątemu i siedemdziesiątemu piątemu procentylowi.

Dodatkowe właściwości komendy NPAR TESTS (K prób zależnych)

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Analiza wielokrotnych odpowiedzi

Dostępne są dwie procedury służące do analizy zestawów wielokrotnych dychotomii oraz zestawów wielokrotnych kategorii. Użycie procedury Częstości wielokrotnych odpowiedzi powoduje wyświetlenie tabel częstości. Użycie procedury Tabele krzyżowe wielokrotnych odpowiedzi powoduje wyświetlenie dwu- i trójwymiarowych tabel krzyżowych. Przed użyciem każdej z tych procedur należy zdefiniować zestawy wielokrotnych odpowiedzi.

Przykład. Ten przykład ilustruje sposób wykorzystania elementów wielokrotnych odpowiedzi w przeprowadzaniu badań marketingowych. Dane są fikcyjne i nie powinny być traktowane jako prawdziwe. Linia lotnicza może przeprowadzać badania wśród pasażerów latających określoną trasą, aby dokonać oceny konkurujących z nią przewoźników. W tym przykładzie linia American Airlines chce poznać wykorzystanie przez jej pasażerów innych linii lotniczych na trasie Chicago-Nowy Jork oraz wagę jaką przywiązują oni do rozkładu lotów i jakości obsługi przy wyborze linii lotniczej. Podczas wchodzenia na pokład personel samolotu wręcza każdemu z pasażerów krótki kwestionariusz. Pierwsze pytanie: proszę zakreślić wszystkie linie lotnicze, którymi leciał (leciała) Pan (Pani) na tej trasie co najmniej raz w przeciągu ostatnich sześciu miesięcy: American, United, TWA, USAir, Inne. Jest to pytanie o wielokrotnej odpowiedzi, ponieważ pasażer może zakreślić kilka odpowiedzi. Pytanie to nie może być jednak kodowane bezpośrednio, ponieważ zmienna dla każdej obserwacji może mieć tylko jedną wartość. Aby odwzorować odpowiedzi do każdego pytania, należy użyć kilku zmiennych. Istnieją dwa sposoby, aby tego dokonać. Pierwszy polega na zdefiniowaniu zmiennej odpowiadającej każdemu z wyborów (na przykład American, United, TWA, USAir i Inne). Jeśli okregi pasażerskie United, zmienna *zjednoczona* ma przypisany kod równy 1, w przeciwnym razie 0. Jest to **metoda wielokrotnych dychotomii** zmiennych odwzorowania. Drugą metodą odwzorowania odpowiedzi jest **metoda wielokrotnych kategorii**, w której szacuje się maksymalną liczbę możliwych odpowiedzi na dane pytanie i ustala się tę samą liczbę zmiennych, z kodami używanymi do określania linii lotniczej, z której skorzystał pasażer. Podczas przeglądania próby kwestionariuszy można by na przykład stwierdzić, że żaden użytkownik nie leciał na tej trasie więcej niż trzema różnymi liniami lotniczymi w ciągu ostatnich sześciu miesięcy. Co więcej, można odkryć, że wskutek dużej liczby linii lotniczych, 10 innych linii lotniczych znajduje się w kategorii Inne. Wykorzystując metodę wielokrotnych odpowiedzi, zostałyby zdefiniowane trzy zmienne, każda kodowana jako 1 = *american*, 2 = *united*, 3 = *twa*, 4 = *usair*, 5 = *delta*, i tak dalej. Jeśli określony pasażer zakreśli linie American i TWA, to pierwsza zmienna otrzyma kod 1, druga kod 3, a trzecia kod braku danych. Kolejny pasażer mógłby zakreślić linię American i Delta. Wskutek tego pierwsza zmienna otrzyma kod 1, druga kod 5, a trzecia kod braku danych. Z drugiej strony, przy metodzie wielokrotnych dychotomii otrzymalibyśmy 14 oddzielnych zmiennych. Chociaż każda z metod odwzorowania jest wykonalna dla potrzeb tego badania, wybór metody zależy od rozkładu odpowiedzi.

Definiowanie zestawów wielokrotnych odpowiedzi

Procedura definiowania zestawów wielokrotnych odpowiedzi grupuje podstawowe zmienne w zestawy wielokrotnych dychotomii i wielokrotnych kategorii, dla których można utworzyć tabele częstości i tabele krzyżowe. Możliwe jest zdefiniowanie do 20 zestawów wielokrotnych odpowiedzi. Każdy zestaw musi posiadać unikatową nazwę. Aby usunąć zestaw, należy go podświetlić na liście zestawów wielokrotnych odpowiedzi i kliknąć przycisk **Usuń**. Aby zmienić zestaw, należy go podświetlić na liście, zmodyfikować dowolne elementy definicji zestawu i kliknąć przycisk **Zmień**.

Zmienne podstawowe można kodować jako dychotomie lub kategorie. Aby używać zmiennych dychotomicznych, należy wybrać opcję **Dychotomie** w celu utworzenia zestawu wielokrotnych dychotomii. Należy wprowadzić wartość całkowitą dla wartości zliczanej. Każda zmienna posiadająca przynajmniej jedno wystąpienie wartości zliczanej staje się kategorią zestawu wielokrotnych dychotomii. Aby utworzyć zestaw wielokrotnych kategorii o tym samym zakresie wartości co zmienne składowe, należy wybrać opcję **Kategorie**. Dla kategorii zestawu wielokrotnych kategorii należy wprowadzić wartości całkowite dla minimalnych i maksymalnych wartości zakresu. Procedura podsumowuje każdą odrębną wartość całkowitą w przedziale zamkniętym poprzez wszystkie zmienne składowe. Puste kategorie nie są ujmowane w tabeli.

Do każdego zestawu wielokrotnych odpowiedzi musi być przydzielona unikatowa nazwa składająca się maksymalnie z siedmiu znaków. Procedura poprzedza przydzieloną nazwę przedrostkiem znaku dolara (\$). Nie można wykorzystać następujących zarezerwowanych nazw: *casenum*, *sysmis*, *jdate*, *date*, *time*, *length* i *width*. Nazwa zestawu wielokrotnych odpowiedzi istnieje wyłącznie w celu wykorzystania w procedurach wielokrotnych odpowiedzi. W innych procedurach nie można nawiązywać do nazw zestawów wielokrotnych odpowiedzi. Opcjonalnie dla zestawu wielokrotnych odpowiedzi można wprowadzić opisową etykietę zmiennej. Etykieta może składać się maksymalnie z 40 znaków.

Definiowanie zestawów wielokrotnych odpowiedzi

1. Z menu wybierz:

Analiza > Wielokrotne odpowiedzi > Definiuj zestawy zmiennych...

2. Wybierz co najmniej dwie zmienne.

3. Jeżeli zmienne kodowane są jako dychotomie, należy wskazać wartość, która ma być zliczana. Jeśli zmienne kodowane są jako kategorie, należy zdefiniować zakres kategorii.

4. Wprowadź unikatową nazwę dla każdego zestawu wielokrotnych odpowiedzi.

5. Kliknij przycisk **Dodaj**, aby dodać zestaw wielokrotnych odpowiedzi do listy zdefiniowanych zestawów.

Częstości wielokrotnych odpowiedzi

Procedura Częstości wielokrotnych odpowiedzi pozwala na tworzenie tabel częstości dla zestawów wielokrotnych odpowiedzi. Należy najpierw zdefiniować co najmniej jeden zestaw wielokrotnych odpowiedzi (patrz: "Definiowanie zestawów wielokrotnych odpowiedzi").

W przypadku zestawów wielokrotnych dychotomii nazwy kategorii pokazywane w raporcie pochodzą z etykiet zmiennych zdefiniowanych dla zmiennych elementarnych w grupie. Jeśli etykiety zmiennych nie są zdefiniowane, to jako etykiety wykorzystane są nazwy zmiennych. W przypadku zestawów wielokrotnych kategorii etykiety kategorii pochodzą z etykiet wartości pierwszej zmiennej w grupie. Jeśli brakujące kategorie dla pierwszej zmiennej są obecne dla innych zmiennych w grupie, to należy zdefiniować etykietę wartości dla brakujących kategorii.

Braki danych. Obserwacje z brakami danych są wyłączone odrębnie dla każdej tabeli. Alternatywnie można wybrać jedną lub obie z następujących opcji:

- **Wyłączanie obserwacji w obrębie dychotomii.** Użycie tej opcji powoduje wyłączenie z analizy obserwacji z brakami danych dla dowolnej zmiennej w zestawie wielokrotnych dychotomii. Ma ona zastosowanie tylko do zestawów wielokrotnych odpowiedzi zdefiniowanych jako zestawy dychotomii. Domyślnie program przyjmuje, że w przypadku zestawu wielokrotnych dychotomii obserwacja jest traktowana jako brakująca, jeśli żadna z jej zmiennych składowych nie zawiera zliczanej wartości. Obserwacje z brakami danych dla niektórych (ale nie wszystkich) zmiennych są włączane do tabeli grupy, jeśli co najmniej jedna zmienna zawiera zliczaną wartość.
- **Wyłączanie obserwacji w obrębie kategorii.** Użycie tej opcji powoduje wyłączenie z analizy obserwacji z brakami danych dla dowolnej zmiennej w zestawie wielokrotnych kategorii. Ma ona zastosowanie tylko do zestawów wielokrotnych odpowiedzi zdefiniowanych jako zestawy kategorii. Domyślnie program przyjmuje, że w przypadku zestawu wielokrotnych kategorii obserwacja jest traktowana jako brakująca tylko wówczas, gdy żaden z jej składników nie ma wartości należącej do zdefiniowanego przedziału wartości.

Przykład. Każda zmienna utworzona z pytania ankietowego jest zmienną elementarną. Aby przeanalizować element wielokrotnej odpowiedzi, należy połączyć zmienne w jeden z dwóch typów zestawów wielokrotnych odpowiedzi: zestaw wielokrotnych dychotomii lub zestaw wielokrotnych kategorii. Na przykład jeśli w ankiecie dotyczącej linii lotniczych pytano, którymi z trzech linii (American, United, TWA) pasażerowie latali w ciągu ostatnich sześciu miesięcy i wykorzystano by zmienne dychotomiczne oraz zdefiniowany **zestaw wielokrotnych dychotomii**, to każda z trzech zmiennych w zestawie stałaby się kategorią zmiennej grupy. Liczebności oraz procenty dla trzech linii wyświetlone są w jednej tabeli częstości. W przypadku gdy żaden z respondentów nie wymienił więcej niż dwóch linii lotniczych, możliwe jest utworzenie dwóch zmiennych, z których każda posiada trzy kody, po jednym dla każdej linii. Jeśli zdefiniowany został **zestaw wielokrotnych kategorii**, wartości ujmowane

są w tabelę poprzez dodawanie do siebie tych samych kodów w zmiennych elementarnych. Wynikowy zestaw wartości jest taki sam jak zestawy dla każdej ze zmiennych elementarnych. Na przykład, 30 odpowiedzi United to suma pięciu odpowiedzi United dla linii lotniczych 1 i 25 odpowiedzi United dla linii lotniczych 2. Liczebności oraz procenty dla trzech linii wyświetlone są w jednej tabeli częstości.

Statystyki. Tabele częstości pokazują liczebności, procenty odpowiedzi, procenty obserwacji, liczbę poprawnych obserwacji oraz liczbę brakujących obserwacji.

Wymagania dotyczące danych w częstościach wielokrotnych odpowiedzi

Dane. Należy używać zestawów wielokrotnych odpowiedzi.

Założenia. Liczebności i procenty stanowią użyteczny opis danych z każdego rozkładu.

Procedury pokrewne. Procedura Zestawy wielokrotnych odpowiedzi umożliwia zdefiniowanie takich zestawów.

Otrzymywanie Częstości wielokrotnych odpowiedzi

1. Z menu wybierz:

Analiza > Wielokrotne odpowiedzi > Częstości...

2. Wybierz co najmniej jeden zestaw wielokrotnych odpowiedzi.

tabele krzyżowe wielokrotnych odpowiedzi

Użycie procedury Tabele krzyżowe wielokrotnych odpowiedzi powoduje umieszczenie w tabeli krzyżowej zdefiniowanych zestawów wielokrotnych odpowiedzi, zmiennych elementarnych lub kombinacji. Możliwe jest również obliczenie odsetka komórek na podstawie obserwacji lub odpowiedzi, modyfikacja obsługi braków danych lub otrzymanie zależnych tabel krzyżowych. Należy najpierw zdefiniować co najmniej jeden zestaw wielokrotnych odpowiedzi (patrz: „Definiowanie zestawów wielokrotnych odpowiedzi”).

W przypadku zestawów wielokrotnych dychotomii nazwy kategorii pokazywane w raporcie pochodzą z etykiet zmiennych zdefiniowanych dla zmiennych elementarnych w grupie. Jeśli etykiety zmiennych nie są zdefiniowane, to jako etykiety wykorzystane są nazwy zmiennych. W przypadku zestawów wielokrotnych kategorii etykiety kategorii pochodzą z etykiet wartości pierwszej zmiennej w grupie. Jeśli brakujące kategorie dla pierwszej zmiennej są obecne dla innych zmiennych w grupie, to należy zdefiniować etykietę wartości dla brakujących kategorii. Procedura pokazuje etykiety kategorii dla kolumn składające się z trzech linii, z maksymalnie ośmioma znakami na linię. Aby uniknąć dzielenia wyrazów, można odwrócić pozycje wiersza i kolumny lub zdefiniować etykiety ponownie.

Przykład. W tej procedurze możliwe jest umieszczanie w tabeli krzyżowej z innymi zmiennymi zarówno zestawów wielokrotnych kategorii, jak i zestawów wielokrotnych dychotomii. W ankiecie przeprowadzanej wśród pasażerów linii lotniczej pytano o następujące informacje: spośród następujących linii lotniczych proszę zakreślić wszystkie, którymi latał/latała Pan/Pani przynajmniej raz w ciągu ostatnich sześciu miesięcy (American, United, TWA). Co jest dla Pana/Pani ważniejsze w wyborze lotu — rozkład lotów czy obsługa? Proszę wybrać tylko jedną odpowiedź. Po wprowadzeniu danych jako dychotomie lub wielokrotne kategorie i połączeniu ich w zestaw możliwe jest umieszczenie w tabeli krzyżowej wyborów linii z pytaniem dotyczącym obsługi lub rozkładu lotów.

Statystyki. Tabela krzyżowa z liczebnościami w komórce, wierszu, kolumnie i ogółem oraz z procentami w komórce, wierszu, kolumnie i ogółem. Odsetki komórek mogą być obliczane na podstawie obserwacji lub odpowiedzi.

Wymagania dotyczące danych w tabelach krzyżowych wielokrotnych odpowiedzi

Dane. Należy używać zestawów wielokrotnych odpowiedzi lub kategorialnych zmiennych numerycznych.

Założenia. Liczebności i procenty zapewniają użyteczny opis danych z każdego rozkładu.

Procedury pokrewne. Procedura Zestawy wielokrotnych odpowiedzi umożliwia zdefiniowanie takich zestawów.

Otrzymywanie tabel krzyżowych wielokrotnych odpowiedzi

1. Wybierz z menu następującą opcję:

Analiza > Wielokrotne odpowiedzi > Tabele krzyżowe...

- Wybierz co najmniej jedną zmienną numeryczną lub zestaw wielokrotnych odpowiedzi dla każdego wymiaru tabeli krzyżowej.
- Zdefiniuj zakres każdej zmiennej elementarnej.

Opcjonalnie możliwe jest otrzymanie dwuwymiarowej tabeli krzyżowej dla każdej kategorii zmiennej sterującej lub zestawu wielokrotnych odpowiedzi. Wybierz co najmniej jedną pozycję i dodaj do listy Warstwy.

Tabele krzyżowe: Definiuj zakres zmiennej

Zakresy wartości muszą być zdefiniowane dla każdej zmiennej elementarnej w tabeli krzyżowej. Należy wprowadzić minimalne i maksymalne wartości całkowite kategorii, które mają być ujęte w tabeli. Kategorie znajdujące się poza zakresem są wyłączone z analizy. Program przyjmuje, że wartości wewnątrz przedziału zamkniętego są liczbami całkowitymi (wartości nie będące liczbami całkowitymi są obcinane).

Tabele krzyżowe wielokrotnych odpowiedzi: Opcje

Procent w komórce. Liczby komórek są zawsze pokazywane. Możliwy jest wybór opcji wyświetlania procentu w wierszu, procentu w kolumnie i procentu tabeli drugiego rzędu (ogółem).

Procenty na podstawie. Możliwe jest obliczanie odsetka komórek na podstawie obserwacji (lub odpowiedzi). Nie jest to możliwe, jeśli wybrano opcję łączenia zmiennych poprzez zestawy wielokrotnych kategorii. Procent w komórce można również obliczać na podstawie odpowiedzi. W przypadku zestawów wielokrotnych dychotomii liczba odpowiedzi równa jest liczbie zliczanych wartości poprzez obserwację. W przypadku zestawów wielokrotnych kategorii liczba odpowiedzi jest liczbą wartości w zdefiniowanym zakresie.

Braki danych. Można wybrać jedną lub obie z następujących opcji:

- Wyłączanie obserwacji w obrębie dychotomii.** Użycie tej opcji powoduje wyłączenie z analizy obserwacji z brakami danych dla dowolnej zmiennej w zestawie wielokrotnych dychotomii. Ma ona zastosowanie tylko do zestawów wielokrotnych odpowiedzi zdefiniowanych jako zestawy dychotomii. Domyślnie program przyjmuje, że w przypadku zestawu wielokrotnych dychotomii obserwacja jest traktowana jako brakująca, jeśli żadna z jej zmiennych składowych nie zawiera zliczanej wartości. Obserwacje z brakami danych dla niektórych (ale nie wszystkich) zmiennych są włączane do tabeli grupy, jeśli co najmniej jedna zmienna zawiera zliczaną wartość.
- Wyłączanie obserwacji w obrębie kategorii.** Użycie tej opcji powoduje wyłączenie z analizy obserwacji z brakami danych dla dowolnej zmiennej w zestawie wielokrotnych kategorii. Ma ona zastosowanie tylko do zestawów wielokrotnych odpowiedzi zdefiniowanych jako zestawy kategorii. Domyślnie program przyjmuje, że w przypadku zestawu wielokrotnych kategorii obserwacja jest traktowana jako brakująca tylko wówczas, gdy żaden z jej składników nie ma wartości należącej do zdefiniowanego przedziału wartości.

Domyślnie program przyjmuje, że podczas umieszczania w tabeli dwóch zestawów wielokrotnych kategorii, procedura ujmuje w tabeli każdą zmienną z pierwszej grupy z każdą zmienną z drugiej grupy i sumuje liczebności dla każdej komórki. Dlatego też niektóre odpowiedzi mogą pojawiać się w tabeli więcej niż jeden raz. Można wybrać następującą opcję:

Połącz zmienne poprzez zestawy odpowiedzi. Pierwsza zmienna z pierwszej grupy jest łączona w parę z pierwszą zmienną z drugiej grupy itd. Wybranie tej opcji powoduje, że program oblicza odsetek komórek na podstawie odpowiedzi, a nie respondentów. Łączenie to nie jest dostępne dla zestawów wielokrotnych dychotomii lub zmiennych elementarnych.

Dodatkowe właściwości komendy Wielokrotne odpowiedzi (MULT RESPONSE)

Język składni komend umożliwia również:

- Otrzymanie tabel krzyżowych z maksymalnie pięcioma wymiarami (za pomocą opcji komendy BY).

- Zmianę opcji formatowania wyników, łącznie z usunięciem etykiet wartości (za pomocą opcji komendy FORMAT).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Przedstawianie wyników

Przedstawianie wyników

Podsumowania obserwacji oraz statystyki opisowe są podstawowymi narzędziami analizy i prezentacji danych. Podsumowanie obserwacji uzyskać można za pomocą Edytora danych lub przy użyciu procedury podsumowania, częstości i statystyki opisowe – przy użyciu procedury częstości, a statystyki podpopulacji przy użyciu procedury średnich. Każda z nich wykorzystuje format pozwalający na przejrzyste przedstawienie informacji. Na przedstawienie informacji w innym formacie pozwalają takie opcje menu Analiza, jak Raport w wierszach... i Raport w kolumnach...

Raport – podsumowania w wierszach

Okno dialogowe Raport: Podsumowania w wierszach... umożliwia tworzenie raportów, w których różne statystyki podsumowujące przedstawione są w wierszach. Dostępne są również podsumowania obserwacji, zarówno razem ze statystykami podsumowującymi, jak i bez nich.

Przykład. Przedsiębiorstwo prowadzące sieć sklepów przechowuje informacje dotyczące pracowników, takie jak płaca, staż pracy oraz sklep i dział, w którym pracownik jest zatrudniony. Można wygenerować raport, umożliwiający przedstawienie listy z informacjami dotyczącymi poszczególnych pracowników z podziałem na poszczególne sklepy i działy (zmiennie dzielące), z uwzględnieniem statystyk podsumowujących (takich jak średnia pensja) dla każdego sklepu, działu oraz dla poszczególnych działów w obrębie każdego sklepu.

Kolumny danych. W tym polu wyświetlana jest lista zmiennych raportu, dla których przygotowano zostanie podsumowanie obserwacji lub statystyki podsumowujące. Można tutaj również określić format wyświetlania kolumn danych.

Kolumny grupujące. W tym polu wyświetlana jest lista opcjonalnych zmiennych dzielących, które pozwalają dzielić raport na grupy. Można tutaj również określić tworzone statystyki podsumowujące i formatowanie kolumn grupujących. W przypadku wielu zmiennych dzielących każdej kategorii każdej zmiennej dzielącej w obrębie kategorii poprzedniej zmiennej dzielącej na liście odpowiada oddzielna grupa. Zmienne dzielące powinny być zmiennymi kategorialnymi typu dyskretnego, które pozwalają na podział obserwacji na ograniczoną liczbę znaczących kategorii. Poszczególne wartości każdej ze zmiennych dzielących wyświetlane są w postaci posortowanej w osobnej kolumnie z lewej strony kolumn danych.

Raport. Ta grupa przycisków pozwala na sterowanie ogólnymi właściwościami raportów, takimi jak całościowe statystyki podsumowujące, wyświetlanie braków danych, numerowanie stron oraz tytuły.

Pokaż obserwacje. Zaznaczenie tego pola wyboru powoduje wyświetlanie wartości (lub etykiet wartości) zmiennych kolumn danych dla każdej obserwacji. Powoduje to utworzenie raportu wyszczególniającego, który może być dużo dłuższy od raportu podsumowującego.

Podgląd. Zaznaczenie tego pola powoduje wyświetlenie jedynie pierwszej strony raportu. Opcja ta jest użyteczna w celu przejrzania formatu raportu, bez potrzeby przetwarzania całego raportu.

Dane są już posortowane. W przypadku raportów ze zmiennymi dzielącymi, plik danych musi być posortowany według wartości zmiennych dzielących przed wygenerowaniem raportu. Jeśli plik danych jest już posortowany według wartości zmiennych dzielących, można zmniejszyć czas przetwarzania przez zaznaczenie tej opcji. Jest ona szczególnie użyteczna po dokonaniu podglądu raportu.

Otrzymywanie raportu podsumowania: Podsumowania w wierszach

1. Z menu wybierz:

Analizuj > Raporty > Raport – Podsumowania w wierszach...

2. Wybierz co najmniej jedną zmienną i umieść ją na liście Kolumny danych. Dla każdej z wybranych zmiennych tworzona jest w raporcie osobna kolumna.
3. W przypadku raportów sortowanych i wyświetlanych według podgrup wybierz co najmniej jedną zmienną i umieść ją na liście Kolumny grupujące.
4. W przypadku raportów ze statystykami podsumowującymi dla grup zdefiniowanych przez zmienne dzielące wybierz zmienną dzielącą z listy Przerwij kolumny z danymi i kliknij przycisk **Podsumowanie** w grupie Kolumny grupujące, aby określić miary podsumowania.
5. W przypadku raportów zawierających ogólne statystyki podsumowujące kliknij przycisk **Podsumowanie**, aby określić miary podsumowania.

Raport: Format kolumny danych/grupującej

Okna dialogowe formatowania pozwalają na określanie tytułów i szerokości kolumn, wyrównania tekstu i sposobu wyświetlania wartości danych lub ich etykiet. Okno formatu kolumny danych pozwala na określenie formatu kolumn danych, znajdujących się po prawej na stronie raportu. Okno formatowania podziału pozwala na określanie formatu kolumn grupujących, znajdujących się po lewej stronie.

Tytuł kolumny. Pole to pozwala na określenie tytułu kolumny dla wybranej zmiennej. Długie tytuły są automatycznie zawijane w kolumnie. Przy użyciu klawisza Enter można ręcznie wstawiać znaki podziału wiersza tam, gdzie tytuły mają się zawijać.

Położenie wartości w kolumnie. Grupa ta pozwala na określanie sposobu wyrównywania wartości danych lub etykiet wartości w kolumnie wybranej zmiennej. Sposób wyrównania wartości lub ich etykiet nie wpływa na sposób wyrównania nagłówków kolumn. Zawartość kolumny można przesunąć o określoną liczbę znaków lub wyśrodkować.

Zawartość kolumny. W grupie tej określić można sposób wyświetlania wartości danych lub zdefiniowanych etykiet wartości dla wybranej zmiennej. Dla każdej wartości danych bez zdefiniowanej etykiety zawsze wyświetlana jest sama wartość (funkcja ta nie jest dostępna w przypadku kolumn danych w raportach podsumowania w kolumnach).

Raport: Wiersze podsumowań dla/Wiersze podsumowań końcowych

Dwa okna dialogowe Raport: Wiersze podsumowań służą do określania sposobu wyświetlania statystyk podsumowujących dla kolumn grupujących i dla całego raportu. W oknie Raport: Wiersze podsumowań dla można określać statystyki podgrup dla każdej kategorii zdefiniowanej przez zmienne dzielące. Okno Raport: Wiersze podsumowań końcowych służy natomiast do określania statystyk całłościowych, które są wyświetlane na końcu raportu.

Dostępne statystyki podsumowujące to: suma wartości, średnia z wartości, wartość minimalna, wartość maksymalna, liczba obserwacji, procent obserwacji powyżej i poniżej określonej wartości, procent obserwacji wewnątrz określonego zakresu wartości, odchylenie standardowe, wariancja, kurtoza oraz skośność.

Raport: Opcje podziału dla

W oknie Raport: Opcje podziału dla określać można odstępy i podział na strony dla informacji wybranej kategorii grupującej.

Ustawienia strony. W grupie tej określać można odstępy i podział na strony dla kategorii wybranej zmiennej dzielącej. Można określić liczbę pustych wierszy między kategoriami grupującymi lub zaznaczyć opcję powodującą rozpoczęcie każdej kategorii na nowej stronie.

Puste wiersze przed podsumowaniami. Opcja ta pozwala na określanie liczby pustych wierszy między etykietami kategorii grupujących lub między danymi a statystykami podsumowującymi. Wprowadzenie pustego wiersza między wykazami obserwacji a statystykami podsumowującymi jest szczególnie użyteczne w przypadku raportów połączonych, składających się zarówno z wykazu pojedynczych obserwacji, jak i ze statystyk podsumowujących dla kategorii grupujących.

Opcje raportu

W oknie Raport: Opcje można określić sposób przetwarzania i wyświetlania braków danych oraz numerację stron raportu.

Wyłączenie obserwacji z brakującymi wartościami. Zaznaczenie tego pola wyboru powoduje eliminację (z raportu) wszelkich obserwacji z brakami danych dla dowolnych zmiennych raportu.

Braki danych pokazywane są jako. W polu tym określić można symbol, który w pliku danych oznaczać będzie brak danych. Ten symbol składający się z jednego tylko znaku jest używany do przedstawiania *systemowych* i *zdefiniowanych* braków danych.

Rozpocznij numerację stron od. W polu tym można określić numer pierwszej strony raportu.

Raport: Układ

Okno dialogowe Raport: Układ pozwala na określanie szerokości i długości każdej strony raportu, a także sposobu rozmieszczenia raportu na stronie oraz wstawiania pustych wierszy i etykiet.

Układ strony. Definiuje marginesy strony, określane w wierszach (od góry i od dołu) i znakach (z lewej i z prawej), i wyrównanie w odniesieniu do marginesów.

Nagłówki i stopki. Ta grupa pozwala na określanie liczby wierszy wstawianych po tytule strony i przed stopką w celu oddzielenia ich od tekstu raportu.

Kolumny grupujące. Ta grupa umożliwia sterowanie sposobem wyświetlania kolumn grupujących. Jeśli określonych jest wiele zmiennych dzielących, mogą one być przedstawione w oddzielnych kolumnach lub w pierwszej kolumnie. Umieszczenie wszystkich zmiennych dzielących w pierwszej kolumnie umożliwia tworzenie węższych raportów.

Tytuły kolumn. W tej grupie określać można sposób wyświetlania tytułów kolumn, z uwzględnieniem podkreślenia tytułów, odstępów między nimi a treścią raportu oraz sposobu ich pionowego wyrównania.

Wiersze kolumny danych i etykiety grup. Określa położenie informacji o kolumnie danych (wartości danych i/lub statystyk podsumowujących) w odniesieniu do etykiet grup na początku każdej kategorii grupowania. Pierwszy wiersz informacji o kolumnie danych może zaczynać się w tym samym wierszu co etykieta kategorii grupowania lub w określonym wierszu po etykietce kategorii grupowania (nie dotyczy podsumowujących raportów kolumnowych).

Raport: Tytuły

Okno Raport: Tytuły pozwala na kontrolowanie zawartości i rozmieszczenia tytułów oraz stopek w raportach. Określić można do 10 wierszy tytułów strony i do 10 wierszy stopek, w każdym z nich określając 3 składniki: Z lewej, W środku i Z prawej.

W przypadku wprowadzenia zmiennych do tytułów lub stopek, w tytule lub stopce wyświetlana będzie aktualna etykieta wartości lub wartość zmiennej. W tytułach wyświetlana jest etykieta wartości odpowiadająca wartości zmiennej na początku strony. W stopkach wyświetlana jest etykieta odpowiadająca wartości zmiennej u dołu strony. Jeżeli etykieta wartości nie istnieje, to wyświetlana jest rzeczywista wartość.

Zmienne specjalne. Zmienne specjalne, takie jak *DATE* i *PAGE*, pozwalają na wstawianie aktualnej daty lub numeru strony do dowolnego wiersza nagłówka lub stopki raportu. Jeśli w pliku danych występują zmienne o nazwach *DATE* lub *PAGE*, nie można ich używać w tytułach i stopkach raportów.

Raport – podsumowania w kolumnach

Okno Raport: Podsumowania w kolumnach umożliwia tworzenie raportów podsumowujących, w których różne statystyki podsumowujące są wyświetlane w oddzielnych kolumnach.

Przykład. Przedsiębiorstwo prowadzące sieć sklepów przechowuje informacje dotyczące pracowników, takie jak płaca, staż pracy oraz dział, w którym pracownik jest zatrudniony. Możliwe jest wygenerowanie raportu zawierającego statystyki podsumowujące dotyczące pensji (na przykład średniej, minimalnej i maksymalnej) dla każdego działu.

Kolumny danych. W tym polu wyświetlana jest lista zmiennych raportu, dla których przygotowano zostaną statystyki podsumowujące. Umożliwia też ono określenie formatu wyświetlania i statystyk podsumowujących wyświetlanych dla każdej zmiennej.

Kolumny grupujące. W tym polu wyświetlana jest lista opcjonalnych zmiennych dzielących, które pozwalają dzielić raport na grupy i określać formatowanie kolumn grupujących. W przypadku wielu zmiennych dzielących każdej kategorii każdej zmiennej dzielącej w obrębie kategorii poprzedniej zmiennej dzielącej na liście odpowiada oddzielna grupa. Zmienne dzielące powinny być zmiennymi kategorialnymi typu dyskretnego, które pozwalają na podział obserwacji na ograniczoną liczbę znaczących kategorii.

Raport. Ta grupa przycisków pozwala na określanie ogólnych właściwości raportów, takich jak wyświetlanie braków danych, numerowanie stron oraz tytuły.

Podgląd. Zaznaczenie tego pola powoduje wyświetlenie jedynie pierwszej strony raportu. Opcja ta jest użyteczna w celu przejrzania formatu raportu, bez potrzeby przetwarzania całego raportu.

Dane są już posortowane. W przypadku raportów ze zmiennymi dzielącymi, plik danych musi być posortowany według wartości zmiennych dzielących przed wygenerowaniem raportu. Jeśli plik danych jest już posortowany według wartości zmiennych dzielących, można zmniejszyć czas przetwarzania przez zaznaczenie tej opcji. Jest ona szczególnie użyteczna po dokonaniu podglądu raportu.

Otrzymywanie raportu podsumowania: Podsumowania w kolumnach

1. Z menu wybierz:

Analiza > Raporty > Raport – Podsumowania w kolumnach...

2. Wybierz co najmniej jedną zmienną i umieść ją na liście Kolumny danych. Dla każdej z wybranych zmiennych tworzona jest w raporcie osobna kolumna.
3. Aby zmienić miarę podsumowania dla zmiennej, wybierz zmienną z listy Zmienne kolumn z danymi i kliknij przycisk **Podsumowanie**.
4. Aby uzyskać więcej miar podsumowań dla zmiennej, wybierz zmienną z listy źródłowej i przenieś ją na listę Zmienne kolumn z danymi tyle razy, ile miar podsumowań potrzebujesz.
5. Aby wyświetlić kolumnę zawierającą sumę, średnią czy inne funkcje istniejących kolumn, kliknij przycisk **Wstaw podsumowanie**. Spowoduje to umieszczenie na liście Kolumny danych zmiennej o nazwie *Ogółem*.
6. W przypadku raportów sortowanych i wyświetlanych według podgrup wybierz co najmniej jedną zmienną i umieść ją na liście Kolumny grupujące.

Funkcja podsumowująca kolumny danych

Okno Raport: Wiersze podsumowań pozwala wybrać statystykę podsumowującą wyświetlaną dla zmiennej wybranej z listy Kolumny danych.

Dostępne statystyki podsumowujące to: suma wartości, średnia z wartości, wartość minimalna, wartość maksymalna, liczba obserwacji, procent obserwacji powyżej i poniżej określonej wartości, procent obserwacji wewnątrz określonego zakresu wartości, odchylenie standardowe, wariancja, kurtoza oraz skośność.

Podsumowanie kolumn danych dla kolumny ogółem

Okno dialogowe Raport: Kolumna podsumowań służy do określania ogólnych statystyk podsumowujących dwie lub więcej kolumn danych.

Dostępne ogólne statystyki podsumowujące to: suma dla kolumn, średnia dla kolumn, minimum dla kolumn, maksimum dla kolumn, różnica między wartościami pierwszej i drugiej kolumny, iloraz wartości pierwszej i drugiej kolumny lub iloczyn wartości pierwszej i drugiej kolumny.

Suma dla kolumn. Wartości w kolumnie *Ogółem* stanowią sumę wartości kolumn występujących na liście Kolumna podsumowań.

Średnia dla kolumn. Wartości w kolumnie *Ogółem* stanowią średnią wartości kolumn występujących na liście Kolumna podsumowań.

Minimum dla kolumn. Wartości w kolumnie *Ogółem* stanowią minimum wartości kolumn występujących na liście Kolumna podsumowań.

Maksimum dla kolumn. Wartości w kolumnie *Ogółem* stanowią maksimum wartości kolumn występujących na liście Kolumna podsumowań.

1. kolumna - 2. kolumna. Wartości w kolumnie *Ogółem* stanowią różnicę wartości kolumn występujących na liście Kolumna podsumowań. Na liście Kolumna podsumowań umieszczone muszą być dokładnie dwie kolumny.

1. kolumna / 2. kolumna. Wartości w kolumnie *Ogółem* stanowią iloraz wartości kolumn występujących na liście Kolumna podsumowań. Na liście Kolumna podsumowań umieszczone muszą być dokładnie dwie kolumny.

% 1. kolumna / 2. kolumna. Wartości w kolumnie *Ogółem* stanowią procenty wartości pierwszej kolumny z wartości drugiej kolumny. Na liście Kolumna podsumowań umieszczone muszą być dokładnie dwie kolumny.

Iloczyn kolumn. Wartości w kolumnie *Ogółem* stanowią iloczyn wartości kolumn występujących na liście Kolumna podsumowań.

Raport: Format kolumny danych

Opcje formatowania kolumn danych i kolumn grupujących dla raportów z podsumowaniami w kolumnach są takie same, jak opisane dla raportów z podsumowaniami w wierszach.

Raport: Opcje podziału dla raportu z podsumowaniami w kolumnach

Opcje podziału umożliwiają sterowanie wyświetlaniem sumy pośredniej, odstępami i podziałem na strony dla kategorii grupujących.

Suma pośrednia. Opcje tej grupy pozwalają na wyświetlanie sum pośrednich dla kategorii grupujących.

Ustawienia strony. W grupie tej określać można odstępy i podział na strony dla kategorii wybranej zmiennej dzielącej. Można określić liczbę pustych wierszy między kategoriami grupującymi lub zaznaczyć opcję powodującą rozpoczynanie każdej kategorii na nowej stronie.

Puste wiersze przed sumą pośrednią. W tej grupie można określić liczbę pustych wierszy, wstawianych między danymi kategorii grupujących i sumami pośrednimi.

Raport: Opcje raportu z podsumowaniami w kolumnach

Opcje pozwalają na kontrolowanie sposobu wyświetlania podsumowania całości, braków danych oraz sposobu numeracji stron w raportach z podsumowaniami w kolumnach.

Podsumowanie całości. Zaznaczenie tej opcji powoduje wyświetlanie podsumowania całości dla każdej kolumny u dołu tejże kolumny.

Brakujące wartości. Braki danych można wyłączyć z raportu lub wskazać pojedynczy znak, za pomocą którego braki te będą symbolizowane w raporcie.

Raport: Układ dla raportu z podsumowaniami w kolumnach

Opcje układu dla raportów z podsumowaniami w kolumnach są takie same, jak opisane dla raportów z podsumowaniem w wierszach.

Dodatkowe właściwości komendy REPORT

Język składni komend umożliwia również:

- Wyświetlanie różnych funkcji podsumowujących w kolumnach pojedynczego wiersza podsumowania.

- Wstawianie wierszy podsumowań do kolumn danych dla zmiennych innych niż występujące na liście Kolumny danych, a także dla różnych kombinacji funkcji podsumowań (funkcji złożonych).
- Korzystanie z funkcji podsumowujących w postaci mediany, dominanty, częstości i procentu.
- Dokładniejsze kontrolowanie formatu wyświetlania statystyk podsumowujących.
- Wstawianie pustych wierszy w różnych miejscach raportów.
- Wstawianie pustych wierszy po każdej n -tej obserwacji w raportach wyszczególniających.

Ze względu na złożoność składni REPORT przy tworzeniu nowego raportu za pomocą składni korzystne może być utworzenie podobnego raportu za pomocą okien dialogowych, a następnie wklejenie odpowiadającej mu składni do okna Edytora komend i dokonanie w nich niezbędnych poprawek, co pozwoli na otrzymanie zamierzonego efektu.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Analiza rzetelności

Analiza rzetelności umożliwia zbadanie właściwości skali pomiarowej oraz elementów, które budują skalę. Procedura analizy rzetelności umożliwia obliczenie liczby zwykle używanych miar skali rzetelności oraz dostarcza informacji na temat związków między pojedynczymi pozycjami na skali. Współczynniki korelacji mogą być wykorzystane do obliczenia rzetelności rozumianej jako zgodność między oceniającymi.

Analiza rzetelności udostępnia także statystykę kappa wielu oceniających Fleissa, która ocenia zgodność między oceniającymi w celu określenia rzetelność w grupie wielu oceniających. Większa zgodność oznacza większą pewność co do tego, że oceny odzwierciedlają prawdziwe okoliczności. Opcje kappa wielu oceniających Fleissa są dostępne w oknie dialogowym “Analiza rzetelności: Statystyki” na stronie 229.

Przykład

Czy kwestionariusz mierzy zadowolenie klientów w użyteczny sposób? Wykorzystując analizę rzetelności można ustalić stopień powiązania ze sobą pozycji kwestionariusza, uzyskać ogólny wskaźnik powtarzalności lub wewnętrznej spójności skali jako całości, a także zidentyfikować pozycje problemowe, które powinny być wyłączone ze skali.

Statystyki

Statystyki opisowe dla każdej zmiennej i dla skali, statystyki podsumowujące dla wielu pozycji, korelacje oraz kowariancje między pozycjami, oszacowania rzetelności, tabela ANOVA, współczynniki korelacji wewnątrzklasowej, T^2 Hotellinga, test addytywności Tukeya i kappa wielu oceniających Fleissa.

Modele

Dostępne są następujące modele rzetelności:

Alfa (Cronbacha)

Ten model jest miarą wewnętrznej spójności, opartą na średniej korelacji między pozycjami.

Omega (McDonald's)

W tym modelu założono, że model jest jednowymiarowy, w tym pojedynczy czynnik bez zależności od pozycji lokalnej w postaci kowariancji błędów. Model zakłada, że kowariancja dwóch różnych elementów jest iloczynem ich obciążeń.

Półówkowy

Ten model dzieli skalę na dwie części i analizuje korelację między nimi.

Guttmana

Ten model oblicza dolne granice Guttmana dla prawdziwej rzetelności.

Równoległe

W tym modelu przyjęto założenie, że wszystkie pozycje mają równe wariancje oraz równe wariancje błędów w replikacjach.

Ściśle równoległy

W tym modelu przyjęto założenia modelu równoległego oraz założenie równej średniej w pozycjach.

Wymagania dotyczące danych do analizy rzetelności

Dane

Dane mogą być dychotomiczne, porządkowe lub interwałowe, ale nie powinny być kodowane liczbowo.

Założenia

Obserwacje powinny być niezależne, a błędy nieskorelowane między pozycjami. Każda para pozycji powinna być skorelowana i wykazywać rozkład normalny. Skale powinny być addytywne, tak aby każda pozycja była liniowo powiązana z ogólną oceną. W odniesieniu do statystyk kappa wielu oceniających Fleissa stosuje się następujące założenia:

- Aby obliczyć statystykę rzetelności, należy określić co najmniej dwie zmienne pozycji.
- Jeśli wybrano co najmniej dwie zmienne oceny, wklejana jest komenda kappa wielu oceniających Fleissa.
- Nie ma żadnego związku pomiędzy oceniającymi.
- Liczba oceniających jest stała.
- Każdy obiekt jest oceniany przez tę samą grupę zawierającą tylko jednego oceniającego.
- Nie można przypisać wag do różnych niezgodności.

Procedury pokrewne

W celu zbadania wymiarowości pozycji skali (sprawdzenia, czy potrzebna jest więcej niż jedna struktura do wyjaśnienia układu ocen pozycji) należy użyć analizy czynnikowej lub skalowania wielowymiarowego. W celu zidentyfikowania homogenicznych grup zmiennych dla zmiennych skupiania można zastosować hierarchiczną analizę skupień.

Wykonywanie analizy rzetelności

1. Z menu wybierz:

Analizuj > Skala > Analiza rzetelności...

2. Wybierz co najmniej dwie zmienne jako potencjalne elementy skali addytywnej.

3. Wybierz model z listy rozwijanej **Model**.

4. Opcjonalnie kliknij opcję **Statystyki**, aby wybrać różne statystyki, które opisują elementy skali lub zgodność między oceniającymi.

Analiza rzetelności: Statystyki

Można wybrać różne statystyki, które opisują skalę, pozycje i zgodność między oceniającymi w celu określenia rzetelności w grupie wielu oceniającymi. Statystyki umieszczane w raportach domyślnie zawierają liczbę obserwacji, liczbę pozycji oraz oceny rzetelności, tak jak zostało to przedstawione poniżej:

Modele alfa

Współczynnik alfa dla danych dychotomicznych; jest równoważny współczynnikowi Kudara-Richardsona 20 (KR20).

Modele omega

Oszacowanie omega McDonalda w celu oceny rzetelności.

Modele połówkowe

Korelacja między połówkami, rzetelność podziału połówkowego Guttmana, rzetelność podziału Spearmana-Browna (równej i nierównej długości) i współczynnik alfa dla każdej połówki.

Modele Guttmana

Współczynniki rzetelności od lambda 1 do lambda 6.

Modele równoległe i ściśle równoległe

Test dobroci dopasowania modelu, oszacowania wariancji błędu, wariancja łączna i prawdziwa, szacowana wspólna korelacja międzypozycyjna, szacowana rzetelność, nieobciążone oszacowanie rzetelności.

Statystyki opisowe dla

Umożliwia tworzenie statystyk opisowych dla skal lub pozycji w obserwacjach.

Pozycja

Umożliwia tworzenie statystyk opisowych dla skal lub pozycji w obserwacjach.

Skala

Procedury statystyk opisowych dla skal.

Skala przy wykluczeniu pozycji

Powoduje wyświetlenie statystyk podsumowujących, które porównują każdą pozycję ze skalą złożoną z innych pozycji. Statystyki obejmują średnią i wariancję skali, jeśli pozycja miała być usunięta ze skali, korelację pomiędzy pozycją a skalą złożoną z innych pozycji i alfę Cronbacha, jeśli pozycja miała zostać usunięta ze skali.

Podsumowania

Udostępnia statystyki opisowe rozkładów pozycji we wszystkich pozycjach na skali.

Średnie

Statystyki podsumowujące dla średnich pozycji. Umożliwia wyświetlanie najmniejszych, największych i przeciętnych wartości średnich z pozycji zakresu i wariancji wartości średnich z pozycji oraz stosunku największych do najmniejszych wartości średnich z pozycji.

Wariancje

Statystyki podsumowujące dla wariancji pozycji. Wyświetlane są: najmniejsza, największa i średnia wariancja pozycji, rozstęp i wariancja z wariancji pozycji oraz stosunek wariancji największej do najmniejszej.

Korelacje

Statystyki podsumowujące korelacje między elementem. Wyświetlane są: najmniejsza, największa i średnia wartość korelacji między pozycjami, zakres i wariancja korelacji między pozycjami oraz stosunek największej do najmniejszej wartości korelacji między pozycjami.

Kowariancje

Statystyki podsumowujące dla kowariancji między pozycją. Wyświetlane są: najmniejsza, największa i średnia wartość kowariancji między pozycjami, zakres i wariancja kowariancji między pozycjami oraz stosunek największej do najmniejszej wartości kowariancji między pozycjami.

Między pozycjami

Umożliwia tworzenie macierzy korelacji lub kowariancji między pozycjami.

Tabela ANOVA

Umożliwia przeprowadzanie testu równych średnich.

Test F

Umożliwia wyświetlanie tabeli analizy wariancji powtarzanych pomiarów.

Chi-kwadrat Friedmana

Wyświetla współczynnik chi-kwadrat Friedmana i współczynnik zgodności Kendall. Opcja ta jest odpowiednia dla danych w formie rang. Test chi-kwadrat zastępuje zwykły test F w tabeli ANOVA.

Chi-kwadrat Cochra

Wyświetla Q Cochra. Ta opcja jest odpowiednia dla danych, które są dychotomiczne. Statystyka Q zastępuje typową statystykę F w tabeli ANOVA.

Zgodność między oceniającymi: kappa Fleissa

Ocenia zgodność między oceniającymi w celu określenia rzetelności w grupie wielu oceniających. Większa zgodność oznacza większą pewność co do tego, że oceny odzwierciedlają prawdziwe okoliczności. Uogólniona nieważona statystyka kappa mierzy zgodność między każdą stałą liczbą osób oceniających przy założeniu, że:

- Aby wyznaczyć statystyki rzetelności, należy określić co najmniej dwie zmienne pozycji.
- Należy określić co najmniej dwie zmienne oceny.
- Zmienne wybrane jako pozycje mogą być również wybrane jako zmienne oceny.
- Nie ma żadnego związku pomiędzy oceniającymi.
- Liczba oceniających jest stała.

- Każdy obiekt jest oceniany przez tę samą grupę zawierającą tylko jednego oceniającego.
- Nie można przypisać wag do różnych niezgodności.

Wyświetlaj zgodność co do poszczególnych kategorii

Określa, czy wyniki mają uwzględniać zgodność dotyczącą poszczególnych kategorii. Domyślnie wyniki przesłaniają oszacowania poszczególnych kategorii. Gdy ta opcja jest włączona, wyniki zawierają wiele tabel.

Ignoruj obserwacje łańcuchowe

Określa, czy w zmiennych łańcuchowych rozróżniana jest wielkość liter. Domyślnie w łańcuchowych wartościach ocen rozróżniana jest wielkość liter.

Etykiety kategorii łańcuchowych wyświetlane są wielkimi literami

Określa, czy etykiety kategorii w tabelach wynikowych są wyświetlane przy użyciu wielkich, czy małych liter. To ustawienie jest domyślnie włączone, co powoduje, że łańcuchowe etykiety kategorii są wyświetlane wielkimi literami.

Asymptotyczny poziom ufności (%)

Określa poziom istotności dla asymptotycznych przedziałów ufności. 95 jest ustawieniem domyślnym.

Brakujące

Wyklucz systemowe i zdefiniowane przez użytkownika braki danych

Kontroluje wykluczanie braków danych zdefiniowanych przez użytkownika i systemowych braków danych. Domyślnie braki danych użytkownika i systemowe braki danych są wykluczane.

Zdefiniowane braki danych są traktowane jako ważne wartości

Gdy ta opcja jest włączona, braki danych użytkownika i systemowe braki danych traktowane są jako poprawne dane. To ustawienie jest domyślnie wyłączone.

T-kwadrat Hotellinga

Umożliwia przeprowadzanie wielowymiarowego testu hipotezy zerowej, że wszystkie pozycje na skali mają tę samą średnią.

Test addytywności Tukey'a

Umożliwia przetestowanie założenia o braku multiplikatywnej interakcji między pozycjami.

Współczynnik korelacji wewnątrzklasowej

Umożliwia przeprowadzanie pomiarów spójności lub zgodności wartości wewnątrz obserwacji.

Model

Należy wybrać model dla obliczenia współczynnika korelacji wewnątrzklasowej. Dostępne modele to: 2-czynnikowy mieszany, 2-czynnikowy losowy, 1-czynnikowy losowy. Należy wybrać model **2-czynnikowy mieszany**, jeśli efekty obiektowe są losowe, a efekty pozycyjne stałe, **2-czynnikowy losowy**, jeśli efekty obiektowe i pozycyjne są losowe, lub **1-czynnikowy losowy**, jeśli obiektowe są losowe.

Typ

Należy wybrać typ wskaźnika. Dostępne typy to Spójność i Bezwzględna zgodność.

Przedział ufności (%)

Należy określić poziom przedziału ufności. Wartością domyślną jest 95%.

Wartość testowa

Należy określić hipotetyczną wartość współczynnika dla testu hipotezy. Jest to wartość, do której porównywana jest wartość obserwowana. Wartość domyślna wynosi 0.

Dodatkowe właściwości komendy RELIABILITY

Język składni komend umożliwia również:

- Odczytywanie i analizę macierzy korelacji.
- Zapisywanie macierzy korelacji do późniejszej analizy.
- Określanie podziałów na części inne niż połowy dla metody potówkowej.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Ważona Kappa

Ważona kappa Cohena jest powszechnie stosowana w klasyfikacji krzyżowej jako miara zgodności pomiędzy ocenającymi. Jest ona właściwym wskaźnikiem zgodności, jeśli oceny są nominalnymi zmiennymi ilościowymi, bez określonej kolejności. Metodę ważonej kappy Cohena opracowano po to, aby uwzględnić fakt, że niektóre przypisania w tabeli kontyngencji mogą mieć większą wagę niż pozostałe. Statystyka ta bazuje na wstępnie zdefiniowanych wagach komórek, odzwierciedlających zarówno zgodność, jak i brak zgodności.

Procedura Ważona Kappa udostępnia opcje estymacji ważonej kappy Cohena, będącej istotnym uogólnieniem statystyki kappa, która mierzy zgodność dwóch porządkowych obiektów o identycznych kategoriach.

Uwaga: Procedura Ważona Kappa zastępuje funkcje realizowane wcześniej przez rozszerzenie STATS WEIGHTED KAPPA.spe.

Przykład

Istnieją sytuacje, w których różnice między ocenającymi nie powinny być traktowane jako równie ważne. Przykładem może być branża medyczna, w której wiele osób gromadzi dane z badań naukowych lub klinicznych. W takich przypadkach rzetelność danych może zostać zakwestionowana ze względu na różnice między osobami, które je gromadzą.

Statystyki

Ważona kappa Cohena, skala liniowa, skala kwadratowa, asymptotyczny przedział ufności.

Wymagania dotyczące danych dla procedury Ważona Kappa

Dane

Do oszacowania ważonej kappy Cohena wymagana jest tabela dwuwymiarowa oparta na aktywnym zbiorze danych.

Zmienne ocenijące muszą być tego samego typu (wszystkie łańcuchowe lub wszystkie liczbowe).

Oszacowanie ważonej kappy Cohena ma sens tylko wtedy, gdy kategorie dwóch zmiennych ocenających, odzwierciedlone przez wiersz i kolumnę tabeli, są odpowiednio uporządkowane (w przypadku pary zmiennych liczbowych stosowana jest kolejność liczbową; w przypadku pary zmiennych łańcuchowych stosowana jest kolejność alfabetyczna).

Założenia

W przypadku wybrania pary zmiennych mieszanych ważona kappa Cohena nie jest szacowana.

Zakłada się, że zmienne ocenijące mają ten sam zestaw kategorii.

Aby uzyskać analizę według procedury Ważona kappa

1. Wybierz z menu następującą opcję:

Analizuj > Skala > Ważona Kappa...

2. Wybierz co najmniej dwie zmienne łańcuchowe lub liczbowe jako **ocenijące parami**.

Uwaga: Wszystkie wybrane zmienne muszą być łańcuchowe albo wszystkie muszą być liczbowe.

3. Opcjonalnie włącz ustawienie **Określ zmienne ocenijące dla wierszy i kolumn**, aby wybrać sposób wyświetlania ocenających: parami lub wiersze/kolumny.

- Gdy ta opcja jest włączona, zmienne ocenijące parami są ukryte, a wyświetlane są zmienne ocenijące wiersze/kolumny. Interfejs użytkownika zmienia się, tak by dostępne były pola **Zmienne ocenijące wiersze** i **Zmienne ocenijące kolumny** (zamiast pola **Zmienne ocenijące parami**).
- Gdy ta opcja jest wyłączona, zmienne ocenijące wiersze/kolumny są ukryte, a wyświetlane są zmienne ocenijące parami (ustawienie domyślne).

Gdy włączona jest opcja **Określ zmienne oceniające dla wierszy i kolumn**, należy określić co najmniej po jednej zmiennej w polu **Zmienne oceniające wiersze** i **Zmienne oceniające kolumny**.

Uwaga: Jeśli zarówno pole **Zmienne oceniające wiersze**, jak i **Zmienne oceniające kolumny** zawiera tylko jedną zmienną, wybrane zmienne nie mogą być takie same w obu polach.

4. Opcjonalnie kliknij opcję **Kryteria**, aby określić ustawienia dotyczące skali wag i braków danych, lub opcję **Drukuj**, aby określić ustawienia formatu wyświetlania i tabeli krzyżowej.

Ważona Kappa: kryteria

Okno dialogowe Kryteria udostępnia opcje służące do określania oszacowania ważonej kappa Cohena.

Skala ważenia

Umożliwia wybranie liniowej albo kwadratowej skali ważenia zgodności. Zastosowanie waży liniowej (zwanej również Cicchetti-Allison) jest ustawieniem domyślnym.

Braki danych

Udostępnia opcje usuwania parami obserwacji z brakami danych i uznawania ważności braków danych zdefiniowanych przez użytkownika.

W łańcuchowych zmiennych oceniających rozróżniana jest wielkość liter

Gdy ta opcja jest wybrana, w zmiennych łańcuchowych jest rozróżniania wielkość liter.

Asymptotyczny przedział ufności (%)

To opcjonalne ustawienie określa poziom ufności dla estymacji asymptotycznych przedziałów ufności. Musi to być jedna wartość typu double z zakresu od 0 do 100 (95 jest ustawieniem domyślnym).

Ważona Kappa: drukuj

Okno dialogowe Drukuj udostępnia opcje służące do sterowania tabelami krzyżowymi.

Prezentacja i format

Udostępnia opcje służące do sterowania wyglądem i formatem tabeli krzyżowej.

Kategorie zmiennych oceniających są prezentowane w porządku rosnącym

Gdy ta opcja jest wybrana, kategorie ocen w tabelach krzyżowych są wyświetlane w porządku rosnącym. Gdy to ustawienie nie jest wybrane, kategorie ocen w tabelach krzyżowych są wyświetlane w porządku malejącym. Ustawienie jest domyślnie włączone.

Etykiety kategorii łańcuchowych wyświetlane są wielkimi literami

Jeśli ta opcja jest wybrana, tabele krzyżowe są wyświetlane wielkimi literami. Gdy to ustawienie nie jest wybrane, tabele krzyżowe są wyświetlane małymi literami. Ustawienie jest domyślnie włączone.

Tabela krzyżowa

Udostępnia opcje służące do określania zmiennych oceniających używanych w tabelach krzyżowych. Domyślnie ustawienia tabel krzyżowych nie są włączone, co powoduje pominięcie tabel krzyżowych zmiennych oceniających.

Prezentacja tabeli krzyżowej zmiennych oceniających

Wybranie tej opcji powoduje włączenie tabeli krzyżowej zmiennych oceniających — wszystkich lub określonych przez użytkownika.

Uwzględnij wszystkie zmienne oceniające

Gdy ta opcja jest wybrana, tabele krzyżowe są prezentowane dla wszystkich zdefiniowanych par zmiennych oceniających.

Uwzględnij zmienne oceniające określone przez użytkownika

Gdy ta opcja jest wybrana, należy użyć pól **Dostępne zmienne**, **Zmienne oceniające wiersze** i **Zmienne oceniające kolumny**, aby wybrać zmienne oceniające uwzględnione w tabelach krzyżowych.

Skalowanie wielowymiarowe

Skalowanie wielowymiarowe służy do znajdowania struktury w zbiorze miar odległości między poszczególnymi obiektami lub obserwacjami. Jest to możliwe dzięki przypisywaniu obserwacji do poszczególnych miejsc w przestrzeni pojęciowej (zwykle dwu- lub trójwymiarowej) w taki sposób, że odległości między punktami w przestrzeni możliwie blisko odpowiadają danym miarom nie-podobieństwa. W wielu przypadkach wymiary tej przestrzeni pojęciowej mogą być interpretowane i wykorzystywane do lepszego zrozumienia danych.

Dokonanie pomiarów zmiennych w sposób obiektywny pozwala na korzystanie ze skalowania wielowymiarowego jako metody redukcji danych (do wyliczania odległości między wieloma zmiennymi można, w razie potrzeby, posłużyć się procedurą skalowania wielowymiarowego). Skalowanie wielowymiarowe można również stosować do subiektywnych ocen niepodobieństwa między obiektami lub pojęciami. Procedura skalowania wielowymiarowego umożliwia ponadto przetwarzanie danych dotyczących niepodobieństw, pochodzących z różnych źródeł, takich jak wielu różnych ankietowanych czy respondentów.

Przykład. W jaki sposób ludzie postrzegają relacje między różnego typu samochodami? Jeśli dane pochodzące od respondentów wskazują na podobieństwa między różnymi markami i modelami aut, to do identyfikacji wymiarów opisujących percepcję konsumentów wykorzystać można skalowanie wielowymiarowe. Z badania wynikać mogłoby, na przykład że cena i rozmiar auta tworzą dwuwymiarową przestrzeń, wyjaśniającą podobieństwa zgłaszane przez respondentów.

Statystyki. Dla każdego modelu danych: macierz danych, optymalnie skalowana macierz danych, obciążenie s (Younga), obciążenie (Kruskala), R kwadrat, współrzędne dla bodźców, średnią wartość obciążenia oraz R kwadrat dla każdego bodźca (modele RMDS). W przypadku modeli różnic indywidualnych (INDSCAL): wagi obiektów i wskaźnik odmienności dla każdego podmiotu. Dla każdej macierzy w replikowanym modelu skalowania wielowymiarowego: obciążenie i R kwadrat dla każdego bodźca. Wykresy: współrzędne dla bodźców (dwu- lub trójwymiarowe), wykres rozproszenia różnic-odległości.

Wymagania dotyczące danych przy skalowaniu wielowymiarowym

Dane. Jeśli dane są danymi dotyczącymi niepodobieństw, wszelkie niepodobieństwa powinny mieć postać ilościową i powinny być mierzone za pomocą tej samej metryki. Jeśli dane zawierają wiele zmiennych, zmienne mogą być danymi ilościowymi, binarnymi lub liczebnościami. Skalowanie zmiennych jest istotne, ponieważ różnice w skalowaniu mogą mieć wpływ na rozwiązanie. Jeżeli występują bardzo znaczne różnice w skalowaniu zmiennych (na przykład jedna zmienna jest mierzona w dolarach, a druga w latach), należy rozważyć możliwość ich standaryzacji (można to zrobić automatycznie za pomocą procedury Skalowanie wielowymiarowe).

Założenia. Procedura skalowania wielowymiarowego jest względnie wolna od założeń dystrybucyjnych. Upewnij się, by w oknie dialogowym Skalowanie wielowymiarowe: Opcje zaznaczyć odpowiedni Poziom pomiaru (porządkowy, interwałowy lub ilorazowy) tak, aby wyniki zostały obliczone poprawnie.

Procedury pokrewne. Jeśli celem jest redukcja danych, to alternatywną metodą, której można użyć w tym celu jest analiza czynnikowa, zwłaszcza kiedy czynności dotyczą zmiennych ilościowych. W razie potrzeby identyfikacji grup podobnych obserwacji, należy rozważyć uzupełnienie analizy skalowania wielowymiarowego o hierarchiczną analizę skupień lub analizę skupień metodą k -średnich.

Wykonywanie analizy skalowania wielowymiarowego

1. Wybierz z menu następującą opcję:

Analiza > Skalowanie > Skalowanie wielowymiarowe...

2. Wybierz co najmniej cztery zmienne numeryczne do analizy.
3. W grupie Odległości wybierz opcję **Dane to odległości** albo **Utwórz odległości na podstawie danych**.
4. Jeśli wybierzesz opcję **Utwórz odległości na podstawie danych**, można również dodać również wybrać zmienną grupującą dla oddzielnych macierzy. Zmienna grupująca może być numeryczna lub łańcuchowa.

Opcjonalnie można również wykonać następujące czynności:

- Określić kształt macierzy odległości, gdy dane są odległościami.
- Określić miarę odległości, która będzie używana do tworzenia odległości na podstawie danych.

Skalowanie wielowymiarowe: Kształt danych

Jeśli aktywny zbiór danych składa się z odległości między obiektami należącymi do zbioru lub odległości między dwoma zbiorami obiektów, to w celu uzyskania poprawnych wyników konieczne jest określenie kształtu macierzy danych.

Uwaga: nie można wybrać opcji **Macierz kwadratowa symetryczna**, jeśli w oknie dialogowym Model określone są warunki porównań wiersza.

Skalowanie wielowymiarowe: Utwórz miarę na podstawie danych

W skalowaniu wielowymiarowym, wynik skalowania tworzony jest na podstawie niepodobieństw. W przypadku korzystania z danych o wielu zmiennych (wartości mierzonych zmiennych), w celu uzyskania wyniku skalowania wielowymiarowego konieczne jest utworzenie danych niepodobieństwa. Można określić szczegóły tworzenia miar niepodobieństw na podstawie danych.

Miara. Pozwala na określanie miary niepodobieństwa dla potrzeb analizy. W obszarze Miara dla danych wybierz jedną z alternatyw, odpowiadającą typowi analizowanych danych, a następnie z odpowiadającej jej listy rozwijanej wybierz stosowną miarę. Dostępne alternatywy to:

- **Interwałowe.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Odległość Czebyszewa, Odległość miejska, Odległość Minkowskiego lub Odległość użytkownika.
- **Liczebności.** Odległość chi-kwadrat lub Odległość phi-kwadrat.
- **Binarne.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Różnica wielkości, Różnica wzoru, Wariancja lub Miara Lance'a i Williamsa.

Utwórz macierz odległości. Pozwala na wybranie jednostki analizy. Dostępne alternatywy to: Pomiędzy zmiennymi i Między obserwacjami.

Przekształcanie wartości. W niektórych przypadkach, na przykład, kiedy zmienne mierzone są przy użyciu różnych skali, wartości można wystandaryzować przed rozpoczęciem wyliczania odległości pomiędzy nimi (nie dotyczy danych binarnych). Z listy rozwijanej Standaryzacja wybierz metodę standaryzacji. Jeżeli nie jest wymagane przeprowadzenie standaryzacji, wybierz opcję **Brak**.

Skalowanie wielowymiarowe: Model

Poprawna estymacja modelu skalowania wielowymiarowego zależy zarówno od danych, jak i od samego modelu.

Poziom pomiaru. Opcja ta pozwala określić poziom pomiaru dla danych. Dostępne alternatywy to: Porządkowy, Interwałowy lub Ilorazowy. Jeśli zmienne są porządkowe, zaznaczenie pola wyboru **Rozwiąż związane obserwacje** sprawia, że dane traktowane są jako zmienne ilościowe, w związku z czym wiązania (czyli równe wartości dla różnych obserwacji) rozwiązywane są optymalnie.

Warunki porównań. Pozwala na określenie, które porównania są znaczące. Dostępne opcje to: Macierz, Wiersz lub Bezwarunkowo.

Wymiary. Pozwala na określenie wymiarowości rozwiązania (lub rozwiązań) skalowania. Dla każdej liczby z zakresu obliczane jest jedno rozwiązanie. Podaj liczby całkowite z zakresu od 1 do 6. Wybranie liczby 1 jako minimum jest dozwolone jedynie po zaznaczeniu opcji **Odległość euklidesowa** jako modelu skalowania. Dla pojedynczego rozwiązania podaj tę samą liczbę jako minimum i maksimum.

Model skalowania. Pozwala na określanie założeń, według których jest realizowane skalowanie. Dostępne alternatywy to: Odległość euklidesowa lub Odległość euklidesowa różnic indywidualnych (określana również mianem INDSCAL). Dla modelu odległości euklidesowej różnic indywidualnych można zaznaczyć pole wyboru **Pozwól na ujemne wagi obiektów**, o ile dane tego wymagają.

Skalowanie wielowymiarowe: Opcje

Można określić opcje dla analizy skalowania wielowymiarowego:

Pokaż. Pozwala na wybranie różnych typów wyników. Dostępne opcje to: Wykresy grupowe, Wykresy indywidualne, Macierz danych i Podsumowanie dla modelu i opcji.

Kryteria. Pozwala na określenie kryterium kończącego iterację. Aby zmienić ustawienia domyślne, należy wprowadzić wartości w pola **Zbieżność s-stress**, **Minimalna wartość s-stress** i **Maksymalna liczba iteracji**.

Traktuj odległości mniejsze od n jako braki. Odległości mniejsze od podanej wartości są wyłączone z analizy.

Dodatkowe właściwości komendy ALSCAL

Język składni komend umożliwia również:

- Korzystanie z trzech dodatkowych typów modeli, znanych z publikacji na temat skalowania wielowymiarowego jako ASCAL, AINDS i GEMSCAL.
- Wykonywanie transformacji wielomianowych na danych typu interwałowego oraz ilorazowego.
- Analizowanie podobieństw (zamiast odległości) z danymi porządkowymi.
- Analizowanie danych nominalnych.
- Zapisywanie różnych macierzy współrzędnych i wag w plikach i odczytywanie ich dla potrzeb późniejszej analizy.
- Ograniczanie rozwijania wielowymiarowego.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

Statystyki ilorazowe

Procedura Statystyki ilorazowe udostępnia pełny zakres statystyk podsumowujących do opisu proporcji (ilorazu) dwóch zmiennych ilościowych.

Wyniki można sortować według wartości zmiennej grupującej, w porządku rosnącym lub malejącym. Istnieje możliwość ukrycia raportu ze statystyk ilorazowych w taki sposób, że nie będzie on pokazywany w Edytorze raportów, a wyniki zostaną zapisane w pliku zewnętrznym.

Przykład

Czy stosunek oszacowanej wartości domów do cen ich sprzedaży jest podobny w każdym z pięciu województw? Na podstawie wyników można stwierdzić, że rozkład proporcji znacznie różni się dla poszczególnych województw.

Statystyki

Mediana, średnia, średnia ważona, przedziały ufności, współczynnik dyspersji (COD), współczynnik zmienności (COV), współczynnik zmienności środkowej zmienności, średni wyśrodkowany współczynnik zmienności, stosunek ceny związanej z ceną (PRB), różnica w stosunku do ceny (PRD), odchylenie standardowe, średnie odchylenie bezwzględne (AAD), zakres, wartości minimalne i maksymalne, wskaźnik koncentracji obliczony dla zakresu określonego przez użytkownika lub procent w stosunku do wskaźnika mediany.

Zagadnienia dotyczące danych

Dane

Użycie kodów numerycznych lub łańcuchów do kodowania zmiennych grupujących (pomiaru na poziomie nominalnym lub porządkowym).

Założenia

Zmienne definiujące licznik i mianownik proporcji powinny być zmiennymi ilościowymi przyjmującymi wartości dodatnie.

Uzyskiwanie statystyk ilorazowych

1. Wybierz z menu następującą opcję:

Analiza > Statystyki opisowe > Statystyki ilorazowe...

2. Wybierz zmienną definiującą licznik.

3. Wybierz zmienną definiującą mianownik.

Opcjonalnie:

- Wybierz zmienną grupującą i określ porządek grup w wynikach.
- Określ, czy wyniki mają być wyświetlane w Edytorze raportów.
- Wybierz, aby dodać słowo kluczowe **N** do podkomendy **PRINT**. To ustawienie powoduje dodanie wielkości próby do tabeli statystyk w wynikach.
- Określ, czy wyniki mają być zapisane w pliku zewnętrznym do wykorzystania w przyszłości oraz podać nazwę pliku, w którym mają zostać zapisane.

Statystyki ilorazowe: Statystyki

Tendencja centralna

Miary tendencji centralnej są statystykami opisującymi rozkład proporcji.

Mediana

Wartość, przy której liczba proporcji mniejszych i większych od niej jest równa.

Średnia

Wynik sumy proporcji podzielony przez ich całkowitą liczbę.

Średnia ważona

Wynik ilorazu średniej licznika i średniej mianownika. Jest również średnią proporcji ważoną mianownikiem.

Przedziały ufności dla PRB i miary tendencji centralnej

Określa przedziały ufności dla obciążenia związanego z ceną (PR.), średniej, mediany i średniej ważonej (w zależności od wyboru użytkownika). Jako poziom ufności należy podać wartość większą lub równą 0 i mniejszą od 100. Ustawieniem domyślnym jest 95%.

Dyspersja

Są to statystyki mierzące zmienność lub rozrzut obserwowanych wartości.

AAD

Przeciętne bezwzględne odchylenie stanowi sumę bezwzględnych odchyłeń proporcji od mediany, podzieloną przez całkowitą liczbę proporcji.

Współczynnik dyspersji

Współczynnik dyspersji jest średnim bezwzględnym odchyleniem wyrażonym w postaci wartości procentowej mediany.

W.Z.

Współczynnik zmienności. Centrowany medianą współczynnik zmienności stanowi pierwiastek ze średniej kwadratów odchyłeń od mediany, wyrażony w postaci wartości procentowej. Centrowany średnią współczynnik zmienności jest odchyleniem standardowym wyrażonym w postaci wartości procentowej średniej.

PR.

Obciążenie związane z ceną (PRB). Określa, czy stosunki wyceny do ceny są ogólnie wyższe lub niższe dla wysoko wycenianych właściwości. Poprzez regresję wyznacza różnice procentowe proporcji ocen z mediany ilorazu w oparciu o logarytmy o podstawie 2 miar zastępczych obliczonych jako średnie cen sprzedaży i ilorazy oszacowanych wartości oraz mediany. Podaje procentową zmianę proporcji ocen przy zmianie wartości o 100 procent.

Wskaźnik regresywności

Wskaźnik regresywności stanowi iloraz średniej i średniej ważonej.

Odchylenie standardowe

Odchylenie standardowe jest to pierwiastek kwadratowy z sumy kwadratów odchyłeń proporcji od średniej, podzielonej przez całkowitą liczbę obserwacji minus jeden.

Wykładnik potęgi od

Przedział jest wynikiem odjęcia różnicy proporcji maksymalnej i minimalnej.

Minimum

minimum to najmniejszy zakres.

Maksimum

Maksimum to największy zakres.

Wskaźnik koncentracji

Wskaźnik koncentracji mierzy procent proporcji mieszczących się w określonym przedziale. Można go wyliczyć na dwa sposoby:

Między proporcjami

W tym przypadku przedział jest definiowany przez określenie wartości maksymalnej i minimalnej. Należy wprowadzić wartości proporcji maksymalnej i minimalnej, po czym kliknąć przycisk **Dodaj**, aby otrzymać przedział.

W obrębie procentu mediany

W tym przypadku przedział jest definiowany przez określenie procenta mediany. Należy wprowadzić wartość z przedziału od 0 do 100, po czym kliknąć przycisk **Dodaj**. Dolna wartość przedziału wynosi $(1-0,01 \times \text{wartość}) \times \text{mediana}$, a górna $(1+0,01 \times \text{wartość}) \times \text{mediana}$.

Statystyki ilorazowe

Procedura Statystyki ilorazowe udostępnia pełny zakres statystyk podsumowujących do opisu proporcji (ilorazu) dwóch zmiennych ilościowych.

Wyniki można sortować według wartości zmiennej grupującej, w porządku rosnącym lub malejącym. Istnieje możliwość ukrycia raportu ze statystyk ilorazowych w taki sposób, że nie będzie on pokazywany w Edytorze raportów, a wyniki zostaną zapisane w pliku zewnętrznym.

Przykład

Czy stosunek oszacowanej wartości domów do cen ich sprzedaży jest podobny w każdym z pięciu województw? Na podstawie wyników można stwierdzić, że rozkład proporcji znacznie różni się dla poszczególnych województw.

Statystyki

Mediana, średnia, średnia ważona, przedziały ufności, współczynnik dyspersji (COD), współczynnik zmienności (COV), współczynnik zmienności środkowej zmienności, średni wyśrodkowany współczynnik zmienności, stosunek ceny związany z ceną (PRB), różnica w stosunku do ceny (PRD), odchylenie standardowe, średnie odchylenie bezwzględne (AAD), zakres, wartości minimalne i maksymalne, wskaźnik koncentracji obliczony dla zakresu określonego przez użytkownika lub procent w stosunku do wskaźnika mediany.

Zagadnienia dotyczące danych

Dane

Użycie kodów numerycznych lub łańcuchów do kodowania zmiennych grupujących (pomiaru na poziomie nominalnym lub porządkowym).

Założenia

Zmienne definiujące licznik i mianownik proporcji powinny być zmiennymi ilościowymi przyjmującymi wartości dodatnie.

Uzyskiwanie statystyk ilorazowych

1. Wybierz z menu następującą opcję:

Analiza > Statystyki opisowe > Statystyki ilorazowe...

2. Wybierz zmienną definiującą licznik.
3. Wybierz zmienną definiującą mianownik.

Opcjonalnie:

- Wybierz zmienną grupującą i określ porządek grup w wynikach.
- Określ, czy wyniki mają być wyświetlane w Edytorze raportów.
- Wybierz, aby dodać słowo kluczowe **N** do podkomendy **PRINT**. To ustawienie powoduje dodanie wielkości próby do tabeli statystyk w wynikach.
- Określ, czy wyniki mają być zapisane w pliku zewnętrznym do wykorzystania w przyszłości oraz podać nazwę pliku, w którym mają zostać zapisane.

Statystyki ilorazowe: Statystyki

Tendencja centralna

Miary tendencji centralnej są statystykami opisującymi rozkład proporcji.

Mediana

Wartość, przy której liczba proporcji mniejszych i większych od niej jest równa.

Średnia

Wynik sumy proporcji podzielony przez ich całkowitą liczbę.

Średnia ważona

Wynik ilorazu średniej licznika i średniej mianownika. Jest również średnią proporcji ważoną mianownikiem.

Przedziały ufności dla PRB i miary tendencji centralnej

Określa przedziały ufności dla obciążenia związanego z ceną (PR.), średniej, mediany i średniej ważonej (w zależności od wyboru użytkownika). Jako poziom ufności należy podać wartość większą lub równą 0 i mniejszą od 100. Ustawieniem domyślnym jest 95%.

Dyspersja

Są to statystyki mierzące zmienność lub rozrzut obserwowanych wartości.

AAD

Przeciętne bezwzględne odchylenie stanowi sumę bezwzględnych odchyłeń proporcji od mediany, podzieloną przez całkowitą liczbę proporcji.

Współczynnik dyspersji

Współczynnik dyspersji jest średnim bezwzględnym odchyleniem wyrażonym w postaci wartości procentowej mediany.

W.Z.

Współczynnik zmienności. Centrowany medianą współczynnik zmienności stanowi pierwiastek ze średniej kwadratów odchyłeń od mediany, wyrażony w postaci wartości procentowej. Centrowany średnią współczynnik zmienności jest odchyleniem standardowym wyrażonym w postaci wartości procentowej średniej.

PR.

Obciążenie związane z ceną (PRB). Określa, czy stosunki wyceny do ceny są ogólnie wyższe lub niższe dla wysoko wycenianych właściwości. Poprzez regresję wyznacza różnice procentowe proporcji ocen z mediany ilorazu w oparciu o logarytmy o podstawie 2 miar zastępczych obliczonych jako średnie cen sprzedaży i ilorazy oszacowanych wartości oraz mediany. Podaje procentową zmianę proporcji ocen przy zmianie wartości o 100 procent.

Wskaźnik regresywności

Wskaźnik regresywności stanowi iloraz średniej i średniej ważonej.

Odchylenie standardowe

Odchylenie standardowe jest to pierwiastek kwadratowy z sumy kwadratów odchyłeń proporcji od średniej, podzielonej przez całkowitą liczbę obserwacji minus jeden.

Wykładnik potęgi od

Przedział jest wynikiem odjęcia różnicy proporcji maksymalnej i minimalnej.

Minimum

minimum to najmniejszy zakres.

Maksimum

Maksimum to największy zakres.

Wskaźnik koncentracji

Wskaźnik koncentracji mierzy procent proporcji mieszczących się w określonym przedziale. Można go wyliczyć na dwa sposoby:

Między proporcjami

W tym przypadku przedział jest definiowany przez określenie wartości maksymalnej i minimalnej. Należy wprowadzić wartości proporcji maksymalnej i minimalnej, po czym kliknąć przycisk **Dodaj**, aby otrzymać przedział.

W obrębie procentu mediany

W tym przypadku przedział jest definiowany przez określenie procenta mediany. Należy wprowadzić wartość z przedziału od 0 do 100, po czym kliknąć przycisk **Dodaj**. Dolna wartość przedziału wynosi $(1-0,01 \times \text{wartość}) \times \text{mediana}$, a górna $(1+0,01 \times \text{wartość}) \times \text{mediana}$.

Wykres P-P (prawdopodobieństwo-prawdopodobieństwo)

Wykresy P-P generują wykresy prawdopodobieństwa dla najmniej jednej sekwencji lub zmiennych szeregu czasowego. Zmienne mogą być standaryzowane, różnicowane i przekształcane przed ich narysowaniem.

Dostępne są następujące rozkłady testowe: beta, chi-kwadrat, wykładniczy, gamma, półnormalny, Laplace'a, logistyczny, lognormalny, normalny, pareto, t-Studenta, Weibulla, jednostajny. Zależnie od wybranego rozkładu można określić ilość stopni swobody oraz inne parametry.

- Wykresy prawdopodobieństwa otrzymuje się dla wartości transformowanych. Opcje przekształcenia obejmują: logarytm naturalny, standaryzację wartości, różnicowanie oraz różnice sezonowe.
- Można określić metodę obliczania rozkładów oczekiwanych oraz rozwiązywania „wiązań” (czyli wielu obserwacji o tych samych wartościach).

Rozkład testowy

Określ typ rozkładu danych. Na liście rozwijanej dostępne są następujące opcje:

Wersja beta

Rozkład Beta. Parametry kształt1 i kształt2, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby. Wszystkie obserwacje muszą mieścić się w zakresie od 0 do 1, włącznie.

Chi-kwadrat

Rozkład chi-kwadrat. Należy podać stopnie swobody (df). Obserwacje ujemne nie są dozwolone.

Wykładniczy

Rozkład wykładniczy. Parametr skali a musi być dodatni. Jeśli parametr nie jest określony, ROZKŁAD szacuje go na podstawie średniej z próby. Obserwacje ujemne nie są dozwolone.

Gamma

Rozkład gamma. Parametry kształtu i skali, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby. Obserwacje ujemne nie są dozwolone.

Półnormalny

Rozkład półnormalny. Przyjmuje się, że dane nie mają lokalizacji lub są scentralizowane. (Parametr lokalizacji=0.) Można określić parametr skali a lub oszacować go przy użyciu opcji ROZKŁAD metodą największej wiarygodności.

Laplace'a

Rozkład Laplace'a lub podwójny rozkład wykładniczy. Opcja LAPLACE wykorzystuje parametr lokalizacji i skali (a i b). Parametr skali (b) musi być dodatni. Jeśli parametry nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Logistyczna

Rozkład logistyczny. Opcja LOGISTIC wykorzystuje parametr lokalizacji i skali (a i b). Parametr skali (b) musi być dodatni. Jeśli parametry nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Lognormalny

Rozkład lognormalny. Parametry skali i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej i odchylenia standardowego logarytmu naturalnego danych próby. Obserwacje ujemne nie są dozwolone.

Normalny

Rozkład normalny. Parametr lokalizacji a może być dowolną wartością liczbową, a parametr skali b musi być liczbą dodatnią. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Pareto

Rozkład Pareto. Parametry progę i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD zakłada, że parametr a jest równy obserwacji minimalnej, a parametr b jest szacowany metodą największej wiarygodności. Obserwacje ujemne nie są dozwolone.

T Studenta

Rozkład t Studenta. Należy podać stopnie swobody (df).

Jednostajny

Rozkład jednostajny. Funkcja UNIFORM przyjmuje parametr minimum i maksimum (a i b). Parametr a musi być równy lub większy niż b . Jeśli parametry nie zostaną określone, DYSTRYBUCJA przyjmie je z przykładowych danych.

Weibulla

Rozkład Weibulla. Parametry skali i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je metodą najmniejszych kwadratów. Obserwacje ujemne nie są dozwolone.

Parametry rozkładu

Udostępnia strategię rozkładu i opcje parametrów.

Oszacuj na podstawie danych

Po wybraniu tego ustawienia parametry rozkładu są szacowane na podstawie danych i wybranego typu rozkładu.

Podaj

Po wybraniu tego ustawienia można oszacować parametry rozkładu dla wybranego typu rozkładu.

Uwaga: Dostępne parametry będą się różnić w zależności od wybranego typu rozkładu.

Przekształcenia

Dostępne opcje umożliwiają wybór ustawienia przekształcenia i cykliczności.

Transformacja logarytmiczna naturalna

Przekształca dane przy użyciu logarytmu naturalnego (podstawa e) w celu usunięcia różnych amplitud.

Standaryzuj wartości

Przekształca sekwencję lub szereg czasowy do próby o średniej 0 i odchyleniu standardowym 1.

Różnice

Określa rząd różnic używany przed wygenerowaniem wykresu do przekształcenia zmiennej niestacjonarnej na zmienną stacjonarną ze stałą średnią i wariancją. Wprowadź odpowiednią wartość w polu.

Różnica sezonowa

Jeśli zmienna wykazuje wzór sezonowy lub okresowy, można użyć tego ustawienia w celu zróżnicowania sezonowego zmiennej przed utworzeniem wykresów.

Uwaga: Ustawienie to jest aktywne tylko wtedy, gdy jako jedną ze zmiennych ilościowych wybrano sekwencję lub szereg czasowy z określoną okresowością.

Formuła estymacji proporcji

Dostępne opcje określają wzór, który jest używany do szacowania proporcji.

Blom

Tworzy nową zmienną rangującą na podstawie oszacowań proporcji obliczonych ze wzoru $(r-3/8) / (w+1/4)$, gdzie w jest sumą wag obserwacji, a r jest rangą.

Rankit

Wykorzystuje wzór $(r-1/2) / w$, gdzie w jest liczbą obserwacji a r jest rangą, przyjmującą wartości od 1 do w .

Tukey

Wykorzystuje wzór $(r-1/3) / (w+1/3)$, gdzie r jest rangą, a w jest sumą wag obserwacji.

Van der Waerden

Transformacja Van der Waerdena określona przez wzór $r/(w+1)$, gdzie w jest sumą wag obserwacji, a r jest rangą od 1 do w .

Rangi przypisane do wiązań

Dostępne opcje umożliwiają wybór sposobu postępowania z wartościami wiązanymi. Poniższa tabela wskazuje sposoby przypisywania rang do wiązanych wartości w różnych metodach.

Tabela 3. Metody i wyniki rangowania

Wartość	Średnia	Niska	Wysoka	Podziel wiązania arbitralnie
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Otrzymywanie wykresów P-P (prawdopodobieństwo-prawdopodobieństwo)

1. Wybierz z menu następującą opcję:

Analizuj > Statystyki opisowe > Wykresy P-P...

2. Zaznacz co najmniej jedną zmienną numeryczną i przenieś ją w pole Zmienne.

3. Wybierz rozkład testowy.

Opcjonalnie można zaznaczyć opcje przekształcenia w celu otrzymania wykresów prawdopodobieństwa wartości transformowanych, a także określić metodę obliczania rozkładów oczekiwanych.

Wykres K-K (kwantyl-kwantyl)

Procedura Wykresy K-K generuje wykresy prawdopodobieństwa wartości transformowanych. Dostępne są następujące rozkłady testowe: beta, chi-kwadrat, wykładniczy, gamma, półnormalny, Laplace'a, logistyczny, lognormalny, normalny, pareto, t-Studenta, Weibulla, jednostajny. Zależnie od wybranego rozkładu można określić ilość stopni swobody oraz inne parametry.

- Wykresy prawdopodobieństwa otrzymuje się dla wartości transformowanych. Opcje przekształcenia obejmują: logarytm naturalny, standaryzację wartości, różnicowanie oraz różnice sezonowe.
- Można określić metodę obliczania rozkładów oczekiwanych oraz rozwiązywania „wiązań” (czyli wielu obserwacji o tych samych wartościach).

Rozkład testowy

Określ typ rozkładu danych. Na liście rozwijanej dostępne są następujące opcje:

Wersja beta

Rozkład Beta. Parametry kształt1 i kształt2, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby. Wszystkie obserwacje muszą mieścić się w zakresie od 0 do 1, włącznie.

Chi-kwadrat

Rozkład chi-kwadrat. Należy podać stopnie swobody (df). Obserwacje ujemne nie są dozwolone.

Wykładniczy

Rozkład wykładniczy. Parametr skali a musi być dodatni. Jeśli parametr nie jest określony, ROZKŁAD szacuje go na podstawie średniej z próby. Obserwacje ujemne nie są dozwolone.

Gamma

Rozkład gamma. Parametry kształtu i skali, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby. Obserwacje ujemne nie są dozwolone.

Pótnormalny

Rozkład pótnormalny. Przyjmuje się, że dane nie mają lokalizacji lub są scentralizowane. (Parametr lokalizacji=0.) Można określić parametr skali a lub oszacować go przy użyciu opcji ROZKŁAD metodą największej wiarygodności.

Laplace'a

Rozkład Laplace'a lub podwójny rozkład wykładniczy. Opcja LAPLACE wykorzystuje parametr lokalizacji i skali (a i b). Parametr skali (b) musi być dodatni. Jeśli parametry nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Logistyczna

Rozkład logistyczny. Opcja LOGISTIC wykorzystuje parametr lokalizacji i skali (a i b). Parametr skali (b) musi być dodatni. Jeśli parametry nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Lognormalny

Rozkład lognormalny. Parametry skali i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej i odchylenia standardowego logarytmu naturalnego danych próby. Obserwacje ujemne nie są dozwolone.

Normalny

Rozkład normalny. Parametr lokalizacji a może być dowolną wartością liczbową, a parametr skali b musi być liczbą dodatnią. Jeśli nie są określone, ROZKŁAD szacuje je na podstawie średniej z próby i odchylenia standardowego próby.

Pareto

Rozkład Pareto. Parametry progu i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD zakłada, że parametr a jest równy obserwacji minimalnej, a parametr b jest szacowany metodą największej wiarygodności. Obserwacje ujemne nie są dozwolone.

T Studenta

Rozkład t Studenta. Należy podać stopnie swobody (df).

Jednostajny

Rozkład jednostajny. Funkcja UNIFORM przyjmuje parametr minimum i maksimum (a i b). Parametr a musi być równy lub większy niż b . Jeśli parametry nie zostaną określone, DYSTRYBUCJA przyjmie je z przykładowych danych.

Weibulla

Rozkład Weibulla. Parametry skali i kształtu, a i b , muszą być dodatnie. Jeśli nie są określone, ROZKŁAD szacuje je metodą najmniejszych kwadratów. Obserwacje ujemne nie są dozwolone.

Parametry rozkładu

Udostępnia strategię rozkładu i opcje parametrów.

Oszacuj na podstawie danych

Po wybraniu tego ustawienia parametry rozkładu są szacowane na podstawie danych i wybranego typu rozkładu.

Podaj

Po wybraniu tego ustawienia można oszacować parametry rozkładu dla wybranego typu rozkładu.

Uwaga: Dostępne parametry będą się różnić w zależności od wybranego typu rozkładu.

Przekształcenia

Dostępne opcje umożliwiają wybór ustawienia przekształcenia i cykliczności.

Transformacja logarytmiczna naturalna

Przekształca dane przy użyciu logarytmu naturalnego (podstawa e) w celu usunięcia różnych amplitud.

Standaryzuj wartości

Przekształca sekwencję lub szereg czasowy do próby o średniej 0 i odchyleniu standardowym 1.

Różnice

Określa rząd różnic używany przed wygenerowaniem wykresu do przekształcenia zmiennej niestacjonarnej na zmienną stacjonarną ze stałą średnią i wariancją. Wprowadź odpowiednią wartość w polu.

Różnica sezonowa

Jeśli zmienna wykazuje wzór sezonowy lub okresowy, można użyć tego ustawienia w celu zróżnicowania sezonowego zmiennej przed utworzeniem wykresów.

Uwaga: Ustawienie to jest aktywne tylko wtedy, gdy jako jedną ze zmiennych ilościowych wybrano sekwencję lub szereg czasowy z określoną okresowością.

Formuła estymacji proporcji

Dostępne opcje określają wzór, który jest używany do szacowania proporcji.

Blom

Tworzy nową zmienną rangującą na podstawie oszacowań proporcji obliczonych ze wzoru $(r-3/8) / (w+1/4)$, gdzie w jest sumą wag obserwacji, a r jest rangą.

Rankit

Wykorzystuje wzór $(r-1/2) / w$, gdzie w jest liczbą obserwacji a r jest rangą, przyjmującą wartości od 1 do w .

Tukey

Wykorzystuje wzór $(r-1/3) / (w+1/3)$, gdzie r jest rangą, a w jest sumą wag obserwacji.

Van der Waerden

Transformacja Van der Waerdena określona przez wzór $r / (w+1)$, gdzie w jest sumą wag obserwacji, a r jest rangą od 1 do w .

Rangi przypisane do wiązań

Dostępne opcje umożliwiają wybór sposobu postępowania z wartościami związanymi. Poniższa tabela wskazuje sposoby przypisywania rang do związanych wartości w różnych metodach.

Tabela 4. Metody i wyniki rangowania

Wartość	Średnia	Niska	Wysoka	Podziel wiązania arbitralnie
10	1	1	1	1
15	3	2	4	2
15	3	2	4	2
15	3	2	4	2
16	5	5	5	3
20	6	6	6	4

Otrzymywanie wykresów K-K (kwantyl-kwantyl)

1. Z menu wybierz:

Analizuj > Statystyki opisowe > Wykresy K-K...

2. Zaznacz co najmniej jedną zmienną numeryczną i przenieś ją w pole Zmienne.

3. Wybierz rozkład testowy.

Opcjonalnie można zaznaczyć opcje przekształcenia w celu otrzymania wykresów prawdopodobieństwa wartości transformowanych, a także określić metodę obliczania rozkładów oczekiwanych.

Analiza ROC

Analiza krzywej charakterystyki operacyjnej odbiornika (ROC – Receiver operating characteristic) jest użytecznym narzędziem oceny dokładności predykcji modelu poprzez wykreślenie czułości wobec (1-swoistości) testu klasyfikacyjnego (przy progu zmieniającym się w całym zakresie wyników testu diagnostycznego). Pełny obszar pod daną krzywą ROC lub pole pod krzywą AUC formułuje ważną statystykę, która reprezentuje prawdopodobieństwo, że predykcja będzie w poprawnej kolejności, gdy obserwowana jest zmienna testowa (dla jednego podmiotu losowo wybranego z grupy przypadków, a drugiego losowo wybranego z grupy kontrolnej). Analiza ROC wspomaga wywodzenie wniosków dotyczących jednego pola pod krzywą (AUC) precyzja-czułość i oferuje opcje porównania dwóch krzywych ROC wygenerowanych na podstawie grup niezależnych lub obserwacji parami.

Stara procedura ROC Curve wspiera statystyczne wnioskowanie o jednolitej krzywej ROC. Może to być również odzyskane w ramach nowej procedury analizy ROC. Ponadto nowa procedura analizy ROC może porównywać dwa krzywe ROC wygenerowane z niezależnych grup lub poddanych sparowanym obiektom.

Krzywe precyzja-czułość (PR – precision-recall) zwykle niosą więcej informacji, gdy obserwowana próba danych jest silnie skośna i stanowią alternatywę dla krzywych ROC w przypadku danych z silnie skośnym rozkładem klas.

Przykład

W interesie banku leży prawidłowe podzielenie klientów na tych, którzy spłacą i na tych, którzy nie spłacą zaciągniętych pożyczek. W celu podejmowania tych decyzji opracowywane są specjalne modele. Analizę ROC można zastosować do oceny dokładności predykcji modelu.

Statystyki

AUC, grupa ujemnych, brakujące wartości, klasyfikacja dodatnia, wartość odcięcia, siła przekonania, dwustronny asymptotyczny przedział ufności, rozkład, błąd standardowy, grupy niezależne, próby parami, założenie nieparametryczne, założenie o rozkładzie wykładniczym dwuujemnym, środek, punkt odcięcia, krzywa PR, interpolacja krokowa, istotność asymptotyczna (dwustronna), czułość i (1-swoistość), precyzja i czułość.

Metody

Porównywane są pola pod dwiema krzywymi ROC wygenerowanymi na podstawie grup niezależnych albo par obiektów. Porównanie dwóch krzywych ROC może dostarczyć więcej informacji na temat dokładności dwóch porównywalnych metod diagnostycznych.

Wymagania dotyczące danych do analizy ROC

Dane

Krzywe precyzja-czułość zwykle niosą więcej informacji, gdy obserwowane próby danych są silnie skośne. Prosta interpolacja liniowa może błędnie wygenerować nadmiernie optymistyczne oszacowanie krzywej precyzja-czułość.

Założenia

Kolejność predykcji na podstawie obserwacji zmiennej testowanej będzie prawidłowa dla jednego obiektu wybranego losowo z grupy obserwacji i drugiego losowo wybranego z grupy kontrolnej. Każda zdefiniowana grupa będzie zawierała co najmniej jedną ważną obserwację. Dla jednej procedury używana jest tylko jedna zmienna grupująca.

Uzyskiwanie analizy ROC

1. Z menu wybierz:

Analiza > Klasyfikuj > Analiza ROC

2. Wybierz co najmniej jedną zmienną testowaną prawdopodobieństwa.

3. Wybierz jedną zmienną stanu.

4. Określ *dotatnią* wartość dla zmiennej stanu.

5. Opcjonalnie wybierz opcję **Próby parami** albo wybierz jedną zmienną grupującą (nie można wybrać obu opcji).

- Ustawienie **Próby parami** wskazuje, że dla zmiennych testowanych ma być używane plan prób parami. W planie prób parami dwie krzywe ROC są porównywane w taki sposób, że wiele wartości testowanych mierzonych jest na tych samych obiektach powiązanych ze zmienną stanu.
- Uwaga:** Gdy wybrana jest opcja **Próby parami** opcje **Zmienna grupująca** i **Założenie o rozkładzie** (w oknie dialogowym Opcje) są wyłączone.
- Po wybraniu liczbowej zmiennej grupującej można kliknąć opcję **Definiuj grupy ...** w celu zażądania projektu niezależnej grupy dla zmiennych testowych, a także do określenia dwóch wartości, punktu środkowo-punktowego lub punktu podziału.
6. Opcjonalnie kliknij przycisk **Opcje**, aby określić klasyfikację, kierunek testu, parametry błędu standardowego i ustawienia braków danych.
 7. Opcjonalnie kliknij opcję **Widok**, aby zdefiniować ustawienia kreślenia i drukowania (obejmują one krzywą ROC, krzywą precyzja-czułość i ustawienia jakości modelu).
 8. Kliknij przycisk **OK**.

Analiza ROC: Opcje

Dostępne są następujące opcje analizy ROC:

Klasyfikacja

Pozwala na określenie, czy wartość odcięcia powinna być uwzględniana czy wyłączana podczas wykonywania *dotatniej* klasyfikacji. Aktualnie ustawienie to nie ma wpływu na wynik.

Kierunek testu

Udostępnia opcje określania kierunku zmian wyników oznaczającego wzrost siły przekonania, że obiekt pozytywnie przechodzi test.

Parametry oceny błędu standardowego powierzchni pod krzywą

Umożliwia określenie metody szacowania błędu standardowego powierzchni pod krzywą. Dostępne opcje to Nieparametryczny i Wykładniczy dwuujemny. Domyślne ustawienie **Nieparametryczny** umożliwia oszacowanie przy założeniu nieparametrycznym. Ustawienie **Wykładniczy dwuujemny** powoduje, że oszacowanie odbywa się przy założeniu rozkładu wykładniczego dwuujemnego.

Ta sekcja umożliwia również określenie poziomu ufności dwustronnego asymptotycznego przedziału ufności pola pod krzywą. Dostępny zakres wynosi od 0,0% do 100,0% (wartość domyślna to 95%).

Uwaga: To ustawienie ma zastosowanie tylko do planu z grupą niezależną i nie ma wpływu na plan Próby parami.

Braki danych

Pozwala na określenie sposobu postępowania z brakami danych. Gdy to ustawienie nie jest wybrane, wykluczane są zarówno braki danych użytkownika, jak i systemowe braki danych. Gdy ustawienie jest wybrane, braki danych zdefiniowane przez użytkownika traktowane są jako ważne, a systemowe braki danych są wykluczane. Obserwacje z systemowymi brakami danych, w zmiennej testującej albo w zmiennej stanu, zawsze są wykluczane z analizy.

Analiza ROC: Prezentacja

W analizie ROC można określić następujące ustawienia prezentacji:

Wykres

Udostępnia opcje kreślenia krzywej ROC i krzywej precyzja-czułość.

Krzywa ROC

Gdy ta opcja jest wybrana wyniki zawierają wykres krzywej ROC. Wybierz opcję **Z przekątną linią referencyjną**, aby nanieść przekątną linię referencyjną na wykres krzywej ROC.

Krzywa precyzja-czułość

Gdy ta opcja jest wybrana wyniki zawierają wykres krzywej precyzja-czułość. Krzywe precyzja-czułość zwykle niosą więcej informacji, gdy obserwowana próba danych jest silnie skośna i stanowią alternatywę dla krzywych ROC w przypadku danych z silnie skośnym rozkładem klas. Domyślne ustawienie **Interpoluj wzduż prawdziwie dodatnich** powoduje wykonanie tej

krokowej interpolacji wzdłuż prawdziwych wyników dodatnich. Domyślne ustawienie **Interpoluj wzdłuż fałszywie dodatnich** powoduje wykonanie tej krokowej interpolacji wzdłuż fałszywych wyników dodatnich.

Jakość modelu ogólnego

To ustawienie określa, czy tworzony jest wykres słupkowy przedstawiający wartość dolnej granicy przedziału ufności oszacowanego pola pod krzywą. Domyślnie to ustawienie nie jest wybrane, co oznacza, że wykres słupkowy jest ukryty.

Drukuj

Udostępnia opcje do definiowania wyników odpowiednich statystyk.

Błąd standardowy i przedział ufności

Ustawienie określa, które statystyki będą widoczne w tabeli „Pole pod krzywą”. Gdy to ustawienie nie jest wybrane, wyświetlane jest tylko szacowane pole pod krzywą. Gdy to ustawienie jest wybrane, wyświetlane są dodatkowe statystyki, w tym błąd standardowy pola pod krzywą, istotność asymptotyczna (dwustronna) i granice asymptotycznego przedziału ufności przy założeniu prawdziwości hipotezy zerowej.

Współrzędne punktów krzywej ROC

Ustawienie wpływa na punkty współrzędnych krzywej ROC oraz wartości odcięcia. Gdy to ustawienie nie jest wybrane, wyniki nie zawierają punktów współrzędnych. Gdy ustawienie jest wybrane, dla każdej krzywej ROC podane są pary wartości Czulość i (1-Swoistość) wraz z wartościami odcięcia.

Indeks Youdena

Wyświetla wartość indeksu Youdena dla każdej wartości odcięcia krzywej ROC.

Współrzędne punktów krzywej precyzja-czulość

Ustawienie wpływa na punkty współrzędnych krzywej precyzja-czulość oraz wartości odcięcia. Gdy to ustawienie nie jest wybrane, wyniki nie zawierają punktów współrzędnych. Gdy ustawienie jest wybrane, dla każdej krzywej precyzja-czulość podane są pary wartości Precyzja i Czulość wraz z wartościami odcięcia.

Metryka oceny klasyfikatora

To ustawienie steruje uwzględnieniem tabeli Metryka oceny klasyfikatora w wynikach. Tabela pokazuje, jak dobrze model klasyfikacji pasuje do danych porównywanych z przypisaniem losowym, i udostępnia następujące informacje:

- Zmienne testowane określone przez użytkownika
- Informacje o grupach
- Wskaźnik Giniego (wskaźnik Giniego wynosi $2 \cdot \text{AUC} - 1$, gdzie AUC jest polem pod krzywą ROC)
- Wartości Maks. K-S i Odcięcie

Analiza ROC: Definiuj grupy (łańcuch)

W przypadku łańcuchowych zmiennych grupujących wprowadź łańcuch dla grupy 1 i inną wartość dla grupy 2, na przykład *tak* i *nie*. Obserwacje z innymi łańcuchami są wykluczane z analizy.

Uwaga: Podane wartości muszą istnieć w zmiennej. W przeciwnym wypadku zostanie wyświetlony komunikat o błędzie w celu wskazania, że co najmniej jedna z grup jest pusta.

Analiza ROC: Definiuj grupy (liczba)

W przypadku liczbowych zmiennych grupujących zdefiniuj dwie grupy dla testu *t*, określając dwie wartości, punkt środkowy lub punkt podziału.

Uwaga: Podane wartości muszą istnieć w zmiennej. W przeciwnym wypadku zostanie wyświetlony komunikat o błędzie w celu wskazania, że co najmniej jedna z grup jest pusta.

- **Użyj określonych wartości.** Wprowadź jedną wartość dla grupy 1 i inną wartość dla grupy 2. Obserwacje z dowolnymi innymi wartościami są wykluczane z analizy. Liczby te nie muszą być liczbami całkowitymi (dopuszczalnymi wartościami są np. 6,25 i 12,5).

- **Użyj wartości środkowej.** W przypadku wyboru tej opcji grupy są podzielone na wartości $< i \geq$ punktu środkowego.
- **Użyj punktu podziału.**
 - **Punkt podziału.** Można również wprowadzić liczbę dzielącą wartości zmiennej grupującej na dwie grupy. Wszystkie obserwacje, których wartości są mniejsze od punktu podziału, tworzą jedną grupę, zaś obserwacje, których wartości są większe od punktu podziału lub mu równe, tworzą drugą grupę.

Krzywe ROC

Ta procedura jest użytecznym sposobem oceny wydajności schematów klasyfikacyjnych, w których podmioty klasyfikowane są według jednej zmiennej z dwiema kategoriami.

Przykład. W interesie banku leży prawidłowe podzielenie klientów na tych, którzy spłacą i na tych, którzy nie spłacą zaciągniętych pożyczek. W celu podejmowania tych decyzji opracowywane są specjalne metody. Za pomocą krzywych ROC możliwa jest ocena jakości działania metod.

Statystyki. Powierzchnia pod krzywą ROC z przedziałem ufności i współrzędnymi punktów krzywej ROC. Wykresy: krzywa ROC.

Metody. Szacunkowa powierzchnia pod krzywą ROC może być obliczona w sposób parametryczny lub nieparametryczny za pomocą dwuwymiarowego modelu wykładniczego.

Wymagania dotyczące danych krzywej ROC

Stara procedura ROC Curve wspiera statystyczny wnioskowanie o jednolitej krzywej ROC. Może to być również odzyskane w ramach nowej procedury analizy ROC. Ponadto nowa procedura analizy ROC może porównywać dwa krzywe ROC wygenerowane z niezależnych grup lub poddanych sparowanym obiektom.

Dane. Zmienne testowane są zmiennymi ilościowymi. Często składają się z prawdopodobieństw z analizy dyskryminacyjnej lub regresji logistycznej lub też wartości skali arbitralnej, wskazującej „stopień przekonania” oceniającego, że dany obiekt należy do danej kategorii. Zmienna stanu może być dowolnego typu i wskazuje prawdziwą kategorię, do której należy obiekt. Wartość zmiennej stanu wskazuje, która kategoria powinna być rozpatrywana jako *dodatnia*.

Założenia. Zakłada się, że wzrastające liczby na skali oceniającego reprezentują zwiększające się przekonanie, że obiekt należy do jednej kategorii, podczas gdy liczby zmniejszające się na skali reprezentują zwiększające się przekonanie, że dany obiekt przynależy do innej kategorii. Użytkownik musi wybrać, który kierunek jest *dodatni*. Zakłada się również, że znana jest *prawdziwa* kategoria, do której należy każdy obiekt.

Otrzymywanie krzywej ROC

1. Z menu wybierz:

Analiza > Klasyfikuj > Krzywa ROC...

2. Wybierz co najmniej jedną zmienną testowaną prawdopodobieństwa.

3. Wybierz jedną zmienną stanu.

4. Określ *dodatnią* wartość dla zmiennej stanu.

Krzywa ROC: Opcje

Dostępne są następujące opcje analizy ROC:

Klasyfikacja

Pozwala na określenie, czy wartość odcięcia powinna być uwzględniana czy wyłączana podczas wykonywania *dodatniej* klasyfikacji. Aktualnie ustawienie to nie ma wpływu na wynik.

Kierunek testu

Pozwala na określenie kierunku skali w odniesieniu do kategorii *dodatniej*.

Parametry oceny błędu standardowego powierzchni pod krzywą

Umożliwia określenie metody szacowania błędu standardowego powierzchni pod krzywą. Dostępne opcje to Nieparametryczny i Wykładniczy dwuujemny. Opcja ta umożliwia również ustalenie poziomu ufności dla przedziału ufności. Dostępny zakres wynosi od 50,1 do 99,9%.

Braki danych

Pozwala na określenie sposobu postępowania z brakami danych.

Symulacja

Modele predykcyjne, takie jak regresji liniowej, wymagają zbioru znanych informacji wejściowych do przewidzenia wartości wyjściowych lub zmiennej przewidywanej. Jednak w wielu rzeczywistych aplikacjach wartości wejściowe są niepewne. Symulacja pozwala na uwzględnienie niepewności danych wejściowych dla modeli predykcyjnych oraz ocenę prawdopodobieństwa różnych danych wyjściowych modelu w obecności tej niepewności. Na przykład: masz model zysku uwzględniający jako zmienne wejściowe koszty materiałów, ale istnieje niepewność tego kosztu związana z niestabilnością rynku. Możesz użyć symulacji do wymodelowania tej niepewności i określić, jak ma ona wpływ na zysk.

Symulacja w produkcie IBM SPSS Statistics używa metody Monte Carlo. Niepewne zmienne wejściowe są modelowane za pomocą rozkładu prawdopodobieństwa (na przykład rozkładu trójkątnego), z którego pobierane są symulowane wartości dla tych danych wejściowych. Zmienne wejściowe, których wartości są znane, są traktowane jako stałe, na poziomie znanych wartości. Model predykcyjny jest tworzony przy użyciu symulowanej wartości dla każdej niepewnej wartości wejściowej oraz stałych wartości dla znanych danych wejściowych, w celu wyliczenia zmiennej przewidywanej (lub zmiennych przewidywanych) modelu. Proces jest powtarzany wiele razy (zwykle dziesiątki tysięcy lub setki tysięcy razy), w wyniku czego otrzymuje się wartości zmiennych przewidywanych, których można użyć do odpowiedzi na pytania natury probabilistycznej. W kontekście produktu IBM SPSS Statistics każde powtórzenie procesu generuje oddzielny przypadek (rekord) danych, składający się ze zbioru symulowanych wartości dla niepewnych danych wejściowych, wartości ustalonych danych wejściowych oraz przewidzianej zmiennej przewidywanej (lub zmiennych przewidywanych) modelu.

Można również przeprowadzać symulację danych bez modelu predykcyjnego poprzez określenie rozkładów prawdopodobieństwa dla zmiennych, które będą uwzględniane w symulacji. Każda wygenerowana obserwacja danych zawiera zestaw symulowanych wartości dla określonych zmiennych.

Aby uruchomić symulację konieczne jest określenie szczegółów takich jak model predykcyjny, prawdopodobieństwo rozkładów dla niepewnych wartości wejściowych, korelacji między tymi danymi wejściowymi a wartościami dowolnych ustalonych wartości wejściowych. Po określeniu wszystkich szczegółów symulacji, można ją uruchomić i, opcjonalnie, zapisać specyfikację w pliku **planu symulacji**. Plan symulacji można udostępnić innym użytkownikom, którzy następnie mogą uruchomić symulację bez potrzeby zrozumienia szczegółów jej tworzenia.

Do pracy z symulacjami dostępne są dwa interfejsy. Kreator symulacji jest zaawansowanym interfejsem dla użytkowników projektujących i uruchamiających symulacje. Dostarcza on pełen zestaw funkcji umożliwiających projektowanie symulacji, zapisywanie specyfikacji w pliku planu symulacji, określanie wartości wyjściowych oraz uruchomienie symulacji. Możesz zbudować symulację opartą o plik modelu IBM SPSS lub o zbiór niestandardowych równań definiowanych w kreatorze symulacji. Możesz także załadować w kreatorze symulacji istniejący plan symulacji, zmodyfikować którekolwiek z ustawień oraz uruchomić symulację, opcjonalnie zapisując zaktualizowany plan. Dla użytkowników posiadających plan symulacji, którzy tylko chcą uruchomić symulację, dostępna jest prostsza wersja interfejsu. Pozwala ona na zmodyfikowanie ustawień pozwalających na uruchomienie symulacji dla różnych warunków, ale nie zapewnia wszystkich możliwości kreatora symulacji dla projektowania aplikacji.

Projektowanie symulacji opartej o plik modelu

1. Z menu wybierz:

Analiza > Symulacja...

2. Kliknij **Wybierz plik modelu SPSS**, a następnie kliknij **Dalej**.

3. Otwórz plik modelu.

Plik modelu to plik w formacie XML zawierający model PMML utworzony przez IBM SPSS Statistics lub IBM SPSS Modeler. Aby uzyskać dodatkowe informacje, patrz temat: "Karta Model" na stronie 252.

4. W karcie Symulacja (w kreatorze symulacji) określ rozkłady prawdopodobieństwa dla symulowanych danych wejściowych oraz wartości dla ustalonych danych wejściowych. Jeśli aktywny zbiór danych zawiera historyczne dane dla symulowanych wartości wejściowych, kliknij **Dopasuj wszystkie**, aby automatycznie określić dla każdej takich danych wejściowych rozkład, który najlepiej do niej pasuje oraz ustala korelacje między danymi. W przypadku wszystkich symulowanych danych wejściowych niepasujących do danych historycznych należy jednoznacznie określić rozkład, wybierając typ rozkładu i wprowadzając żądane parametry.
5. Kliknij przycisk **Uruchom**, aby uruchomić symulację. Domyślnie plan symulacji określający szczegóły symulacji jest zapisywany w lokalizacji określonej w ustawieniach Zapisywania.

Dostępne są następujące opcje:

- Zmodyfikować lokalizację dla zapisanego planu symulacji.
- Określić znane korelacje między symulowanymi danymi wejściowymi.
- Automatycznie obliczyć wartości w tabeli kontyngencji dla powiązań pomiędzy jakościowymi danymi wejściowymi i użyć tych powiązań podczas generowania informacji dla tych danych wejściowych.
- Określić analizę czułości, aby zbadać wpływ zmieniającej się wartości ustalonych danych wejściowych lub zmieniającego się parametru rozkładu dla symulowanych danych wejściowych.
- Określić ustawienia opcji zaawansowanych takich, jak ustawienie maksymalnej liczby obserwacji do wygenerowania lub żądanie próbkowania wartości krańcowych.
- Dopasowywać dane wyjściowe.
- Zapisać symulowane dane w pliku danych.

Projektowanie symulacji opartej o niestandardowe równania

1. Z menu wybierz:

Analiza > Symulacja...

2. Kliknij **Wpisz równania** i kliknij **Dalej**.
3. W karcie Modelu (w kreatorze symulacji) kliknij **Nowe równanie**, aby zdefiniować każde z równań swojego modelu predykcyjnego.
4. Kliknij kartę Symulacja i określ rozkład prawdopodobieństwa dla symulowanych danych wejściowych oraz wartości dla ustalonych danych wejściowych. Jeśli aktywny zbiór danych zawiera historyczne dane dla symulowanych wartości wejściowych, kliknij **Dopasuj wszystkie**, aby automatycznie określić dla każdej takich danych wejściowych rozkład, który najlepiej do niej pasuje oraz ustala korelacje między danymi. W przypadku wszystkich symulowanych danych wejściowych niepasujących do danych historycznych należy jednoznacznie określić rozkład, wybierając typ rozkładu i wprowadzając żądane parametry.
5. Kliknij przycisk **Uruchom**, aby uruchomić symulację. Domyślnie plan symulacji określający szczegóły symulacji jest zapisywany w lokalizacji określonej w ustawieniach Zapisywania.

Dostępne są następujące opcje:

- Zmodyfikować lokalizację dla zapisanego planu symulacji.
- Określić znane korelacje między symulowanymi danymi wejściowymi.
- Automatycznie obliczyć wartości w tabeli kontyngencji dla powiązań pomiędzy jakościowymi danymi wejściowymi i użyć tych powiązań podczas generowania informacji dla tych danych wejściowych.
- Określić analizę czułości, aby zbadać wpływ zmieniającej się wartości ustalonych danych wejściowych lub zmieniającego się parametru rozkładu dla symulowanych danych wejściowych.
- Określić ustawienia opcji zaawansowanych takich, jak ustawienie maksymalnej liczby obserwacji do wygenerowania lub żądanie próbkowania wartości krańcowych.
- Dopasowywać dane wyjściowe.

- Zapisać symulowane dane w pliku danych.

Projektowanie symulacji bez uwzględnienia modelu predykcyjnego

1. Wybierz z menu następującą opcję:

Analiza > Symulacja...

2. Kliknij opcję **Utwórz dane symulowane**, a następnie kliknij **Dalej**.
3. W karcie Model (w kreatorze symulacji) wybierz zmienne, które mają zostać uwzględnione w symulacji. Możesz wybrać zmienne z aktywnego zbioru danych lub zdefiniować nowe zmienne, klikając opcję **Nowe**.
4. Kliknij zakładkę Symulacja i określ rozkłady prawdopodobieństwa dla zmiennych, które będą poddane symulacji. Jeśli aktywny zbiór danych zawiera dane historyczne dotyczące którejś zmiennej, kliknij **Dopasuj wszystkie**, aby automatycznie określić rozkład, który najlepiej pasuje do danych i ustala korelacje pomiędzy zmiennymi. W przypadku zmiennych niepasujących do danych historycznych należy jednoznacznie określić rozkład, wybierając typ rozkładu i wprowadzając żądane parametry.
5. Kliknij przycisk **Uruchom**, aby uruchomić symulację. Domyślnie symulowane dane są zapisywane w nowym zbiorze danych określonym w ustawieniach opcji Zapisz. Ponadto plan symulacji określający szczegóły symulacji jest zapisywany w lokalizacji określonej w ustawieniach Zapisywania.

Dostępne są następujące opcje:

- Modyfikowanie lokalizacji dla symulowanych danych lub zapisanego planu symulacji.
- Określanie znanych korelacji między symulowanymi zmiennymi.
- Automatyczne obliczanie wartości w tabeli kontyngencji dla powiązań pomiędzy jakościowymi danymi wejściowymi i używanie tych powiązań podczas generowania informacji dla tych zmiennych.
- Określanie analizy czułości w celu zbadania wpływu zmieniającego się parametru rozkładu dla symulowanej zmiennej.
- Określanie ustawień opcji zaawansowanych, takich jak ustawianie liczby obserwacji do wygenerowania.

Uruchamianie symulacji z wykorzystaniem planu symulacji

Dla uruchamiania symulacji stworzonej według planu symulacji dostępne są dwie opcje. Możesz użyć okna dialogowego Uruchom symulację, który przede wszystkim został zaprojektowany do uruchamiania z wykorzystaniem planu symulacji, możesz też skorzystać z kreatora symulacji.

Aby użyć okna dialogowego kreatora symulacji:

1. Z menu wybierz:

Analiza > Symulacja...

2. Kliknij **Otwórz istniejący plan symulacji**.
3. Upewnij się, że pole wyboru **Otwórz w kreatorze symulacji** nie jest zaznaczone i kliknij przycisk **Dalej**.
4. Otwórz plan symulacji.
5. W oknie Uruchom symulację kliknij przycisk **Uruchom**.

Aby uruchomić symulację korzystając z kreatora symulacji:

1. Z menu wybierz:

Analiza > Symulacja...

2. Kliknij **Otwórz istniejący plan symulacji**.
3. Zaznacz pole wyboru **Otwórz w kreatorze symulacji** i kliknij **Dalej**.
4. Otwórz plan symulacji.
5. Na karcie Symulacja zmodyfikuj wszystkie żądane ustawienia.
6. Kliknij przycisk **Uruchom**, aby uruchomić symulację.

Opcjonalnie można wykonać następujące czynności:

- Skonfigurować lub zmodyfikować analizę czułości, aby zbadać wpływ zmieniającej się wartości ustalonych danych wejściowych lub zmieniającego się parametru rozkładu dla symulowanych danych wejściowych.
- Ponownie dopasować rozkłady i korelacje symulowanych danych wejściowych do nowych danych.
- Zmienić rozkład dla symulowanych danych wejściowych.
- Dopasowywać dane wyjściowe.
- Zapisać symulowane dane w pliku danych.

Kreator symulacji

Kreator symulacji zapewnia cały zestaw funkcji służących do projektowania i uruchamiania symulacji. Pozwala na przeprowadzanie następujących zadań ogólnych:

- Projektowanie i uruchamianie symulacji dla modelu IBM SPSS zdefiniowanego w pliku modelu PMML.
- Projektowanie i uruchamianie symulacji dla modelu predykcyjnego zdefiniowanego przez zestaw określonych przez użytkownika równań niestandardowych.
- Projektowanie i uruchamianie symulacji generujących dane bez uwzględniania modelu predykcyjnego.
- Uruchamianie symulacji opartej o istniejący plan symulacji, opcjonalnie modyfikując dowolne ustawienia planu.

Karta Model

W przypadku symulacji na podstawie modelu predykcyjnego karta Model określa źródło modelu.

W przypadku symulacji nieobejmujących modelu predykcyjnego karta Model określa pola, które nie będą poddawane symulacji.

Wybierz plik modelu SPSS. Ta opcja określa, że model predykcyjny jest zdefiniowany w pliku modelu IBM SPSS. Plik modelu IBM SPSS to plik XML lub plik skompresowany (.zip), który zawiera model PMML utworzony przez program IBM SPSS Statistics lub IBM SPSS Modeler. Modele predykcyjne są tworzone przez procedury takie, jak procedura Regresji liniowej oraz Drzew decyzyjnych w obrębie IBM SPSS Statistics i mogą zostać wyeksportowane do pliku modelu. Można użyć innego pliku modelu, klikając przycisk **Przełączaj** i przechodząc do żądanego pliku.

Modele PMML obsługiwane przez Symulację

- Regresja liniowa
- Automatyczny model liniowy
- Uogólniony model liniowy
- Uogólniony liniowy model mieszany
- Uogólniony model liniowy
- Binarna regresja logistyczna
- Wielomianowa regresja logistyczna
- Wielomianowa regresja porządkowa
- Regresja Coxa
- Drzewo
- Wzmocnione drzewo (C5)
- Dyskryminacja
- Grupowanie dwustopniowe
- Grupowanie metodą k-średnich
- Sieć neuronowa
- Zestaw reguł (lista decyzyjna)

Uwaga:

- Modele PMML zawierające wiele pól zmiennych przewidywanych (zmiennych) lub podziałów i nie są obsługiwane w Symulacji.
- Wartości łańcuchowych danych wejściowych dla modeli binarnej regresji logistycznej są ograniczone do 8 bajtów w modelu. Podczas dopasowywania tego typu łańcuchowych danych wejściowych do aktywnego zbioru danych upewnij się, czy wartości danych nie są dłuższe niż 8 bajtów. Wartości danych przekraczające 8 bajtów są wykluczane z powiązanego rozkładu kategorialnego danych wejściowych i w tabeli kategorii niedopasowanych są wyświetlane jako niedopasowane.

Wpisz równania dla modelu. Ta opcja precyzuje, że model predykcyjny składa się z jednego lub więcej równań niestandardowych stworzonych przez użytkownika. Stwórz równania, klikając **Nowe równanie**. Spowoduje to otwarcie Edytora równań. Możesz zmodyfikować istniejące równania, skopiować je, by móc użyć je jako szablonu dla nowych równań, zmienić ich kolejność lub usunąć je.

- Kreator symulacji nie obsługuje systemów równań współzależnych ani równań, które są nieliniowe ze względu na zmienną przewidywaną.
- Równania niestandardowe są oceniane według kolejności, w jakiej są określone. Jeśli równanie dla danej zmiennej przewidywanej zależy od innej zmiennej przewidywanej, wtedy zmienna ta musi zostać określona przez poprzednie równanie.

Na przykład: biorąc pod uwagę poniższe trzy równania, równanie dla *zysku* zależy od wartości *przychodu* i *wydatków*, więc równania dla *przychodu* i *wydatków* muszą znaleźć się przed równaniem dla *zysku*.

$\text{przychód} = \text{cena} \times \text{rozmiar}$

$\text{wydatki} = \text{stałe} + \text{rozmiar} \times (\text{koszt_materiałów_dla_jednostki} + \text{koszt_pracy_dla_jednostki})$

$\text{zysk} = \text{przychód} - \text{wydatki}$

Utwórz dane symulowane bez modelu. Zaznacz tę opcję, aby przeprowadzić symulację danych bez uwzględniania modelu predykcyjnego. Określ zmienne, jakie mają zostać zasymulowane, zaznaczając pola w aktywnym zbiorze danych lub klikając opcję **Nowe**, umożliwiającą zdefiniowanie nowych zmiennych.

Edytor równań

Edytor równań pozwala na tworzenie lub modyfikowanie równań niestandardowych dla modelu predykcyjnego.

- Wyrażenie dla równania może zawierać pola z aktywnego zbioru danych lub nowe pola danych wejściowych, które definiuje się w Edytorze równań.
 - Możesz określić właściwości zmiennej przewidywanej takie, jak jej poziom pomiaru, etykiety wartości i czy dla zmiennej przewidywanej generowane są dane wyjściowe.
 - Możesz użyć zmiennych przewidywanych pochodzących ze zdefiniowanych wcześniej równań jako zmienne wejściowe dla bieżącego równania i w ten sposób tworzyć równania sprzężone.
 - Do równania możesz dołączyć komentarz z opisem. Komentarze są wyświetlane razem z równaniem w karcie Modelu.
1. Wprowadź nazwę zmiennej przewidywanej. Można też kliknąć opcję **Edytuj** znajdującą się w polu tekstowym Zmienna przewidywana, aby otworzyć okno dialogowe Zdefiniowane zmienne, co pozwoli na zmianę domyślnych właściwości przewidywanej.
 2. Aby utworzyć wyrażenie, można wkleić składowe w pole Wyrażenie liczbowe lub wpisać je tam bezpośrednio.
- Możesz utworzyć własne wyrażenie używając pól z aktywnego zbioru danych bądź też możesz zdefiniować nowe zmienne wejściowe klikając przycisk **Nowe**. Spowoduje to otwarcie okna dialogowego Zdefiniuj zmienne wejściowe.
 - Możesz wkleić zmienne poprzez wybranie grupy z grupy z listy grup Funkcji i dwukrotne kliknięcie funkcji lub zmiennej na liście Funkcji (lub wybrać funkcję lub zmienną i kliknąć przycisk ze strzałką znajdujący się obok listy funkcji grupy). Wprowadź wszystkie parametry oznaczone znakami zapytania.

Funkcja grupy o nazwie **Wszystkie** zapewnia listę wszystkich dostępnych funkcji. Krótki opis aktualnie wybranej funkcji jest wyświetlony w zarezerwowanym obszarze okna dialogowego.

- Stałe łańcuchowe muszą być ujęte w cudzysłów.
- Jeśli wartości zawierają dziesiętne, do wskazania dziesiętnej należy użyć kropki (.).

Uwaga: symulacja nie obsługuje równań niestandardowych ze zmiennymi przewidywanymi w formie łańcuchów znaków.

Zdefiniowane zmienne

Okno dialogowe Zdefiniuj zmienne wejściowe pozwala na zdefiniowanie nowych danych wejściowych i na ustawienie właściwości zmiennych przewidywanych.

- Jeśli zmienne wejściowe przeznaczone do użycia w równaniu nie istnieją w aktywnym zbiorze danych, należy określić je przed użyciem w równaniu.
- Jeśli symulacja dotyczy danych bez użycia modelu predykcyjnego, należy zdefiniować wszystkie symulowane zmienne wejściowe, które nie istnieją w aktywnym zbiorze danych.

Name. Określ nazwę dla zmiennej przewidywanej lub danych wejściowych.

Zmienna przewidywana. Możesz określić poziom pomiaru dla przewidywanej. Domyślny poziom pomiaru to: ciągły. Możesz także określić, czy dla tej docelowej zostaną stworzone dane wyjściowe. Na przykład: w przypadku zbioru równań sprzężonych mogą Cię tylko interesować dane wyjściowe dla zmiennej przewidywanej w końcowym równaniu, więc będziesz chciał ukryć dane wyjściowe innych zmiennych przewidywanych.

Dane wejściowe do zasymulowania. Określa, czy wartości danych wejściowych zostaną zasymulowane z wykorzystaniem określonego rozkładu prawdopodobieństwa (rozkład prawdopodobieństwa określa się w karcie Symulacji). Poziom pomiaru określa domyślny zbiór rozkładów, które są brane pod uwagę podczas wyszukiwania najlepiej dopasowanego rozkładu dla danych wejściowych (po kliknięciu pozycji **Dopasuj** lub **Dopasuj wszystkie** znajdujących się w karcie Symulacji). Przykładowo: jeśli poziom pomiaru jest ciągły, to rozważany będzie rozkład normalny (odpowiedni dla danych ciągłych), a nie będzie rozważany rozkład dwumianowy.

Uwaga: Dla danych wejściowych typu łańcuchowego wybierz poziom pomiaru typu łańcuch. Poddawane symulacji zmienne wejściowe typu łańcuchowego są ograniczone do rozkładu kategoryjnego.

Dane wejściowe o ustalonej wartości. Określa, czy wartość wejścia jest znana i czy zostanie ona ustalona na poziomie wartości znanej. Ustalane wartości wejścia mogą być numeryczne lub łańcuchowe. Określ wartość dla ustalonej wartości wejścia. Wartości łańcuchowe nie powinny być ujęte w cudzysłów.

Etykiety wartości. Etykiety wartości można określić dla zmiennych przewidywanych, symulowanych wartości wejściowych oraz ustalonych wartości wejściowych. Etykiety wartości są używane w wykresach i tabelach zmiennych wyjściowych.

Karta Symulacji

Karta Symulacji określa wszystkie właściwości symulacji, poza modelem predykcyjnym. Korzystając z karty Symulacji, można wykonywać następujące zadania ogólne:

- Określać rozkład prawdopodobieństwa dla symulowanych danych wejściowych oraz wartości dla ustalonych danych wejściowych.
- Określać korelacje między symulowanymi danymi wejściowymi. W przypadku jakościowych danych wejściowych można określić, czy powiązania, które istnieją pomiędzy tymi danymi wejściowymi w aktywnym zbiorze danych będą stosowane podczas generowania danych dla tych danych wejściowych.
- Określać zaawansowane opcje takie, jak próbkowanie wartości krańcowych i kryteria dopasowania rozkładów do danych historycznych.
- Dopasowywać dane wyjściowe.
- Określać, gdzie zapisywać plan symulacji i, opcjonalnie, zapisywać zasymulowane dane.

Symulowane zmienne

Aby uruchomić symulację, każda zmienna wejściowa musi zostać określona jako ustalona lub symulowana. Symulowane zmienne wejściowe to takie, których wartości są niepewne i zostaną wygenerowane z rysunku określonego rozkładu prawdopodobieństwa. Gdy w przypadku danych wejściowych przeznaczonych do symulacji są dostępne dane historyczne, można określić rozkład najlepiej pasujący do danych oraz korelacje między danymi wejściowymi. Można także ręcznie określić rozkłady lub korelacje, jeśli historyczne dane nie są dostępne lub wymagane są konkretne rozkłady bądź korelacje.

Stałe wartości wejściowe to takie, których wartości są znane i pozostają stałe dla każdego przypadku wygenerowanego na drodze symulacji. Na przykład: masz model regresji liniowej dla sprzedaży, który jest funkcją liczby wejść, łącznie z ceną, a chcesz utrzymać cenę na stałym poziomie bieżącej ceny rynkowej. W tym przypadku cenę określa się jako ustalone zmienne wejściowe.

W przypadku symulacji na podstawie modelu predykcyjnego każdy predyktor w modelu stanowi zmienną wejściową dla symulacji. W przypadku symulacji nieobjętych modelem predykcyjnym zmiennymi wejściowymi symulacji są zmienne określone na karcie Model.

Automatycznie dopasowywane rozkłady i obliczanie korelacji dla symulowanych danych wejściowych. Jeśli aktywny zbiór danych zawiera dane historyczne dla wejść, które zostaną zasymulowane, to można wtedy automatycznie znaleźć najlepiej pasujące rozkłady dla tych danych wejściowych, a także określić dowolne korelacje między nimi. Należy wykonać następujące kroki:

1. Sprawdź, czy każde z danych wejściowych, które chcesz zasymulować, jest dopasowane do odpowiedniego pola w aktywnym zbiorze danych. Zmienne wejściowe zostały wymienione w kolumnie Zmienne wejściowe, a kolumna Pasują do wyświetla dopasowane pola w aktywnym zbiorze danych. Możesz dopasować daną wejściową do innego pola w aktywnym zbiorze danych wybierając inny element z rozwijanej listy kolumny Pasują do.

Wartość *-Brak-* w kolumnie Pasują do wskazuje, że nie udało się dopasować zmienne wejściowe do pola w aktywnym zbiorze danych. Domyślnie zmienne wejściowe są dopasowywane do pól zbioru danych, biorąc pod uwagę poziom nazwy i pomiaru oraz typ (wartość liczbowa lub łańcuchowa). Jeśli aktywny zbiór danych nie zawiera danych historycznych dla wejścia, możesz wtedy ręcznie określić rozkład dla danych wejściowych lub oznaczyć wejście jako ustalone zmienne wejściowe tak, jak opisano poniżej.

2. Kliknij przycisk **Dopasuj wszystkie**.

W kolumnie Rozkład wyświetlany jest najlepiej pasujący rozkład razem z powiązanimi z nim parametrami, a także wykres rozkładu nałożony na histogram (lub wykres słupkowy) danych historycznych. Korelacje między symulowanymi wejściami są wyświetlane w ustawieniach Korelacji. Możesz sprawdzić wyniki dopasowywania i podać wartości niestandardowe automatycznego dopasowania rozkładu dla konkretnych danych wejściowych, wybierając wiersz danych wejściowych i klikając opcję **Szczegóły dopasowania**. Aby uzyskać dodatkowe informacje, patrz temat: [“Szczegóły dopasowania”](#) na stronie 257.

Dla konkretnych danych wejściowych możesz uruchomić automatyczne dopasowywanie rozkładu wybierając wiersz danych wejściowych i klikając opcję **Szczegóły dopasowania**. Automatycznie wyliczane są także korelacje dla wszystkich zasymulowanych danych wejściowych dopasowanych do pól w aktywnym zbiorze danych.

Uwaga:

- Obserwacje z brakami danych we wszelkich symulowanych danych wejściowych zostaną wykluczone z dopasowania rozkładu, obliczania korelacji oraz obliczania opcjonalnej tabeli kontyngencji (dla danych wejściowych z rozkładem jakościowym). Opcjonalnie można określić, czy brakujące wartości użytkownika dla danych wejściowych z rozkładem jakościowym są traktowane jako poprawne. Domyślnie są traktowane jako brakujące. Więcej informacji można znaleźć w temacie [“Opcje zaawansowane”](#) na stronie 259.
- W przypadku ilościowych i jakościowych danych wejściowych, jeśli nie można znaleźć dopuszczalnego dopasowania dla żadnego testowanego rozkładu, jako najlepsze dopasowanie sugeruje się użycie rozkładu empirycznego. W przypadku danych wejściowych ciągłych rozkład Empiryczny jest funkcją

skumulowanego rozkładu danych historycznych. W przypadku danych wejściowych porządkowych rozkład Empiryczny jest skumulowanym rozkładem kategorialnym danych historycznych.

Ręczne określanie rozkładów. Dla dowolnej symulowanych danych wejściowych można ręcznie określić rozkład prawdopodobieństwa, wybierając rozkład z rozwijanej listy **Typ** i wprowadzając parametry rozkładu w siatce Parametrów. Po wprowadzeniu parametrów rozkładu obok siatki Parametrów zostanie wyświetlony wykres próbny rozkładu nakreślony na podstawie podanych parametrów. Poniżej przedstawiono uwagi na temat konkretnych rozkładów:

- **Jakościowy.** Jakościowy rozkład kategorii określa zmienną wejściową, która ma ustaloną liczbę wartości zwanych kategoriami. Każda kategoria posiada powiązane prawdopodobieństwo przypisane w taki sposób, by suma wszystkich prawdopodobieństw dla wszystkich kategorii wynosi jeden. Aby wprowadzić kategorię, kliknij kolumnę po lewej stronie w siatce Parametry i określ kategorię jako wartość liczbową. W kolumnie znajdującej się z prawej strony wprowadź prawdopodobieństwo powiązane z kategorią.

Uwaga: Dla jakościowych danych wejściowych z modelu PMML dostępne są kategorie ustalone na podstawie modelu, których nie można modyfikować.

- **Dwumianowy ujemny – niepowodzenia.** Opisuje rozkład liczby niepowodzeń w sekwencji prób zanim osiągnięto określoną liczbę sukcesów. Parametr *thresh* jest określoną liczbą sukcesów, a parametr *prob* to prawdopodobieństwo sukcesu dla podanej próby.
- **Dwumianowy ujemny – próby.** Opisuje rozkład liczby wymaganych prób przed osiągnięciem określonej liczby sukcesów. Parametr *thresh* jest określoną liczbą sukcesów, a parametr *prob* to prawdopodobieństwo sukcesu dla podanej próby.
- **Przedział.** Ta dystrybucja składa się z zestawu przedziałów z prawdopodobieństwem przypisanym do każdego z nich, tak aby suma prawdopodobieństw dla wszystkich przedziałów wynosiła 1. Wartości w danym przedziale czasu wywodzą się z rozkładu jednostajnego zdefiniowanego w tym przedziale czasu. Przedziały określa się wprowadzając wartość minimalną, wartość maksymalną oraz powiązane z nim prawdopodobieństwo.

Na przykład: uważasz, że koszt surowca ma 40% szans na znalezienie się w przedziale od \$10 do \$15 za jednostkę i 60% szans na znalezienie się w przedziale od \$15 do \$20 za jednostkę. Koszt zostałby zamodelowany za pomocą rozkładu Przedziału składającego się z dwóch przedziałów [10 - 15] i [15 - 20] z ustawieniem prawdopodobieństwa powiązanego z pierwszym przedziałem jako 0,4 oraz prawdopodobieństwa skojarzonego z drugim przedziałem jako 0,6. Przedziały nie muszą być przedziałami sąsiadującymi ze sobą, a nawet mogą na siebie zachodzić. Na przykład: mógłbyś określić przedziały: \$10 - \$15 i \$20 - \$25 lub \$10 - \$15 i \$13 - \$16.

- **Rozkład Weibulla.** Parametr *c* jest opcjonalnym parametrem lokalizacji określającym, w którym miejscu znajduje się początek rozkładu.

Parametry poniższych rozkładów mają takie same znaczenie jak w powiązanych funkcjach zmiennych losowych dostępnych w oknie dialogowym Oblicz wartości zmiennej: Bernoulliego, Beta, Dwumianowy, Wykładniczy, Gamma, Lognormalny, Ujemny dwumianowy (Próby i Niepowodzenia), Normalny, Poissona i jednostajny.

Określanie ustalonych danych wejściowych. Określ ustaloną zmienną wejściową wybierając Stałe z rozwijanej listy **Typ** kolumny Rozkład i wprowadzając stałą wartość. Wartość może być typu liczbowego lub łańcuchowego, w zależności od tego, czy dana wejściowa jest typu liczbowego czy łańcuchowego. Wartości łańcuchowe nie powinny być ujęte w cudzysłów.

Określanie granic wartości symulowanych. Większość rozkładów obsługuje określanie górnych i dolnych granic symulowanych wartości. Możesz określić dolną granicę wpisując wartość w polu tekstowym **Min**, a górną granicę - wpisując wartość w polu tekstowym **Max**.

Blokowanie danych wejściowych. Blokowanie danych wejściowych, przeprowadzane przez zaznaczenie w kolumnie pola wyboru z ikoną kłódki, wyklucza daną wejściową z automatycznego dopasowywania rozkładu. Jest to najbardziej użyteczne przy ręcznym określaniu rozkładu lub wartości stałej, gdy chcesz się upewnić, że automatyczne dopasowywanie rozkładu nie będzie mieć na niego wpływu. Blokowanie jest też pomocne, gdy masz zamiar udostępnić swój plan symulacji użytkownikom, którzy będą uruchamiać ją w oknie dialogowym Uruchom symulację i nie chcesz, by nie dokonywali jakichkolwiek






zmian w konkretnych danych wejściowych. W związku z tym specyfikacje dla zablokowanych danych wejściowych nie mogą być modyfikowane przez okno dialogowe Uruchom symulację.

Analiza czułościowa. Analiza czułości pozwala na sprawdzenie skutków systematycznych zmian ustalonych danych wejściowych lub parametru rozkładu na Symulowane zmienne wejściowe przez generowanie niezależnego zbioru symulowanych obserwacji (czyli osobnej symulacji) dla każdej z określonych wartości. Aby określić analizę czułości, zaznacz stałe lub symulowane zmienne wejściowe i kliknij **Analiza czułości**. Analiza czułości jest ograniczona do pojedynczej statych danych wejściowych lub pojedynczego parametru rozkładu dla symulowanych danych wejściowych. Więcej informacji można znaleźć w temacie [“Analiza czułości”](#) na stronie 258.

Ikony stanu dopasowania

Ikony znajdujące się w kolumnie Dopasuj do wskazują stan dopasowania każdego pola wejściowego.

Tabela 5. Ikony statusu

Ikona	Opis
	Dla danych wejściowych nie określono żadnego rozkładu ani dana wejściowa nie została określona jako stała. Aby uruchomić symulację, musisz albo określić rozkład dla tej zmiennej wejściowej lub zdefiniować ją jako stałą i określić jej stałą wartość.
	Dana wejściowa została wcześniej dopasowana do pola, które nie istnieje w aktywnym zbiorze danych. Nie jest konieczne żadne działanie, jeśli nie chcesz ponownie dopasowywać rozkładu danych wejściowych do aktywnego zbioru danych.
	Najlepiej dopasowany rozkład został zastąpiony alternatywnym rozkładem z okna dialogowego Dopasuj szczegóły.
	Zmienne wejściowe zostały ustawione jako najlepiej dopasowany rozkład.
	Rozkład został określony ręcznie lub dla tych danych wejściowych określono iteracje analizy czułości

Szczegóły dopasowania

Okno dialogowe Szczegóły dopasowywania wyświetla wyniki automatycznego dopasowania rozkładu dla konkretnych danych wejściowych. Rozkłady są uporządkowane według dobroci dopasowania, a najlepiej pasujący rozkład wymieniony jest jako pierwszy. Można zastąpić najlepiej dopasowany rozkład innym, wybierając przycisk opcji żądanego rozkładu w kolumnie Użyj. Wybór wartości dla przycisku opcji w kolumnie Użyj wyświetla także wykres rozkładu nałożony na histogram (lub wykres słupkowy) danych historycznych dla tych wartości wejściowych.

Statystyka dopasowania. Aby określić dobroć dopasowania domyślnie oraz dla zmiennych ciągłych, używany jest test Andersona-Darlinga. Opcjonalnie (tylko w przypadku zmiennych ciągłych) do sprawdzenia dobroci dopasowania można użyć testu Kołmogorowa-Smirnowa, zaznaczając swój wybór w ustawieniach Opcji zaawansowanych. Dla ilościowych danych wejściowych wyniki obydwu testów pokazywane są w kolumnie Statystyka dopasowania (A oznacza test Andersona-Darlinga, a K test Kołmogorowa-Smirnowa), a do uporządkowania kolejności rozkładów używany jest test wybrany przez użytkownika. Dla porządkowych i nominalnych danych wejściowych używany jest test chi-kwadrat. Pokazane są także powiązane z testami wartości p.

Parametry. Parametry rozkładu powiązane z każdym dopasowanym rozkładem są wyświetlane w kolumnie Parametry. Parametry poniższych rozkładów mają takie same znaczenie jak w powiązanych funkcjach zmiennych losowych dostępnych w oknie dialogowym Oblicz wartości zmiennej: Bernoulliego, Beta, Dwumianowy, Wykładniczy, Gamma, Lognormalny, Ujemny dwumianowy (Próby i Niepowodzenia), Normalny, Poissona i jednostajny. Dla rozkładu jakościowego kategorii nazwy parametrów są kategoriami, a wartości parametrów to powiązane prawdopodobieństwa.

Ponowne dopasowanie z użyciem dostosowanego zbioru rozkładów. Domyślnie do ustalenia zbioru rozkładów, które mogą zostać użyte do automatycznego dopasowania rozkładów, używany jest poziom pomiaru danych wejściowych. Przykładowo: rozkłady ilościowe takie, jak Lognormalny i Gamma są brane pod uwagę, gdy dopasowanie jest przeprowadzane dla ciągłego wejścia, a rozkłady dyskretne takie jak rozkład Poissona czy Dwumianowy nie są brane pod uwagę. Można wybrać podzbiór rozkładów domyślnych, wybierając rozkłady w kolumnie Dopasuj ponownie. Można także zastąpić domyślny zbiór rozkładów, wybierając inny poziom pomiaru z listy rozwijanej **Przyjmij (Poziom pomiaru)** i zaznaczając żądane rozkłady w kolumnie Dopasuj ponownie. Kliknij **Uruchom ponowne dopasowanie**, aby ponownie dopasować niestandardowy zbiór rozkładów.

Uwaga:

- Obserwacje z brakami danych we wszelkich symulowanych danych wejściowych zostaną wykluczone z dopasowania rozkładu, obliczania korelacji oraz obliczania opcjonalnej tabeli kontyngencji (dla danych wejściowych z rozkładem jakościowym). Opcjonalnie można określić, czy brakujące wartości użytkownika dla danych wejściowych z rozkładem jakościowym są traktowane jako poprawne. Domyślnie są traktowane jako brakujące. Więcej informacji można znaleźć w temacie [“Opcje zaawansowane”](#) na stronie 259.
- W przypadku ilościowych i jakościowych danych wejściowych, jeśli nie można znaleźć dopuszczalnego dopasowania dla żadnego testowanego rozkładu, jako najlepsze dopasowanie sugeruje się użycie rozkładu empirycznego. W przypadku danych wejściowych ciągłych rozkład Empiryczny jest funkcją skumulowanego rozkładu danych historycznych. W przypadku danych wejściowych porządkowych rozkład Empiryczny jest skumulowanym rozkładem kategorialnym danych historycznych.

Analiza czułości

Analiza czułości pozwala na sprawdzenie skutków zmiany wartości stałych danych wejściowych lub parametru rozkładu symulowanych danych wejściowych dla określonego zbioru wartości. Dla każdej z określonych wartości generowany jest niezależny zbiór symulowanych obserwacji (w efekcie - osobna symulacja), co pozwala na sprawdzenie skutków zmieniających się danych wejściowych. Każdy zbiór symulowanych obserwacji jest określany jako **iteracja**.

Iteracja. Wybór ten pozwala na określenie zbioru wartości, które będą wartościami zmieniających się danych wejściowych.

- Jeśli zmieniana jest wartość parametru rozkładu, parametr należy zaznaczyć, wybierając go z rozwijanej listy. Wprowadź zbiór wartości w wartości Parametrów obok siatki iteracji. Kliknięcie **Dalej** doda określone wartości do siatki Parametrów dla powiązanych danych wejściowych z indeksem określającym liczbę iteracji dla wartości.
- Dla Jakościowego rozkładu kategorii oraz rozkładu rozstępu prawdopodobieństwa kategorii lub przedziałów (odpowiednio) mogą być różne, ale wartości kategorii i granice przedziałów nie mogą ulegać zmianie. Wybierz kategorię lub przedział z listy rozwijanej i określ zbiór prawdopodobieństw w wartości Parametrów obok siatki iteracji. Prawdopodobieństwa dla innych kategorii bądź przedziałów zostaną automatycznie odpowiednio dopasowane.

Brak iteracji. Użyj tej opcji, aby anulować iteracje dla danych wejściowych. Kliknięcie **Dalej** spowoduje usunięcie iteracji.

Korelacje

Często uważa się, że zmienne wejściowe dla symulacji są skorelowane - na przykład wysokość i waga. Korelacje między danymi wejściowymi, które zostaną zasymulowane muszą zostać wyjaśnione, aby upewnić się, że symulowane wartości zachowają te korelacje.

Ponownie przelicz korelacje podczas dopasowywania. Ta opcja określa, że korelacje między symulowanymi danymi wejściowymi są wyliczane automatycznie podczas dopasowywania rozkładów do aktywnego zbioru danych przy użyciu poleceń **Dopasuj wszystkie** lub **Dopasuj** w ustawieniach Pól symulowanych.

Nie przeliczaj ponownie korelacji podczas dopasowywania. Zaznacz tę opcję, jeśli chcesz ręcznie określić korelacje i zapobiec temu, że zostaną one zastąpione podczas automatycznego dopasowywania

rozkładów do aktywnego zbioru danych. Wartości wprowadzone w siatce Korelacje muszą należeć do zakresu od -1 do 1. Wartość 0 oznacza, że nie istnieje korelacja między powiązaną parą danych wejściowych.

Resetuj. Powoduje to ustawienie dla wszystkich korelacji wartości 0.

Użyj dopasowanej wielozdzielczej tabeli kontyngencji dla wejściowych zmiennych z rozkładem kategorialnym. W przypadku zmiennych wejściowych z rozkładem kategorialnym można automatycznie obliczać wartości w tabeli kontyngencji wielu rzędów, korzystając z aktywnego zbioru danych, który określa powiązania pomiędzy tymi zmiennymi. Tabela kontyngencji jest wówczas używana podczas generowania danych dla zmiennych wejściowych. Po wybraniu opcji zapisu planu symulacji tabela kontyngencji jest zapisywana w pliku planu i zostaje użyta po uruchomieniu danego planu.

- **Oblicz tabelę kontyngencji dla danych z aktywnego zbioru** W przypadku pracy z istniejącym planem symulacji zawierającym tabelę kontyngencji można ponownie obliczyć wartości w tej tabeli, korzystając z aktywnego zbioru danych. Wykonanie tej czynności spowoduje zastąpienie tabeli kontyngencji przez wartości z załadowanego pliku planu.
- **Użyj tabeli kontyngencji z załadowanego planu symulacji.** Domyślnie podczas ładowania planu symulacji zawierającego tabelę kontyngencji używana jest tabela z tego planu. Wartości w tabeli kontyngencji z aktywnego zbioru danych można ponownie obliczyć, zaznaczając opcję **Oblicz tabelę kontyngencji dla danych z aktywnego zbioru.**

Opcje zaawansowane

Maksymalna liczba obserwacji. Określa maksymalną liczbę obserwacji danych symulowanych oraz powiązanych wartości zmiennych przewidywanych, które mają zostać wygenerowane. Gdy określona została analiza czułości, jest to maksymalna liczba obserwacji dla każdej iteracji.

Kryteria zatrzymania dla zmiennej przewidywanej. Jeśli Twój model predykcyjny zawiera więcej niż jedną zmienną przewidywaną, możesz wtedy wybrać, do której z tych zmiennych zastosowane zostaną kryteria zatrzymywania.

Kryteria zatrzymania. Wybrane wartości określają kryteria zatrzymania symulacji potencjalnie zanim wygenerowana zostanie maksymalna dozwolona liczba obserwacji.

- **Kontynuuj do czasu osiągnięcia maksimum.** Określa, że obserwacje symulacji będą generowane do momentu osiągnięcia maksymalnej liczby obserwacji.
- **Zatrzymaj, gdy zostaną wylosowane wartości skrajne z rozkładu.** Użyj tej opcji, gdy chcesz się upewnić, czy jeden z krańców określonego rozkładu zmiennej przewidywanej został odpowiednio spróbkowany. Symulowane obserwacje będą generowane do momentu zakończenia określonego próbkowania wartości krańcowej lub do momentu osiągnięcia maksymalnej liczby obserwacji. Jeśli Twój model predykcyjny zawiera wiele zmiennych przewidywanych, używając rozwijanej listy **Zmienna przewidywana dla kryteriów zatrzymywania** zaznacz tę zmienną, do której zastosowane zostaną kryteria.

Typ. Możesz zdefiniować granice obszaru krańcowego określając wartość zmiennej przewidywanej na przykład na poziomie 10 000 000 lub wartość percentylu jako 99-ty. Jeśli z listy rozwijanej **Typ** wybierzesz **Wartość**, wprowadź wartość brzegu w polu tekstowym **Wartość** i użyj rozwijanej listy **Strona**, aby określić, czy jest to brzeg obszaru krańcowego z Lewej czy z Prawej strony. Jeśli z rozwijanej listy **Typ** wybierzesz opcję **Percentyl**, musisz wprowadzić wartość w polu tekstowym **Percentyl**.

Częstość. Określ liczbę wartości zmiennej przewidywanej, które muszą leżeć w obszarze krańcowym, aby móc upewnić się, że kraniec został odpowiednio spróbkowany. Gdy liczba ta zostanie osiągnięta, zakończy się generowanie obserwacji.

- **Zatrzymaj, gdy przedział ufności średniej jest w granicach wyznaczonych progami.** Użyj tej opcji, gdy chcesz upewnić się, że średnia podanej zmiennej przewidywanej jest znana z określonym stopniem dokładności. Symulowane obserwacje będą generowane do momentu osiągnięcia określonego stopnia dokładności lub do momentu osiągnięcia maksymalnej liczby obserwacji. Aby użyć tej opcji, należy określić poziom pewności i próg. Symulowane obserwacje będą generowane do momentu, aż przedział pewności powiązany z określonym poziomem znajdzie się w granicach progów. Na przykład: możesz użyć tej opcji do określenia, że obserwacje będą generowane do momentu, aż przedział pewności średniej na

poziomie pewności 95% znajdzie się w granicach 5% wartości średniej. Jeśli Twój model predykcyjny zawiera wiele zmiennych przewidywanych, używając rozwijanej listy **Zmienna przewidywana dla kryteriów zatrzymywania** zaznacz tę zmienną, do której zastosowane zostaną kryteria.

Typ prog. Możesz określić próg jako wartość numeryczną lub jako procent wartości średniej. Jeśli z rozwijanej listy **Typ prog** wybierzesz opcję Wartość, musisz określić próg w polu tekstowym Próg jako wartość. Jeśli z rozwijanej listy **Typ prog** wybierzesz opcję Procent, musisz wprowadzić wartość w polu tekstowym Próg jako procent.

Liczba losowanych obserwacji. Określa to liczbę obserwacji, które zostaną użyte podczas automatycznego dopasowywania rozkładów dla symulowanych danych wejściowych w aktywnym zbiorze danych. Jeśli Twój zbiór danych jest bardzo duży, być może warto rozważyć ograniczenie liczby obserwacji używanych do dopasowywania rozkładu. Jeśli wybierzesz **Ogranicz do N obserwacji**, użytych zostanie pierwszych N obserwacji.

Kryteria poprawności dopasowania (ciągłe). W przypadku danych wejściowych ciągłych do uszeregowania rozkładów podczas dopasowywania rozkładów dla symulowanych danych wejściowych do aktywnego zbioru danych można użyć testu Andersona-Darlinga lub testu Kołmogorowa-Smirnowa określających dobroć dopasowania. Domyślnie wybrany zostaje test Andersona-Darlinga i jest on polecany zwłaszcza w przypadkach, gdy chcesz zapewnić najlepsze możliwe dopasowanie w obszarach krańcowych.

Rozkład empiryczny. W przypadku danych wejściowych ciągłych rozkład Empiryczny jest funkcją skumulowanego rozkładu danych historycznych. Możesz określić liczbę kontenerów użytych do wyliczenia rozkładu Empirycznego dla ciągłych danych wejściowych. Domyślną wartością jest 100, a maksymalną 1000.

Replikacja wyników. Ustawienie wartości początkowej generatora liczb losowych umożliwia powielenie symulacji. Podaj liczbę całkowitą lub kliknij przycisk **Generuj**, co spowoduje utworzenie pseudolosowej liczby całkowitej między 1 a 2147483647, włącznie. Domyślną wartością jest 629111597.

Uwaga: Dla określonej wartości początkowej generatora liczb losowych wyniki są powtarzalne, o ile liczba wątków pozostaje niezmienną. Na konkretnym komputerze liczba wątków jest niezmienna, o ile nie zostanie zmieniona za pomocą komendy SET THREADS. Liczba wątków może się zmienić w przypadku uruchomienia symulacji na innym komputerze, ponieważ do określania liczby wątków na każdej maszynie używany jest algorytm wewnętrzny.

Braki danych użytkownika dla zmiennych wejściowych z rozkładem kategoryjnym. Te elementy pozwalają określić, czy brakujące wartości użytkownika dla danych wejściowych z rozkładem kategoryjnym są traktowane jako poprawne. Systemowe braki danych i brakujące wartości użytkownika dla wszystkich typów danych wejściowych są zawsze traktowane jako nieprawidłowe. Wszystkie zmienne wejściowe muszą mieć poprawne wartości, na wypadek gdyby miały zostać uwzględnione w dopasowaniu rozkładu, obliczeniu korelacji oraz obliczeniu wartości w opcjonalnej tabeli kontyngencji.

Funkcje gęstości

Ustawienia te pozwalają na dostosowanie danych wyjściowych dla funkcji gęstości prawdopodobieństwa i funkcji skumulowanego rozkładu dla ciągłych zmiennych przewidywanych, a także wykresów słupkowych wartości przewidywanych dla jakościowych zmiennych przewidywanych.

Funkcja gęstości prawdopodobieństwa (PDF). Funkcja gęstości prawdopodobieństwa wyświetla rozkład wartości zmiennych przewidywanych. Dla ciągłych zmiennych przewidywanych pozwala określić prawdopodobieństwo, że zmienna przewidywana znajduje się w granicach podanego obszaru. Dla jakościowych zmiennych przewidywanych (zmiennych przewidywanych z poziomem pomiaru nominalnym lub porządkowym) generowany jest wykres słupkowy, który wyświetla procent obserwacji przypadających na każdą z kategorii zmiennej przewidywanej. Dla jakościowych zmiennych przewidywanych modeli PMML dostępne są dodatkowe opcje z wartościami Kategorii do opisanych poniżej ustawień raportowania.

Dla modeli dwustopniowego skupienia oraz modeli skupień metodą k-średnich tworzony jest wykres słupkowy przynależności do grupy.

Skumulowana funkcja gęstości (CDF). Skumulowana funkcja gęstości wyświetla prawdopodobieństwo, że wartość zmiennej przewidywanej jest mniejsza lub równa określonej wartości. Jest ona dostępna tylko dla ciągłych zmiennych przewidywanych.

Pozycje suwaków. Można określić pozycje początkowe przesuwalnych linii odniesienia na wykresach PDF i CDF. Wartości podane dla dolnej i górnej linii odnoszą się do pozycji wzdłuż osi poziomej, nie są to percentyle. Można usunąć dolną linię, wybierając **-Nieskończoność** lub górną linię, wybierając **Nieskończoność**. Domyślnie linie są ustawione na 5 i 95 percentylach. Jeśli na jednym wykresie wyświetlanych jest wiele funkcji rozkładu (z uwagi na wiele zmiennych przewidywanych lub wyników pochodzących z iteracji analizy czułości), wartości domyślne odnoszą się do rozkładu dla pierwszej iteracji lub pierwszej zmiennej przewidywanej.

Linie referencyjne (dane ciągłe). Możesz zażądać różnych pionowych linii odniesienia, które powinny zostać dodane do funkcji gęstości prawdopodobieństwa i funkcji skumulowanego rozkładu dla ciągłych zmiennych przewidywanych.

- **Sigmy.** Możesz dodać linie odniesienia plus/minus określona liczba odchyłeń standardowych od średniej zmiennej przewidywanej.
- **Percentyle.** Możesz dodać linie odniesienia dla jednej lub dwóch wartości percentyli rozkładu dla zmiennej docelowej, wprowadzając wartości w polach tekstowych Dolny i Górny. Na przykład: wartość 95 w Górnym polu tekstowym reprezentuje 95-ty percentyl, co jest wartością, poniżej której znajduje się 95% obserwacji. Tak samo wartość 5 w Dolnym polu tekstowym reprezentuje 5-ty percentyl, co jest wartością, powyżej której znajduje się 5% obserwacji.
- **Niestandardowe linie referencyjne.** Możesz dodać linie odniesienia o określonych wartościach zmiennej przewidywanej.

Uwaga: Jeśli na jednym wykresie wyświetlanych jest wiele funkcji rozkładu (z uwagi na wiele zmiennych przewidywanych lub wyników pochodzących z iteracji analizy czułości), linie odniesienia mają zastosowanie tylko do rozkładu dla pierwszej iteracji lub pierwszej zmiennej przewidywanej. Można dodać linie odniesienia do pozostałych rozkładów, korzystając z okna dialogowego Opcje, dostępnego w obszarze wykresu PDF lub CDF.

Nakładaj wyniki pochodzące z oddzielnych ciągłych przewidywanych zmiennych. W przypadku wielu ciągłych zmiennych przewidywanych, opcja ta określa, czy funkcje rozkładu dla tych wszystkich zmiennych przewidywanych będą wyświetlane na jednym rysunku z jednym wykresem funkcji gęstości prawdopodobieństwa i drugim wykresem funkcji skumulowanego rozkładu. Gdy opcja ta nie jest zaznaczona, wyniki dla każdej ze zmiennych przewidywanych będą wyświetlane na osobnym wykresie.

Wartości kategorii do zaraportowania. W przypadku modeli PMML z jakościowymi zmiennymi przewidywanymi, wynik dla modelu jest zbiorem prawdopodobieństw - po jednym dla każdej kategorii w taki sposób, że wartość zmiennej przewidywanej pojawia się w każdej kategorii. Kategoria o najwyższym prawdopodobieństwie jest brana jako kategoria przewidywana i używa się jej do generowania wykresu słupkowego opisanego dla powyższego ustawienia **Funkcji gęstości prawdopodobieństwa**. Wybór **Przewidywanej kategorii** wygeneruje wykres słupkowy. Wybór **Przewidywanych prawdopodobieństw** wygeneruje histogramy rozkładu prawdopodobieństw przewidywanych dla każdej z kategorii zmiennych przewidywanych.

Grupowanie w analizie czułości. Symulacje zawierające analizę czułości generują niezależny zbiór przewidywanych wartości zmiennej przewidywanej dla każdej iteracji zdefiniowany przez analizę (jedna iteracja dla każdej wartości danych wejściowych, która ulega zmianie). Gdy istnieją iteracje, wykres słupkowy kategorii przewidywanej dla jakościowej zmiennej przewidywanej jest wyświetlany jako zgrupowany wykres słupkowy, zawierający wyniki dla wszystkich iteracji. Możesz zdecydować się na zgrupowanie kategorii razem lub zgrupowanie razem iteracji.

Dane wyjściowe

Wykresy tornado. Wykresy tornado są wykresami słupkowymi wyświetlającymi relacje między zmiennymi przewidywanymi i symulowanymi danymi wejściowymi przy użyciu wielu różnych miar.

- **Korelacja przewidywanej i danych wejściowych.** Ta opcja tworzy wykresy tornado współczynników korelacji między podaną zmienną przewidywaną i każdą z jej symulowanych danych wejściowych.

Ten typ wykresu tornado nie obsługuje zmiennych przewidywanych o nominalnym lub porządkowym poziomie pomiaru lub symulowanych danych wejściowych z rozkładem kategoryjnym.

- **Udział w wariancji.** Ta opcja tworzy wykres tornado wyświetlający udział w wariancji zmiennej przewidywanej pochodzący od każdej z symulowanych danych wejściowych, pozwalając na oszacowanie, w jakim stopniu każda z danych wejściowych wpływa na całkowitą niepewność zmiennych przewidywanych. Ten typ wykresu tornado nie obsługuje zmiennych przewidywanych z porządkowymi lub nominalnymi poziomami pomiarów lub symulowanych danych wejściowych o następujących rozkładach: jakościowy, Bernoulliego, dwumianowy, Poissona lub ujemny dwumianowy.
- **Czułość na zmiany przewidywanej.** Opcja ta tworzy wykres tornado, który wyświetla wpływ na wartość przewidywaną modulacji każdego symulowanego wejścia przez dodanie lub odjęcie określonej liczby odchylenia standardowego rozkładu powiązanego z wejściem. Ten typ wykresu tornado nie obsługuje zmiennych przewidywanych z porządkowymi lub nominalnymi poziomami pomiarów lub symulowanych danych wejściowych o następujących rozkładach: jakościowy, Bernoulliego, dwumianowy, Poissona lub ujemny dwumianowy.

Wykresy skrzynkowe rozkładów przewidywanych. Wykresy skrzynkowe są dostępne dla ciągłych zmiennych przewidywanych. Zaznacz **Nakładaj wyniki pochodzące od osobnych przewidywanych**, jeśli Twój model predykcyjny posiada kilka ciągłych zmiennych przewidywanych, a chcesz wyświetlić wykresy skrzynkowe dla wszystkich zmiennych przewidywanych na jednym rysunku.

Wykresy rozrzucone przewidywanych w stosunku do danych. Wykresy rozrzutu zmiennych przewidywanych w stosunku do symulowanych danych wejściowych są dostępne zarówno dla ciągłych, jak i jakościowych zmiennych przewidywanych i zawierają rozrzuty zmiennych przewidywanych zarówno dla ciągłych, jak i jakościowych danych wejściowych. Rozrzuty obejmujące jakościowe zmienne przewidywane lub jakościowe zmienne wejściowe również są wyświetlane w postaci mapy natężeń.

Stwórz tabelę wartości percentyli. W przypadku ciągłych zmiennych przewidywanych możesz otrzymać tabelę zbudowaną z określonych percentyli rozkładu zmiennych przewidywanych. Kwartyłe (percentyle 25., 50. i 75.) dzielą obserwacje na cztery grupy jednakowej wielkości. Jeśli istnieje konieczność podzielenia obserwacji na liczbę równych grup różną od czterech, należy wybrać opcję **Przedziały** i określić liczbę. Zaznacz **Percentyle niestandardowe**, aby określić pojedyncze percentyle, na przykład 99-ty percentyl.

Statystyki opisowe rozkładów przewidywanych. Opcja ta tworzy tabele ze statystykami opisowymi dla ciągłych i jakościowych zmiennych przewidywanych oraz dla ciągłych danych wejściowych. W przypadku ciągłych zmiennych przewidywanych tabela zawiera średnią, odchylenie standardowe, medianę, minimum oraz maksimum, przedział pewności średniej na określonym poziomie oraz 5-ty i 95-ty percentyl rozkładu zmiennej przewidywanej. Dla jakościowych zmiennych przewidywanych tabela zawiera procent obserwacji, odpowiadającej każdej z kategorii zmiennej przewidywanej. Dla jakościowych zmiennych przewidywanych modeli PMML tabela zawiera także średnie prawdopodobieństwo wystąpienia każdej z kategorii zmiennej przewidywanej. Dla ciągłych danych wejściowych tabela zawiera średnią, odchylenie standardowe, minimum oraz maksimum.

Korelacje i tabela kontyngencji dla danych wejściowych. Ta opcja umożliwi wyświetlenie tabeli współczynników korelacji pomiędzy symulowanymi danymi wejściowymi. W przypadku generowania z tabeli kontyngencji danych wejściowych z rozkładem kategoryjnym wyświetlana jest również tabela kontyngencji danych wygenerowanych na podstawie tych danych wejściowych.

Symulowane dane wejściowe, które zostaną zawarte w raportach. Domyślnie w danych wyjściowych zawarte zostają wszystkie symulowane zmienne wejściowe. Możesz wykluczyć wybrane zmienne wejściowe z wyniku. Spowoduje to wykluczenie ich z wykresów tornado, wykresów rozrzuconych i wyników w tabelach.

Ograniczenie przedziałów dla przewidywanych zmiennych ilościowych. Możesz określić zakres poprawnych wartości dla co najmniej jednej ciągłej zmiennej przewidywanej. Wartości spoza określonego zakresu nie będą uwzględniane w wynikach i analizach powiązanych ze zmiennymi przewidywanymi. Aby ustawić dolną granicę, wybierz wartość **Dolny** w kolumnie Ograniczenie i wprowadź wartość w kolumnie Minimum. Aby ustawić górną granicę, wybierz wartość **Górny** w kolumnie Ograniczenie i wprowadź wartość w kolumnie Maksimum. Aby ustawić górną i dolną granicę, wybierz wartość **Łącznie** w kolumnie Ograniczenie i wprowadź wartości w kolumnach Minimum i Maksimum.

Formaty. Możesz ustawić formaty używane podczas wyświetlania wartości zmiennych przewidywanych i danych wejściowych (zarówno stałych, jak i symulowanych danych wejściowych).

Zapisz

Zapisz plan dla tej symulacji. Możesz zapisać bieżące specyfikacje dla swojej symulacji w pliku planu symulacji. Pliki planu symulacji mają rozszerzenie *.splan*. Możesz otworzyć ponownie plan w kreatorze symulacji, opcjonalnie dokonać modyfikacji i uruchomić symulację. Możesz udostępnić plan symulacji innym użytkownikom, którzy następnie mogą uruchomić go w oknie dialogowym Uruchom symulację. Plany symulacji zawierają wszystkie specyfikacje, poza poniższym: ustawieniami dla Funkcji gęstości, ustawieniami wyniku dla wykresów i tabeli, ustawieniami Opcji zaawansowanych dla Dopasowywania, Rozkładu empirycznego i Wartości początkowej generatora liczb losowych.

Zapisz symulowane dane w nowym pliku danych. Możesz zapisać symulowane zmienne wejściowe, stałe zmienne wejściowe i przewidywane wartości zmiennych przewidywanych w pliku danych SPSS Statistics, nowym zbiorze danych w aktualnej sesji lub pliku Excel. Każda obserwacja (lub wiersz) pliku danych składa się z przewidywanych wartości zmiennych przewidywanych razem z symulowanymi danymi wejściowymi i stałymi danymi wejściowymi generującymi wartości zmiennych przewidywanych. Gdy określona jest analiza czułości, każda iteracja powiększa zgrupowany zbiór obserwacji oznaczonych etykietą z numerem iteracji.

Okno dialogowe Uruchom symulację

Okno dialogowe Uruchom symulację zostało zaprojektowane z myślą o użytkownikach posiadających plan symulacji i przede wszystkim chcą uruchomić symulację. Zapewnia ono także funkcje, które potrzebujesz do uruchomienia symulacji w różnych warunkach. Pozwala na przeprowadzanie następujących zadań ogólnych:

- Skonfigurować lub zmodyfikować analizę czułości, aby zbadać wpływ zmieniającej się wartości ustalonych danych wejściowych lub zmieniającego się parametru rozkładu dla symulowanych danych wejściowych.
- Dopasować ponownie rozkłady prawdopodobieństwa dla niepewnych danych wejściowych (oraz korelacji między tymi danymi wejściowymi) do nowych danych.
- Modyfikować rozkład dla symulowanych danych wejściowych.
- Dopasowywać dane wyjściowe.
- Uruchomić symulację.

Karta Symulacji

Karta Symulacji pozwala na określenie analizy czułości, ponowne dopasowanie rozkładów prawdopodobieństwa dla symulowanych danych wejściowych, a także korelacji między symulowanymi danymi wejściowymi do nowych danych oraz na modyfikowanie rozkładu prawdopodobieństwa powiązanego z symulowanymi danymi wejściowymi.

Siatka symulowane zmienne wejściowe zawiera wpis dla każdej zmiennej wejściowej zdefiniowanej w planie symulacji. Każdy wpis wyświetla nazwę danych wejściowych i typ rozkładu prawdopodobieństwa powiązany z tym wejściem, razem z próbnym wykresem krzywej skojarzonego rozkładu. Każda z danych wejściowych posiada także powiązaną ikonę stanu (kolorowy okrąg ze znakiem zaznaczenia), który przydaje się podczas ponownego dopasowywania rozkładów do nowych danych. Dodatkowo, wejścia mogą zawierać ikonę blokowania, wskazującą, że dana wejściowa jest zablokowana i nie może być modyfikowana ani ponownie dopasowywana do nowych danych w oknie dialogowym Uruchom symulację. Aby zmodyfikować zablokowane wejście, będziesz musiał otworzyć plan symulacji w kreatorze symulacji.

Każde z wejść jest symulowane bądź stałe. Symulowane zmienne wejściowe to takie, których wartości są niepewne i zostaną wygenerowane z rysunku określonego rozkładu prawdopodobieństwa. Stałe wartości wejściowe to takie, których wartości są znane i pozostają stałe dla każdego przypadku wygenerowanego na drodze symulacji. Aby pracować z konkretną daną wejściową, zaznacz w siatce Symulowanych danych wejściowych wpis dla wejścia.

Określanie analizy czułości

Analiza czułości pozwala na sprawdzenie skutków systematycznych zmian ustalonych danych wejściowych lub parametru rozkładu na Symulowane zmienne wejściowe przez generowanie niezależnego zbioru symulowanych obserwacji (czyli osobnej symulacji) dla każdej z określonych wartości. Aby określić analizę czułości, zaznacz stałe lub symulowane zmienne wejściowe i kliknij **Analiza czułości**. Analiza czułości jest ograniczona do pojedynczej statych danych wejściowych lub pojedynczego parametru rozkładu dla symulowanych danych wejściowych. Więcej informacji można znaleźć w temacie [“Analiza czułości”](#) na stronie 258.

Ponowne dopasowanie rozkładów do nowych danych

Aby automatycznie ponownie dopasować rozkład prawdopodobieństwa dla symulowanych danych wejściowych (oraz korelacji między symulowanymi wejściami) do danych w aktywnym zbiorze danych:

1. Sprawdź, czy każde z wejść modelu jest dopasowane do odpowiedniego pola w aktywnym zbiorze danych. Każda symulowana dana wejściowa jest dopasowana do pola w aktywnym zbiorze danych określonym w rozwijanej liście **Pole** powiązanej z tym wejściem. Możesz w łatwy sposób zidentyfikować zmienne wejściowe, które nie zostały dopasowane, szukając danych wejściowych z ikoną stanu zawierającą znak zaznaczenia ze znakiem zapytania, jak pokazano poniżej.



2. Zmodyfikuj wszystkie potrzebne dopasowania pól, zaznaczając opcję **Dopasuj pole w zbiorze danych** i wybierając pole z listy.
3. Kliknij przycisk **Dopasuj wszystkie**.

Dla każdego dopasowanego wejścia wyświetlany jest najlepiej pasujący rozkład razem z wykresem rozkładu nałożonym na histogram (lub wykres słupkowy) danych historycznych. Jeśli nie można znaleźć dopuszczalnego dopasowania, używany jest rozkład Empiryczny. Dla danych wejściowych, które zostały dopasowane do rozkładu Empirycznego, zobaczysz tylko histogram danych historycznych, ponieważ rozkład Empiryczny jest w rzeczywistości reprezentowany przez ten histogram.

Uwaga: pełna lista ikon stanu znajduje się w temacie [“Symulowane zmienne”](#) na stronie 255.

Modyfikowanie rozkładów prawdopodobieństwa

Możesz zmodyfikować rozkład prawdopodobieństwa dla symulowanych danych wejściowych i opcjonalnie zmienić symulowaną daną wejściową na stałą daną wejściową i na odwrót.

1. Wybierz zmienne wejściowe i zaznacz opcję **Ustaw rozkład ręcznie**.
2. Wybierz typ rozkładu i określ parametry rozkładu. Aby zmienić symulowane zmienne wejściowe na stałe zmienne wejściowe, zaznacz **Stać** w rozwijanej liście **Typ**.

Po wprowadzeniu parametrów dla rozkładu próbny wykres rozkładu (wyświetlany we wpisie dla danych wejściowych) zostanie zaktualizowany, aby odzwierciedlić wprowadzone zmiany. Więcej informacji na temat ręcznego określania rozkładów prawdopodobieństwa można znaleźć w temacie [“Symulowane zmienne”](#) na stronie 255.

Podczas dopasowania uwzględnij braki danych użytkownika dla wejściowych zmiennych jakościowych. Wybór tej opcji określa, czy brakujące wartości użytkownika dla danych wejściowych z rozkładem kategoryjnym są traktowane jako poprawne podczas ponownego dopasowywania do danych w aktywnym zbiorze danych. Systemowe braki danych i brakujące wartości użytkownika dla wszystkich typów danych wejściowych są zawsze traktowane jako nieprawidłowe. Wszystkie zmienne wejściowe muszą mieć poprawne wartości, na wypadek gdyby miały zostać uwzględnione w dopasowaniu rozkładu lub obliczeniu korelacji.

Karta Raport

Karta Raport pozwala na dostosowanie danych wyjściowych wygenerowanych na drodze symulacji.

Funkcje gęstości. Funkcje gęstości są głównymi średnimi sondowania zbioru wyników Twojej symulacji.

- **Funkcja gęstości prawdopodobieństwa.** Funkcja gęstości prawdopodobieństwa wyświetla rozkład wartości przewidywanych, pozwalając na określenie prawdopodobieństwa tego, że zmienna przewidywana znajduje się na podanym obszarze. Dla zmiennych przewidywanych o stałym zbiorze wyników ("zła obsługa", "średnia obsługa", "dobra obsługa" i "doskonała obsługa") generowany jest wykres słupkowy wyświetlający procent obserwacji przypadających dla każdej z kategorii zmiennej przewidywanej.
- **Skumulowana funkcja gęstości.** Skumulowana funkcja gęstości wyświetla prawdopodobieństwo, że wartość zmiennej przewidywanej jest mniejsza lub równa określonej wartości.

Wykresy tornado. Wykresy tornado są wykresami słupkowymi wyświetlającymi relacje między zmiennymi przewidywanymi i symulowanymi danymi wejściowymi przy użyciu wielu różnych miar.

- **Korelacja przewidywanej i danych wejściowych.** Ta opcja tworzy wykresy tornado współczynników korelacji między podaną zmienną przewidywaną i każdą z jej symulowanych danych wejściowych.
- **Udział w wariancji.** Ta opcja tworzy wykres tornado wyświetlający udział w wariancji zmiennej przewidywanej pochodzący od każdej z symulowanych danych wejściowych, pozwalając na oszacowanie, w jakim stopniu każda z danych wejściowych wpływa na całkowitą niepewność zmiennych przewidywanych.
- **Czułość na zmiany przewidywanej.** Opcja ta tworzy wykres tornado, który wyświetla wpływ na wartość przewidywaną modulacji każdego symulowanego wejścia przez dodanie lub odjęcie odchylenia standardowego rozkładu powiązanego z wejściem.

Wykresy rozrzucone przewidywanych w stosunku do danych. Opcja ta generuje wykresy rozrzutu zmiennych przewidywanych w stosunku do symulowanych danych wejściowych.

Wykresy skrzynkowe rozkładów przewidywanych. Opcja ta generuje wykresy skrzynkowe rozkładów prawdopodobieństwa.

Tabela kwartyli. Opcja ta generuje tabelę kwartyli rozkładów prawdopodobieństwa. Kwartyle rozkładu to 25-ty, 50-ty i 75-ty percentyl rozkładu i dzielią one obserwacje na cztery grupy jednakowej wielkości.

Korelacje i tabela kontyngencji dla danych wejściowych. Ta opcja umożliwia wyświetlenie tabeli współczynników korelacji pomiędzy symulowanymi danymi wejściowymi. Tabela kontyngencji powiązań pomiędzy danymi wejściowymi z rozkładem kategorialnym jest wyświetlana, jeśli plan symulacji określa generowanie danych kategorialnych na podstawie tabeli kontyngencji.

Nakładaj wyniki pochodzące od osobnych przewidywanych. Jeśli model predykcyjny, który symulujesz, zawiera kilka zmiennych przewidywanych, możesz określić, czy wyniki pochodzące od oddzielnych zmiennych przewidywanych są wyświetlane na pojedynczym wykresie. Ustawienie to dotyczy wykresów funkcji gęstości prawdopodobieństwa, funkcji skumulowanego rozkładu i wykresy skrzynkowe. Na przykład: jeśli zaznaczysz tę opcję, to funkcje gęstości prawdopodobieństwa dla wszystkich zmiennych przewidywanych zostaną wyświetlone na jednym wykresie.

Zapisz plan dla tej symulacji. Możesz zapisać wszystkie modyfikacje dla swojej symulacji w pliku planu symulacji. Pliki planu symulacji mają rozszerzenie *.splan*. Możesz ponownie otworzyć plan w oknie dialogowym Uruchom symulację lub w kreatorze symulacji. Plany symulacji zawierają wszystkie specyfikacje, poza ustawieniami wyników.

Zapisz symulowane dane w nowym pliku danych. Możesz zapisać symulowane zmienne wejściowe, stałe zmienne wejściowe i przewidywane wartości zmiennych przewidywanych w pliku danych SPSS Statistics, nowym zbiorze danych w aktualnej sesji lub pliku Excel. Każda obserwacja (lub wiersz) pliku danych składa się z przewidywanych wartości zmiennych przewidywanych razem z symulowanymi danymi wejściowymi i stałymi danymi wejściowymi generującymi wartości zmiennych przewidywanych. Gdy określona jest analiza czułości, każda iteracja powiększa zgrupowany zbiór obserwacji oznaczonych etykietą z numerem iteracji.

Jeśli potrzebujesz większego stopnia dopasowania wyniku, niż jest tu dostępny, rozważ uruchomienie symulacji w kreatorze symulacji. Więcej informacji można znaleźć w temacie: ["Uruchamianie symulacji z wykorzystaniem planu symulacji"](#) na stronie 251.

Praca z wynikami Symulacji w formie wykresu

Wiele z wykresów wygenerowanych przez symulację posiada interaktywne funkcje pozwalające Ci na dostosowanie wyświetlania. Interaktywne funkcje są dostępne po aktywowaniu (dwukrotnym kliknięciu) obiektu wykresu w przeglądarce wyników. Wszystkie wykresy symulacji są wizualizacjami danych. .

Wykresy funkcji gęstości prawdopodobieństwa dla ciągłych zmiennych przewidywanych. Wykres ten posiada dwie przesuwające się pionowe linie odniesienia dzielące wykres na dwa osobne obszary. Tabela poniżej wykresu przedstawia prawdopodobieństwo, że zmienna przewidywana znajduje się w każdym z obszarów. Jeśli na tym samym wykresie wyświetlanych jest kilka funkcji gęstości, tabela posiada oddzielne wiersze dla prawdopodobieństw powiązanych z każdą z funkcji gęstości. Każda linia odniesienia posiada suwak (odwrócony trójkąt) umożliwiający łatwe przesuwanie linii. Po kliknięciu przycisku **Opcje wykresu** znajdującego się na wykresie pojawia się wiele dostępnych funkcji. W szczególności, można wyraźnie ustawić pozycje suwaków, dodać stałe linie odniesienia i zmienić widok wykresu z krzywej ciągłej na histogram lub na odwrót. Aby uzyskać dodatkowe informacje, patrz temat: [“Opcje wykresu ” na stronie 266.](#)

Wykresy funkcji skumulowanego rozkładu dla ciągłych zmiennych przewidywanych. Wykres ten posiada takie same dwie przesuwalne pionowe linie odniesienia oraz powiązaną tabelę, które zostały opisane powyżej dla przypadku wykresu funkcji gęstości prawdopodobieństwa. Zapewnia też dostęp do okna dialogowego Opcje wykresu, które umożliwi na bezpośrednie ustawienie pozycji suwaków, dodanie stałych linii odniesienia i na określenie, czy funkcja skumulowanego rozkładu jest wyświetlana jako funkcja rosnąca (domyślnie) czy malejąca. Aby uzyskać dodatkowe informacje, patrz temat: [“Opcje wykresu ” na stronie 266.](#)

Wykresy słupkowe dla jakościowych zmiennych przewidywanych z iteracjami analizy czułości. Dla jakościowych zmiennych przewidywanych z iteracjami analizy czułości wyniki dla przewidywanej kategorii zmiennej przewidywanej są wyświetlane jako zgrupowany wykres słupkowy zawierający wyniki dla wszystkich iteracji. Wykres zawiera rozwijaną listę pozwalającą na grupowanie ze względu na kategorię lub iterację. Dla modeli dwustopniowego skupienia oraz modeli skupień metodą k-średnich możesz wybrać grupowanie ze względu na numer grupy lub iterację.

Wykresy skrzynkowe dla kilku zmiennych przewidywanych z iteracjami analizy czułości. Dla modeli predykcyjnych z kilkoma ciągłymi zmiennymi przewidywanymi i iteracjami analizy czułości, wybór wyświetlania wykresów skrzynkowych dla wszystkich zmiennych przewidywanych na jednym wykresie spowoduje wyświetlenie zgrupowanego wykresu skrzynkowego. Wykres zawiera rozwijaną listę pozwalającą na grupowanie ze względu na zmienną przewidywaną lub iterację.

Opcje wykresu

Okno dialogowe Opcje wykresu pozwala na dostosowanie wyświetlania włączonych wykresów funkcji gęstości prawdopodobieństwa i funkcji skumulowanego rozkładu wygenerowanych za pomocą symulacji.

Widok. Rozwijana lista **Widok** stosuje się tylko do wykresu funkcji gęstości prawdopodobieństwa. Pozwala na przełączanie widoku wykresu między krzywą ciągłą a histogramem. Funkcja ta nie jest dostępna, gdy na tym samym wykresie wyświetlanych jest kilka funkcji gęstości. W tym przypadku funkcje gęstości można przeglądać tylko jako krzywe ciągłe.

Porządek. Rozwijana lista **Porządek** stosuje się tylko do wykresu funkcji skumulowanego rozkładu. Określa ona, czy funkcja skumulowanego rozkładu jest wyświetlana jako funkcja rosnąca (domyślnie) czy malejąca. Gdy wyświetlana jest ona jako funkcja malejąca, wartość funkcji w danym punkcie na osi poziomej jest prawdopodobieństwem, że zmienna przewidywana znajduje się na prawo od tego punktu.

Pozycje suwaków. Możesz bezpośrednio ustawić pozycję przesuwanych linii odniesienia wprowadzając odpowiednią wartość w polach tekstowych Górny i Dolny. Możesz usunąć linie znajdującą się z lewej strony wybierając **-Nieskończoność**, praktycznie ustawiając jej pozycję na minus nieskończoność, możesz też usunąć prawą linię, wybierając **Nieskończoność**, ustawiając jej pozycję na poziomie nieskończoność

Linie referencyjne. Do funkcji gęstości prawdopodobieństwa i funkcji skumulowanego rozkładu możesz dodać wiele stałych pionowych linii odniesienia. Jeśli na jednym wykresie wyświetlanych jest wiele funkcji (z uwagi na wiele zmiennych przewidywanych lub wyników pochodzących z iteracji analizy czułości), można określić konkretne funkcje, dla których linie odniesienia są stosowane.

- **Sigmy.** Możesz dodać linie odniesienia plus/minus określona liczba odchyłeń standardowych od średniej zmiennej przewidywanej.
- **Percentyle.** Możesz dodać linie odniesienia dla jednej lub dwóch wartości percentyli rozkładu dla zmiennej docelowej, wprowadzając wartości w polach tekstowych Dolny i Górny. Na przykład: wartość 95 w Górnym polu tekstowym reprezentuje 95-ty percentyl, co jest wartością, poniżej której znajduje się 95% obserwacji. Tak samo wartość 5 w Dolnym polu tekstowym reprezentuje 5-ty percentyl, co jest wartością, powyżej której znajduje się 5% obserwacji.
- **Pozycje niestandardowe.** Możesz dodać linie odniesienia o określonych wartościach wzdłuż osi poziomej.

Linie referencyjne etykiet. Ta opcja określa, czy dla wybranych linii odniesienia są stosowane etykiety.

Linie referencyjne są usuwane przez usunięcie zaznaczenia powiązanego z linią wyboru w oknie dialogowym Opcje wykresu i kliknięcie **Dalej**.

Modelowanie geoprzestrzenne

Techniki modelowania geoprzestrzennego służą do wykrywania wzorów w danych zawierających składnik geoprzestrzenny (mapę). Kreator modelowania geoprzestrzennego oferuje metody analizowania danych geoprzestrzennych ze składnikiem czasu lub bez niego.

Znajdź skojarzenia w oparciu o dane zdarzenia i dane geoprzestrzenne (reguły związków geoprzestrzennych)

Za pomocą reguł skojarzeń geoprzestrzennych można znaleźć wzory w danych na bazie zarówno właściwości przestrzennych, jak i innych niż przestrzenne. Na przykład można wykryć wzory w danych dotyczących przestępczości, związując dane o lokalizacji z atrybutami demograficznymi. Następnie na podstawie tych wzorów można zbudować reguły pozwalające określić miejsce wystąpienia niektórych typów przestępstw.

Twórz predykcje z wykorzystaniem szeregu czasowego i danych geoprzestrzennych (predykcja przestrzenno-czasowa)

Predykcja przestrzenno-czasowa wykorzystuje dane zawierające informacje o lokalizacji, zmienne wejściowe predykcji (predyktory), jedną lub wiele zmiennych czasu i zmienną docelową. Każda lokalizacja ma wiele wierszy danych przedstawiających wartości każdego predyktora i zmiennej docelowej w poszczególnych przedziałach czasu.

Używanie kreatora modelowania geoprzestrzennego

1. Wybierz z menu następującą opcję:

Analiza > Modelowanie przestrzenne i temporalne > Modelowania przestrzenne

2. Postępuj zgodnie ze poleceniami kreatora.

Przykłady

Szczegółowe przykłady są dostępne w systemie pomocy.

- Reguły asocjacji geoprzestrzennych: **Pomoc > Tematy > Studia przypadków > Statistics Base > Reguły asocjacji przestrzennych**
- Predykcja przestrzenna szeregów czasowych: **Pomoc > Tematy > Studia przypadków > Statistics Base > Predykcja przestrzenna szeregów czasowych**

Wybieranie map

W modelowaniu geoprzestrzennym źródłami danych może być dowolna liczba map. Źródła danych oparte na mapach używają informacji określających obszary geograficzne i inne cechy geograficzne, np. drogi lub rzeki. Wiele map zawiera także dane geograficzne lub opisowe oraz dane o zdarzeniach, np. raporty dotyczące przestępczości lub wskaźniki bezrobocia. Istnieje możliwość użycia uprzednio zdefiniowanego pliku specyfikacji mapy lub zdefiniowania specyfikacji w tym miejscu i zapisaniu ich na przyszłość.

Wczytaj specyfikację mapy

Wczytuje plik z uprzednio zdefiniowaną specyfikacją mapy (.mplan). Źródła danych mapy zdefiniowane w tym miejscu mogą być zapisane w pliku specyfikacji mapy. Gdy przy predykcji przestrzenno-czasowej zostanie wybrany plik specyfikacji mapy wskazujący na więcej niż jedną mapę, pojawi się monit o wybranie mapy z pliku.

Dodaj plik mapy

Dodaj plik kształtu ESRI (.shp) lub archiwum .zip zawierające plik kształtu ESRI.

- W lokalizacji z plikiem .shp musi istnieć odpowiedni plik .dbf o takim samym trzonie nazwy.
- Jeśli plik jest archiwum .zip, pliki .shp i .dbf muszą mieć ten sam trzon nazwy, co archiwum .zip.
- Jeśli nie istnieje odpowiedni plik odwzorowania (.prj), pojawi się monit o wybranie układu odwzorowania.

Związek

Przy regułach skojarzeń geoprzestrzennych ta kolumna definiuje powiązanie zdarzeń z właściwościami mapy. To ustawienie jest niedostępne przy predykcji przestrzennej i szeregów czasowych.

Przesuń w górę, Przesuń w dół

Porządek warstw elementów mapy jest określony kolejnością ich pojawiania się na liście. Pierwsza mapa na liście jest dolną warstwą.

Wybieranie mapy

Gdy przy predykcji przestrzenno-czasowej zostanie wybrany plik specyfikacji mapy wskazujący na więcej niż jedną mapę, pojawi się monit o wybranie mapy z pliku. Predykcja przestrzenno-czasowa nie obsługuje wielu map.

Relacja geoprzestrzenna

Przy regułach skojarzeń geoprzestrzennych okno dialogowe Relacja geoprzestrzenna definiuje powiązanie zdarzeń z właściwościami mapy.

- To ustawienie ma zastosowanie tylko do reguł skojarzeń geoprzestrzennych.
- To ustawienie wpływa wyłącznie na źródła danych powiązane z mapami, które w kroku wyboru źródła danych określone zostały jako dane kontekstowe.

Związek

W pobliżu

Zdarzenie wystąpiło blisko określonego punktu na mapie.

Jest wewnątrz

Zdarzenie występuje w określonym obszarze na mapie.

Zawiera

Obszar zdarzenia zawiera obiekt kontekstu mapy.

Na przecięciu

Lokalizacje, w których linie lub regiony z różnych map przecinają się.

Krzyżowanie

Przy wielu mapach są to lokalizacje, w których linie (dróg, rzek, linii kolejowych itp.) z różnych map przecinają się.

Na północ od, Na południe od, Na wschód od, Na zachód od

To zdarzenie wystąpiło w określonym kierunku od określonego punktu na mapie.

Ustawianie układu współrzędnych

Jeśli nie istnieje plik odwzorowania (.prj) z mapą lub po zdefiniowaniu dwóch zmiennych ze źródła danych jako współrzędnych, należy ustawić układ współrzędnych.

Domyślny geograficzny (długość i szerokość geograficzna)

Układ współrzędnych jest określany przez długość i szerokość geograficzną.

Prosty kartezjański (X i Y)

Układ współrzędnych to proste współrzędne X i Y.

Użyj dobrze znanego identyfikatora (WKID)

„Dobrze znany identyfikator” dla popularnych odwzorowań.

Użyj nazwy układu współrzędnych

Układ współrzędnych jest oparty na nazwanym odwzorowaniu. Nazwa jest ujęta w nawiasach.

Ustawianie odwzorowania

Jeśli nie jest możliwe określenie układu odwzorowania na podstawie informacji dostarczonych razem z mapą, należy jawnie określić układ odwzorowania. Najczęstszą przyczyną takiej sytuacji jest brak pliku odwzorowania (.prj) skojarzonego z mapą lub brak możliwości użycia tego pliku.

- **Miejscowość, region lub kraj (Merkatora)**
- **Duży kraj, kilka krajów lub kontynenty (Winkel Tripel)**
- **Obszar położony blisko równika (Merkatora)**
- **Obszar położony blisko jednego z biegunów (Stereograficzne)**

Odwzorowanie Merkatora jest typowym odwzorowaniem stosowanym na wielu mapach. W odwzorowaniu tym kula ziemską przedstawiona jest jako walec rozwinięty na płaskiej powierzchni. Odwzorowanie Merkatora zniekształca rozmiar i kształt większych obiektów. Zniekształcenie to zwiększa się w miarę oddalania się od równika i zbliżania do biegunów. W odwzorowaniach Winkel Tripel i stereograficznym stosowane są korekty uwzględniające fakt, że mapa przedstawia fragment powierzchni trójwymiarowej sfery na powierzchni dwuwymiarowej.

Odwzorowanie i układ współrzędnych

Po wybraniu kilku map z różnymi odwzorowaniami i układami współrzędnych należy wybrać mapę z tym systemem odwzorowania, który ma być używany. Zostanie on zastosowany do wszystkich map podlegających łączeniu w ramach określania wyniku.

Źródła danych

Źródłem danych może być plik dBase podany z plikiem kształtu, plik danych produktu IBM SPSS Statistics lub otwarty zbiór danych z bieżącej sesji.

Dane kontekstowe. Dane kontekstowe identyfikują elementy na mapie. Mogą również zawierać zmienne używane jako zmienne wejściowe dla modelu. Aby można było użyć kontekstowego pliku dBase (.dbf) powiązanego z plikiem kształtu mapy (.shp), kontekstowy plik dBase musi mieć tę samą lokalizację i ten sam trzon nazwy, co plik kształtu. Na przykład, jeśli plik kształtu ma nazwę geodata .shp, plik dBase musi mieć nazwę geodata .dbf.

Dane zdarzenia. Dane zdarzenia zawierają informacje o zdarzeniach, do których doszło, takich jak przestępstwa lub wypadki. Ta opcja jest dostępna wyłącznie dla reguł skojarzeń geoprzestrzennych.

Gęstość punktu. Przedział czasu i współrzędne do oszacowań gęstości algorytmu domyślnego. Ta opcja jest dostępna wyłącznie dla predykcji przestrzenno-czasowej.

Dodaj. Otwiera okno dialogowe służące do dodawania źródeł danych. Źródłem danych może być plik dBase podany z plikiem kształtu, plik danych produktu IBM SPSS Statistics lub otwarty zbiór danych z bieżącej sesji.

Powiąz. Otwiera okno dialogowe w celu określenia identyfikatorów (współrzędnych lub kluczy) wiążących dane z mapami. Każde źródło danych musi zawierać jeden lub większą liczbę identyfikatorów, które wiążą dane z mapą. Pliki dBase, które pochodzą z pliku kształtu, zwykle zawierają pole, które jest automatycznie używane jako identyfikator domyślny. W innych źródłach danych należy określić zmienne używane jako identyfikatory.

Sprawdź poprawność klucza. Otwiera okno dialogowe służące do sprawdzania dopasowania klucza do mapy i źródła danych.

Reguły asocjacji geoprzestrzennych

- Jako źródło danych o zdarzeniu należy wybrać przynajmniej jedno źródło danych.
- Wszystkie źródła danych o zdarzeniu muszą używać identyfikatorów powiązań mapy o tej samej postaci: współrzędnych lub wartości kluczowych.
- Jeśli źródła danych o zdarzeniach są powiązane z mapami zawierającymi wartości kluczy, we wszystkich źródłach muszą być używane te same właściwości mapy (np. wielokąty, punkty lub linie).

Modelowanie przestrzenno-czasowe

- Musi istnieć kontekstowe źródło danych.
- Jeśli istnieje tylko jedno źródło danych (plik danych bez powiązanej mapy), muszą w nim występować wartości współrzędnych.
- Jeśli istnieją dwa źródła danych, jedno musi zawierać dane kontekstowe, a drugie – dane o gęstości punktów.
- Nie można uwzględnić więcej niż dwóch źródeł danych.

Dodawanie źródła danych

Źródłem danych może być plik dBase podany z plikiem kształtu i plikiem kontekstu, plik danych produktu IBM SPSS Statistics lub otwarty zbiór danych z bieżącej sesji.

To samo źródło danych może być dodane wielokrotnie w celu użycia różnych skojarzeń przestrzennych.

Skojarzenia danych i map

Każde źródło danych musi mieć przynajmniej jeden identyfikator wiążący dane z mapą.

Współrzędne

Jeśli źródło danych zawiera zmienne przedstawiające współrzędne kartezjańskie, wybierz zmienną wskazującą na współrzędne X i Y. W przypadku reguł skojarzeń geoprzestrzennych może występować także współrzędna Z.

Wartości kluczy

Kluczowe wartości w zmiennych w źródle danych odpowiadają wybranym kluczom mapy. Na przykład mapa regionów może mieć identyfikator z nazwą (klucz mapy) stanowiący etykietę każdego regionu. Identyfikator ten odpowiada zmiennej w danych również zawierającej nazwy regionów (klucz danych). Zmienne są dopasowywane do kluczy map na podstawie kolejności ich wyświetlania na obu listach.

Sprawdzanie poprawności kluczy

Okno dialogowe Sprawdź poprawność kluczy przedstawia podsumowanie dopasowania rekordów między mapą i źródłem danych na podstawie wybranych kluczy identyfikujących. Jeśli występują niedopasowane wartości kluczy, można je ręcznie dopasować do wartości kluczy mapy.

Reguły asocjacji geoprzestrzennych

Przy regułach skojarzeń geoprzestrzennych po zdefiniowaniu map i źródeł danych pozostałe kroki w kreatorze to:

- Jeśli występuje wiele źródeł danych, zdefiniuj sposób ich łączenia.
- Wybierz zmienne, które mają być użyte jako zmienne warunków i predykcji w analizie.

Opcjonalnie można również wykonać następujące czynności:

- Wybierz różne opcje wyników.
- Zapisz plik z modelem oceniania.
- Utwórz nowe zmienne dla wartości przewidywanych i reguł w źródle danych używanych w modelu.
- Dostosuj ustawienia budowania reguł asocjacyjnych.

- Dostosuj ustawienia histogramu i agregacji.

Definiowanie zmiennych danych o zdarzeniu

Jeśli w regułach skojarzeń geoprzestrzennych istnieje więcej niż jedno źródło danych o zdarzeniach, źródła zostaną połączone.

- Domyślnie uwzględnione są wyłącznie zmienne występujące we wszystkich źródłach danych o zdarzeniach.
- Można wyświetlić listy wspólnych zmiennych, zmiennych z określonego źródła danych lub zmiennych ze wszystkich źródeł danych, a następnie wybrać zmienne wyznaczone do uwzględnienia.
- Przy wspólnych zmiennych wartości **Typ** oraz **Pomiar** muszą być takie same dla wszystkich źródeł danych. Przy sprzeczności można określić typ i poziom pomiaru używany we wszystkich wspólnych zmiennych.

Wybieranie zmiennych

Lista dostępnych zmiennych obejmuje zmienne ze źródła danych o zdarzeniu oraz ze źródeł danych kontekstowych.

- Listą wyświetlanych zmiennych można sterować, wybierając źródło danych z listy **Źródła danych**.
- Należy wybrać co najmniej dwie zmienne. Przynajmniej jedna z nich musi być warunkiem, a jedna – predykcją. Istnieje kilka sposobów spełnienia tego ograniczenia, takich jak wybranie dwóch zmiennych z listy **Oba (warunek i predykcja)**.
- Reguły asocjacyjne przewidują wartości zmiennych predykcyjnych opartych na wartościach zmiennych warunkowych. Na przykład w regule „jeśli $x=1$ i $y=2$, to $z=3$ ”, wartości x i y są warunkami, a wartość z jest predykcją.

Wynik

Tabele reguł

Każda tabela reguł przedstawia górną liczbę reguł, wartości ufności, pokrycie reguł, przyrost, pokrycie warunków i wdrażalność. Każda tabela jest sortowana wg wartości wybranego kryterium. Można zarówno wyświetlić wszystkie reguły lub górną **liczbę** reguł według wybranego kryterium.

Sortowalna chmura słów

Górna liczba reguł oparta na wartościach wybranego kryterium. Rozmiar tekstu wskazuje na względną ważność reguły. Wyjście interaktywne przedstawia górne reguły ufności, pokrycie reguł, przyrost, pokrycie warunków i wdrażalność. Wybrane kryterium określa listę reguł wyświetlanych domyślnie. W wynikach można interaktywnie wybrać inne kryteria. **Maks. liczba reguł do wyświetlenia** określa liczbę reguł wyświetlanych w wyniku.

Odwzorowanie

Interaktywny wykres słupkowy i mapa górnych reguł określonych na podstawie wartości wybranego kryterium. Każde wyjście interaktywne przedstawia górne reguły ufności, obsługę reguł, przyrost, pokrycie warunków i wdrażalność. Wybrane kryterium określa listę reguł wyświetlanych domyślnie. W wynikach można interaktywnie wybrać inne kryteria. **Maks. liczba reguł do wyświetlenia** określa liczbę reguł wyświetlanych w wyniku.

Tabele informacji o modelu

Transformacje zmiennych.

Opisuje przekształcenia stosowane do zmiennych używanych w analizie.

Podsumowanie rekordów.

Liczba i wartość procentowa uwzględnionych i wykluczonych rekordów.

Statystyki reguł.

Statystyka podsumowująca pokrycie warunków, ufność, pokrycie reguł, przyrost i wdrażalność. Statystyki obejmują średnią, odchylenie standardowe, minimum, maksimum i odchylenie standardowe.

Najczęstsze elementy.

Pozycje, które występują najczęściej. Pozycja jest uwzględniona w warunku lub predykcji w regule. Na przykład wiek < 18 lub płeć=kobieta.

Najczęstsze zmienne.

Zmienne, które występują najczęściej.

Wykluczone dane wejściowe.

Zmienne wykluczone z analizy i powód wykluczenia każdej zmiennej.

Kryteria tabel reguł, chmury słów i map

Ufność.

Wartość procentowa określająca poprawną predykcję reguł.

Pokrycie reguł.

Wartość procentowa obserwacji, dla których reguła jest prawdziwa. Na przykład, jeśli treść reguły to „jeśli x=1 i y=2, to z=3”, pokrycie reguł jest wartością procentową obserwacji, w których dane mają wartość x=1, y=2 i z=3.

Przyrost.

Przyrost jest miarą stopnia, w którym reguła oferuje lepszą predykcję w stosunku do przypadkowej. Jest to iloraz poprawnych predykcji do łącznej liczby wystąpień wartości przewidywanej. Wartość musi być większa od 1. Jeśli, na przykład wartość przewidywana występuje w 20% czasu, a ufność predykcji wynosi 80%, przyrost ma wartość 4.

Pokrycie warunków.

Wartość procentowa obserwacji, dla których warunek reguły jest prawdziwy. Na przykład, jeśli treść reguły to „jeśli x=1 i y=2, to z=3”, pokrycie warunków to proporcja obserwacji, w których dane mają wartość x=1 i y=2.

Wdrażalność.

Odsetek nieprawidłowych predykcji, gdy warunki są spełnione. Wdrażalność jest równa (1-ufność) razy pokrycie warunków lub pokrycie warunków minus pokrycie reguł.

Zapisywanie

Zapisz mapy i dane kontekstowe jako specyfikację mapy

Zapisz specyfikację mapy do pliku zewnętrznego (.mplan). Plik specyfikacji mapy można wczytać do kreatora w celu dalszej analizy. Pliku specyfikacji mapy można także użyć z poleceniem SPATIAL ASSOCIATION RULES.

Skopiuj pliki mapy i danych do specyfikacji

Dane z plików kształtu map, zewnętrzne dane plików oraz zbiory danych używane w specyfikacji map są zapisywane w pliku specyfikacji mapy.

Ocenianie

Zapisuje wartości najlepszych reguł, współczynnik ufności reguł oraz wartości identyfikatorów liczbowych reguł jako nowe zmienne w określonym źródle danych.

Źródło danych do oceny

Źródło lub źródła danych, w których tworzone są nowe zmienne. Jeśli źródło danych nie zostało jeszcze otwarte w bieżącej sesji, zostanie to zrobione. Aby zapisać nowe zmienne, należy bezpośrednio zapisać zmodyfikowany plik.

Wartości docelowe

Utwórz nowe zmienne dla wybranych zmiennych docelowych (predykcji).

- Dla każdej zmiennej docelowej tworzone są dwie nowe zmienne: z wartością przewidywaną i współczynnikiem ufności.
- Dla ilościowych zmiennych docelowych wartość przewidywana jest ciągiem opisującym przedział wartości. Wartość o postaci „(wartość1, wartość2]” oznacza przedział większy od wartości1 i mniejszy lub równy wartości2.

Liczba najlepszych reguł

Tworzy nowe zmienne dla określonej liczby najlepszych reguł. Dla każdej zmiennej tworzone są trzy reguły: wartość reguły, współczynnik ufności i wartość identyfikatora liczbowego reguły.

Przedrostek nazwy

Przedrostek do używania z nazwami nowych zmiennych.

Reguły asocjacji geoprzestrzennych

Parametry budowania reguł określają kryteria generowanych reguł asocjacyjnych.

Elementy dla reguł

Liczba wartości zmiennych, które można uwzględnić w warunkach reguł i predykcjach. Maksymalna łączna liczba elementów to 10. Na przykład w regule „jeśli $x=1$ i $y=2$, to $z=3$ ” istnieją dwie pozycje warunków i jedna wartość przewidywana.

Maksymalna liczba predykcji.

Maksymalna liczba wartości zmiennych, które mogą wystąpić w wartościach przewidywanych reguły.

Maksymalna liczba warunków.

Maksymalna liczba wartości zmiennych, które mogą wystąpić w warunkach reguły.

Wyklucz parę

Wyklucza określoną parę zmiennych z uwzględnienia w tej samej regule.

Kryteria reguł

Ufność.

Minimalna ufność reguły wymagana do jej uwzględnienia w wyniku. Ufność to wartość procentowa określająca poprawnie przewidziane wartości.

Pokrycie reguł.

Minimalne pokrycie reguły wymagane do jej uwzględnienia w wyniku. Wartość procentowa przedstawiająca udział obserwacji, w których reguła jest prawdziwa dla obserwowanych danych. Na przykład, jeśli treść reguły to „jeśli $x=1$ i $y=2$, to $z=3$ ”, pokrycie reguł jest wartością procentową obserwacji, w których dane mają wartość $x=1$, $y=2$ i $z=3$.

Pokrycie warunków.

Minimalne pokrycie warunków przez regułę wymagane do jej uwzględnienia w wyniku. Ta wartość przedstawia wartość procentową obserwacji, dla których warunek istnieje. Na przykład, jeśli treść reguły to „jeśli $x=1$ i $y=2$, to $z=3$ ”, pokrycie warunków to wartość procentowa obserwacji, w których dane mają wartość $x=1$ i $y=2$.

Przyrost.

Minimalny przyrost reguły wymagany do jej uwzględnienia w wyniku. Przyrost jest miarą stopnia, w którym reguła oferuje lepszą predykcję w stosunku do przypadkowej. Jest to iloraz poprawnych predykcji do łącznej liczby wystąpień wartości przewidywanej. Jeśli, na przykład wartość przewidywana występuje w 20% czasu, a ufność predykcji wynosi 80%, przyrost ma wartość 4.

Traktuj jako jednakowe

Identyfikuje pary zmiennych, które powinny być traktowane jako ta sama zmienna.

Opcje histogramu i agregacja

- Agregacja jest niezbędna, gdy liczba rekordów w danych przekracza liczbę właściwości mapy. Na przykład masz rekordy danych dla poszczególnych gmin, lecz mapę dla całych województw.
- W takiej sytuacji możesz określić metody tworzenia miary podsumowania dla zmiennych ilościowych i jakościowych. Zmienne nominalne są agregowane w oparciu o dominantę.

Ciągle

Przy zmiennej ciągłej (ilościowej) miarą podsumowania może być średnia, mediana lub suma.

Porządkowy

W zmiennych jakościowych miarą podsumowania może być mediana, dominanta, wartość maksymalna lub wartość minimalna.

Liczba kategorii

Ustawia maksymalną liczbę kategorii dla zmiennych ciągłych (ilościowych). Zmienne ciągłe są zawsze grupowane lub „kategoryzowane” w różne przedziały wartości. Na przykład: mniejsze lub równe 5, większe niż 5 oraz mniejsze lub równe 10 lub większe niż 10.

Agreguj mapę

Zastosuj agregację zarówno dla danych, jak i map.

Ustawienia niestandardowe dla określonych zmiennych

Możesz zastąpić domyślne miary podsumowania i liczbę kategorii dla określonych zmiennych.

- Kliknij ikonę, aby otworzyć okno dialogowe **Wybór zmiennych** i wybrać zmienną do dodania do listy.
- W kolumnie **Agregacja** wybierz miarę podsumowania.
- W zmiennych ciągłych kliknij przycisk w kolumnie **Kategorie** i określ dostosowaną liczbę kategorii zmiennej w oknie dialogowym **Kategorie**.

Predykcja przestrzenno-czasowa

Przy predykcji przestrzennej i szeregów czasowych po zdefiniowaniu map i źródeł danych pozostałe kroki w kreatorze to:

- Wybierz zmienne docelowe, zmienne czasu i opcjonalne predyktory.
- Zdefiniuj przedziały czasowe lub okresy cykliczne dla zmiennych czasu.

Opcjonalnie można również wykonać następujące czynności:

- Wybierz różne opcje wyników.
- Dostosuj parametry budowania modelu.
- Dostosuj ustawienia agregacji.
- Zapisz wartości przewidywane w zbiorze danych w bieżącej sesji lub w pliku danych w formacie produktu IBM SPSS Statistics.

Wybieranie zmiennych

Lista dostępnych zmiennych obejmuje zmienne z wybranych źródeł danych. Listą wyświetlanych zmiennych można sterować, wybierając źródło danych z listy **Źródła danych**.

Cel

Zmienna przewidywana jest wymagana. Zmienna przewidywana to zmienna, której wartości są przewidywane.

- Zmienna przewidywana musi być liczbową zmienną ilościową (ciągłą).
- Jeśli istnieją dwa źródła danych, wartość przewidywana jest oszacowaniem gęstości algorytmu domyślnego, a jako nazwa zmiennej przewidywanej wyświetlane jest słowo „Density”. Nie można tego zmienić.

Predyktory

Wymagane jest określenie co najmniej jednej zmiennej predyktora. To ustawienie jest opcjonalne.

Zmienne czasu

Należy wybrać co najmniej jedną zmienną przedstawiającą przedział czasu lub ustawić opcję **Okresy cykliczne**.

- Jeśli istnieją dwa źródła danych, należy wybrać zmienne czasu dla obu tych źródeł. Należy określić poziom przedziału ufności.
- W okresach cyklicznych należy określić zmienne definiujące cykle okresowości w panelu **Przedziały czasu** kreatora.

Przedziały czasowe

Opcje dostępne w tym panelu są oparte na opcjach wybranych w ustawieniu **Zmienne czasu** lub **Okres cykliczny** w kroku wybierania zmiennych.

Zmienne czasu

Wybrane zmienne czasowe. Zmienne czasu wybrane w kroku wybierania zmiennych pojawiają się na tej liście.

Przedział ufności. Wybierz dowolny szablon z rozwijanej listy. Przy pewnych przedziałach czasu można wprowadzić inne ustawienia, takie jak interwał między obserwacjami (przyrost) i wartość początkową. Przedział czasu jest używany dla wszystkich wybranych zmiennych czasu.

- W tej procedurze przyjęto założenie, że wszystkie obserwacje (rekordy) przedstawiają równomiernie rozłożone przedziały.
- W wybranym przedziale czasu procedura może wykryć brakujące obserwacje lub wielokrotne obserwacje z tego samego przedziału, które muszą zostać zagregowane. Na przykład, jeśli przedziałem czasu są dni, a po dacie 2014-10-27 występuje 2014-10-29, istnieje brakująca obserwacja dla 2014-10-28. Jeśli przedziałem czasu jest miesiąc, to wiele dat z tego samego miesiąca zostanie zagregowanych.
- Przy niektórych przedziałach czasu istnieje dodatkowe ustawienie umożliwiające zdefiniowanie odstępów w normalnie rozmieszczonych interwałach. Na przykład, jeśli przedziałem czasu są dni, ale istotne są tylko dni robocze, można określić pięciodniowy tydzień rozpoczynający się od poniedziałku.
- Jeśli wybrana zmienna czasu nie ma formatu daty ani godziny, przedział czasu zostanie automatycznie ustawiony na **Okresy** bez możliwości jego zmiany.

Zmienne cykli

Po wybraniu opcji **Okresy cykliczne** w kroku wybierania zmiennych należy wskazać zmienne definiujące okresy cykliczne. Okres cykliczny identyfikuje powtarzającą się cykliczną zmienność, taką jak liczba miesięcy w roku lub liczba dni w tygodniu.

- Można określić do trzech zmiennych definiujących okresy cykliczne.
- Pierwsza zmienna cykliczna określa najwyższy poziom cyklu. Na przykład, jeśli zmienność cykliczna zachodzi wg roku, kwartału i miesiąca, pierwszą zmienną cykliczną jest zmienna przedstawiająca rok.
- Długość cyklu pierwszej i drugiej zmiennej cyklu jest okresowością na kolejnych poziomach. Na przykład, jeśli zmiennymi cyklu jest rok, kwartał i miesiąc, długość pierwszego cyklu to 4, a drugiego — 3.
- Wartość początkowa zmiennej drugiego i trzeciego cyklu jest pierwszą wartością każdego z tych okresów cyklicznych.
- Długość cyklu i wartość początkowa muszą być dodatnią liczbą całkowitą.

Agregacja

- Po wybraniu dowolnej opcji **Predyktory** w kroku wybierania zmiennych istnieje możliwość wybrania metody agregacji predyktorów.
- Agregacja jest potrzebna, gdy w zdefiniowanym przedziale czasu istnieje więcej niż jeden rekord. Na przykład, jeśli przedziałem czasu jest miesiąc, to zagregowane zostanie wiele dat z tego samego miesiąca.
- Istnieje możliwość określenia metody tworzenia miary podsumowania dla zmiennych ilościowych i jakościowych. Zmienne nominalne są agregowane w oparciu o dominantę.

Ciągle

Przy zmiennej ciągłej (ilościowej) miarą podsumowania może być średnia, mediana lub suma.

Porządkowy

W zmiennych jakościowych miarą podsumowania może być mediana, dominanta, wartość maksymalna lub wartość minimalna.

Ustawienia niestandardowe dla określonych zmiennych

Istnieje możliwość zastąpienia domyślnych miar podsumowania dla określonych predyktorów.

- Kliknij ikonę, aby otworzyć okno dialogowe **Wybór zmiennych** i wybrać zmienną do dodania do listy.
- W kolumnie **Agregacja** wybierz miarę podsumowania.

Wynik

Mapy

Wartości przewidywane.

Mapa wartości dla wybranej zmiennej docelowej.

Korelacja

Mapa korelacji.

Skupienia

Mapa wyróżniająca podobne do siebie skupienia lokalizacji. Mapy skupień są dostępne tylko w przypadku modeli empirycznych.

Próg podobieństwa położenia.

Podobieństwo wymagane do tworzenia skupień. Wprowadzana wartość musi być liczbą większą od zera i mniejszą od 1.

Określ maksymalną liczbę skupień.

Maksymalna liczba skupień do wyświetlenia.

Tabele ewaluacji modelu

Specyfikacja modelu.

Podsumowanie specyfikacji użytej do uruchomienia analizy, w tym zmiennych przewidywanych, zmiennych wejściowych i lokalizacji.

Podsumowanie informacji czasowych.

Identyfikuje zmienne i przedziały czasowe używane w modelu.

Testy wpływów na strukturę średnich.

Wynik obejmuje wartości statystyki testu, stopnie swobody oraz poziom istotności dla modelu i każdego efektu.

Współczynniki struktury średnich modelu.

Wynik obejmuje wartości współczynników, błąd standardowy, wartość statystyki testu, poziom istotności oraz przedziały ufności dla każdego składnika modelu.

Współczynniki autoregresywne.

Wynik obejmuje wartości współczynników, błąd standardowy, wartość statystyki testu, poziom istotności oraz przedziały ufności dla każdego przesunięcia.

Testy kowariancji przestrzennej.

Dla modeli parametrycznych opartych na wariogramie wyświetla wyniki testu dobroci dopasowania dla struktury kowariancji przestrzennej. Wyniki testu mogą określać, czy należy modelować strukturę kowariancji przestrzennej parametrycznie, czy użyć modelu nieparametrycznego.

Parametryczna kowariancja przestrzenna.

Dla modeli parametrycznych opartych na wariogramie wyświetla oszacowania parametru dla parametrycznej kowariancji przestrzennej.

Opcje modelu

Ustawienia modelu

Automatycznie uwzględniaj wyraz wolny

Uwzględnij wyraz wolny w modelu.

Maksymalne przesunięcie autoregresyjne

Maksymalne przesunięcie autoregresyjne. Wartość musi być liczbą całkowitą w przedziale od 1 do 5.

Kowariancja przestrzenna

Określa metodę estymacji dla kowariancji przestrzennej.

Parametryczna

Metoda estymacji jest parametryczna. Dostępne są metody **Gausa**, **wykładnicza** lub **wykładniczo-potęgową**. Przy metodzie wykładniczo-potęgowej można określić wartość **Potęga**.

Nieparametryczny

Metoda estymacji jest nieparametryczna.

Zapisywanie

Zapisz mapy i dane kontekstowe jako specyfikację mapy

Zapisz specyfikację mapy do pliku zewnętrznego (.mplan). Plik specyfikacji mapy można wczytać do kreatora w celu dalszej analizy. Pliku specyfikacji mapy można także użyć z poleceniem SPATIAL TEMPORAL PREDICTION.

Skopiuj pliki mapy i danych do specyfikacji

Dane z plików kształtu map, zewnętrzne dane plików oraz zbiory danych używane w specyfikacji map są zapisywane w pliku specyfikacji mapy.

Ocenianie

Zapisuje wartości przewidywane, wariancję oraz górną i dolną granicę ufności zmiennej docelowej w wybranym pliku danych.

- Możesz zapisać wartości przewidywane w otwartym zbiorze danych w bieżącej sesji lub w pliku danych w formacie produktu IBM SPSS Statistics.
- Plik danych nie może być źródłem danych używanych w modelu.
- Plik danych musi zawierać wszystkie zmienne czasu i predyktory używane w modelu.
- Wartości czasu muszą być większe od wartości czasu używanych w modelu.

Zaawansowane

Maksymalna liczba obserwacji z brakami danych (%)

Maksymalna wartość procentowa obserwacji z brakującymi wartościami.

Poziom istotności

Poziom istotności pozwala określić, czy parametryczny model oparty na wariogramie jest odpowiedni. Wartość musi być liczbą większą od 0 i mniejszą od 1. Wartością domyślną jest 0,05. Poziom istotności jest używany w teście dobroci dopasowania dla struktury kowariancji przestrzennej. Dobroć dopasowania używana jest do określenia, czy należy użyć modelu parametrycznego, czy nieparametrycznego.

Współczynnik niepewności (%)

Współczynnik niepewności to wartość procentowa odzwierciedlająca wzrost niepewności przyszłych prognoz. Górna i dolna granica niepewności prognozy zwiększa się o tę wartość procentowo wraz z każdym krokiem w przyszłość.

Zakończenie

W ostatnim kroku kreatora modelowania geoprzestrzennego można uruchomić model lub wkleić do okna edytora komend wygenerowaną komendę. Istnieje możliwość zmodyfikowania i zapisania wygenerowanej składni komend na przyszłość.

Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. IBM może udostępniać ten materiał w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na produkty, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przesyłać na adres:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

Intellectual Property Licensing

*Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japonia*

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA NIE NARUSZA PRAW STRON TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w tej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych firm innych niż IBM zostały wprowadzone wyłącznie dla wygody użytkowników i w żadnym razie nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przystanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjodawcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) umożliwienia wymiany informacji między niezależnie utworzonymi programami i innymi programami

(łącznie z opisywanym) oraz (ii) wykorzystywania wymienianych informacji, powinni skontaktować się z:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z produktami firm innych niż IBM. Pytania dotyczące możliwości produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego zilustrowania tej działalności podane przykłady zawierają nazwy osób, firm i ich produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

LICENCJA W ZAKRESIE PRAW AUTORSKICH:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i rozpowszechniać te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu rozbudowy, użytkowania, handlowego lub w celu rozpowszechniania aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla którego napisane były programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować ani sugerować niezawodności, użyteczności i funkcjonalności tych programów. Programy przykładowe są dostarczane w stanie, w jakim się znajdują ("AS IS"), bez jakichkolwiek gwarancji (rękojmię również wyłącza się). IBM nie ponosi odpowiedzialności za jakiegokolwiek szkody wynikające z używania programów przykładowych.

Każda kopia programu przykładowego lub jakiegokolwiek jego fragment, jak też jakiegokolwiek prace pochodne muszą zawierać następujące uwagi dotyczące praw autorskich:

© Copyright IBM Corp. 2021. Fragmenty tego kodu pochodzą z przykładowych programów produktu IBM Corp. Programy przykładowe.

© Copyright IBM Corp. 1989-2021. Wszelkie prawa zastrzeżone.

Znaki towarowe

IBM, logo IBM i ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp., zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW, w sekcji "Copyright and trademark

information" (Informacje o prawach autorskich i znakach towarowych), pod adresem www.ibm.com/legal/copytrade.shtml.

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych w Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące Java są znakami towarowymi firmy i jej firm zależnych.

Indeks

A

- agregacja metodą bootstrap
 - w modelach liniowych [140](#)
- Agrestiego-Caffo
 - w procedurze Proporcje dla prób niezależnych [112](#)
- Agrestiego-Coulla
 - w procedurze Proporcje dla jednej próby [106](#)
- Agrestiego-Mina
 - w procedurze Proporcje dla prób zależnych [109](#)
- alfa Cronbacha
 - w analizie rzetelności [228](#), [229](#)
- analiza czułości
 - podczas symulacji [258](#)
- Analiza czynnikowa
 - brakujące wartości [185](#)
 - dodatkowe właściwości komendy [185](#)
 - format wyświetlania współczynników [185](#)
 - metody rotacji [184](#)
 - metody wyodrębniania [183](#)
 - oceny czynnikowe [184](#)
 - przeгляд [181](#)
 - przykład [181](#)
 - statystyki [181](#), [182](#)
 - statystyki opisowe [182](#)
 - wybór obserwacji [182](#)
 - wykresy ładunków [184](#)
 - zbieżność [183](#), [184](#)
- Analiza dyskryminacyjna
 - braki danych [180](#)
 - definiowanie zakresów [178](#)
 - dodatkowe właściwości komendy [181](#)
 - eksportowanie informacji o modelu [181](#)
 - kryteria [179](#)
 - Lambda Wilksa [179](#)
 - macierz kowariancji [180](#)
 - macierze [178](#)
 - metody dyskryminacyjne [179](#)
 - metody krokowe [177](#)
 - Odległość Mahalanobisa [179](#)
 - opcje wyświetlania [179](#), [180](#)
 - prawdopodobieństwa a priori [180](#)
 - przykład [177](#)
 - statystyki [177](#), [178](#)
 - statystyki opisowe [178](#)
 - V Rao [179](#)
 - współczynniki funkcji [178](#)
 - wybór obserwacji [178](#)
 - wykresy [180](#)
 - zapisywanie zmiennych klasyfikacyjnych [181](#)
 - zmiennie grupujące [177](#)
 - zmiennie niezależne [177](#)
- analiza głównych składowych [181](#), [183](#)
- Analiza mocy
 - statystyki [1](#)
- analiza najbliższego sąsiedztwa
 - najbliższe sąsiedztwo [171](#)
 - analiza najbliższego sąsiedztwa (*kontynuacja*)
 - zapisywanie zmiennych [173](#)
- Analiza najbliższego sąsiedztwa
 - opcje [174](#)
 - partycje [172](#)
 - widok modelu [174](#)
 - wybór predyktorów [172](#)
 - wyniki [173](#)
- Analiza ROC
 - statystyki i wykresy [246](#)
- Analiza rzetelności
 - dodatkowe właściwości komendy [231](#)
 - korelacje i kowariancje między pozycjami [229](#)
 - przykład [228](#)
 - statystyki [228](#), [229](#)
 - statystyki opisowe [229](#)
 - t-kwadrat Hotellinga [229](#)
 - Tabela ANOVA [229](#)
 - Test addytywności Tukey'a [229](#)
 - współczynnik korelacji wewnątrzklasowej [229](#)
 - współczynnik Kuder-Richardsona 20 [229](#)
- analiza skupień
 - Analiza skupień metodą k-średnich [195](#)
 - Hierarchiczna analiza skupień [193](#)
 - skuteczność [196](#)
- analiza skupień metodą k-średnich
 - skuteczność [196](#)
- Analiza skupień metodą k-średnich
 - brakujące wartości [197](#)
 - dodatkowe właściwości komendy [198](#)
 - iteracje [197](#)
 - kryteria zbieżności [197](#)
 - metody [195](#)
 - odległości od centrów skupień [197](#)
 - przeгляд [195](#)
 - przykłady [195](#)
 - przynależność do skupień [197](#)
 - statystyki [195](#), [197](#)
 - zapisywanie informacji o skupieniach [197](#)
- analiza szeregów czasowych
 - prognoza [166](#)
 - prognozowanie obserwacji [166](#)
- analiza wariancji
 - w estymacji krzywej [164](#)
 - w jednoczynnikowej analizie wariancji ANOVA [118](#)
 - w procedurze Regresja liniowa [150](#)
 - w procedurze Średnie [100](#)
- analiza what-if
 - podczas symulacji [258](#)
- analiza wielokrotnych odpowiedzi
 - Częstości wielokrotnych odpowiedzi [220](#)
 - tabela krzyżowa [221](#)
 - tabele częstości [220](#)
 - tabele krzyżowe wielokrotnych odpowiedzi [221](#)
- ANOVA
 - w GLM jednej zmiennej [122](#)
 - w modelach liniowych [144](#)

D

- d
 - w tabelach krzyżowych [94](#)
- D Somersa
 - w tabelach krzyżowych [94](#)
- dendrogramy
 - w hierarchicznej analizie skupień [195](#)
- DfBeta
 - w procedurze Regresja liniowa [148](#)
- DfFit
 - w procedurze Regresja liniowa [148](#)
- division
 - iloraz kolumn raportu [226](#)
- dobroć dopasowania
 - w regresji porządkowej [153](#)
- Dokładny dwumianowy
 - w procedurze Proporcje dla prób zależnych [109](#)
- dominanta
 - w częstościach [86](#)
- dopasowywanie rozkładu
 - podczas symulacji [255](#)
- Dwustopniowa analiza skupień
 - opcje [187](#)
 - statystyki [188](#)
 - zapisz do pliku zewnętrznego [188](#)
 - zapisz do roboczego pliku [188](#)

E

- eksploracja
 - statystyki [91](#)
- Eksploruj
 - braki danych [92](#)
 - dodatkowe właściwości komendy [92](#)
 - opcje [92](#)
 - transformacje potęgi [92](#)
 - wykresy [91](#)
- elementy równorzędne
 - w analizie najbliższego sąsiedztwa [176](#)
- eliminacja wsteczna
 - w procedurze Regresja liniowa [147](#)
- Estymacja krzywej
 - analiza wariancji [164](#)
 - modele [165](#)
 - prognoza [166](#)
 - uwzględnianie stałej [164](#)
 - zapisywanie przedziałów predykcji [166](#)
 - zapisywanie reszt [166](#)
 - zapisywanie wartości przewidywanych [166](#)
- estymator dwuwagi Tukey'a
 - w eksploracji [91](#)
- estymator fali Andrewsa
 - w eksploracji [91](#)
- estymatory Hodgesa-Lehmana
 - testy nieparametryczne dla prób zależnych [205](#)
- eta
 - w procedurze Średnie [100](#)
 - w tabelach krzyżowych [94](#)
- eta kwadrat
 - w procedurze Średnie [100](#)

F

- formatowanie
 - kolumn raportu [224](#)
- funkcje gęstości prawdopodobieństwa
 - podczas symulacji [260](#)
- funkcje skumulowanego rozkładu
 - podczas symulacji [260](#)

G

- gamma
 - w tabelach krzyżowych [94](#)
- gamma Goodmana i Kruskala
 - w tabelach krzyżowych [94](#)
- głębokość drzewa
 - w dwustopniowej analizie skupień [187](#)
- grupowanie
 - widok ogólny [189](#)
- GT2 Hochberga
 - w jednoczynnikowej analizie wariancji ANOVA [120](#)
 - w OML [128](#)

H

- H Kruskala-Wallisa
 - w testach dla dwóch prób niezależnych [216](#)
- Haucka-Andersona
 - w procedurze Proporcje dla prób niezależnych [112](#)
- Hierarchiczna analiza skupień
 - dendrogramy [195](#)
 - dodatkowe właściwości komendy [195](#)
 - macierze odległości [195](#)
 - metoda skupiania (aglomeracji) [194](#)
 - miary odległości [194](#)
 - miary podobieństwa [194](#)
 - położenie wykresu [195](#)
 - przeгляд aglomeracji [195](#)
 - przekształcanie wartości [194](#)
 - przykład [193](#)
 - przynależność do skupień [195](#)
 - skupianie obserwacji [193](#)
 - skupianie zmiennych [193](#)
 - statystyki [193](#), [195](#)
 - transformacja miar [194](#)
 - wykresy soplekowe [195](#)
 - zapisywanie zmiennych wyników [195](#)
- hierarchiczna dekompozycja [124](#)
- histogramy
 - w częstościach [87](#)
 - w eksploracji [91](#)
 - w regresji liniowej [148](#)

I

- ICC. Patrz współczynnik korelacji wewnątrzklasowej [229](#)
- iloraz kowariancji
 - w procedurze Regresja liniowa [148](#)
- iloraz wiarygodności chi-kwadrat
 - w regresji porządkowej [153](#)
 - w tabelach krzyżowych [94](#)
- informacje o diagnostyce obserwacjami
 - w procedurze Regresja liniowa [150](#)

informacje o teście współliniowości
w procedurze Regresja liniowa [150](#)

iteracje

w analizie czynnikowej [183](#), [184](#)
w analizie skupień metodą k-średnich [197](#)

J

Jednoczynnikowa ANOVA

brakujące wartości [121](#)
czynniki [118](#)
dodatkowe właściwości komendy [122](#)
kontrasty [119](#)
kontrasty wielomianowe [119](#)
model [124](#)
opcje [121](#)
statystyki [121](#)
testy post hoc [120](#)
w jednoczynnikowej analizie wariancji ANOVA [118](#)
wielokrotne porównania [120](#)

Jeffreysa

w procedurze Proporcje dla jednej próby [106](#)

K

kappa

w tabelach krzyżowych [94](#)

kappa Cohena

w tabelach krzyżowych [94](#)

Kappa wielu ocenających Fleissa [228](#), [229](#)

Karta L' Abb'e Plot

Metaanaliza wielkości efektu (binarnie) [81](#)

Karta Wykres bąbelkowy

Metaanaliza (binarnie) [78](#)
Metaanaliza ciągłej wielkości efektu [78](#)
Metaanaliza ilościowa [78](#)
Metaanaliza wielkości efektu (binarnie) [78](#)

Karta Wykres L' Abba

Metaanaliza (binarnie) [81](#)
Metaanaliza ciągłej wielkości efektu [81](#)
Metaanaliza ilościowa [81](#)

Karta Wykres lasu

Metaanaliza binarna [75](#)
Metaanaliza binarna wielkości efektu [75](#)
Metaanaliza ciągłej wielkości efektu [75](#)
Metaanaliza ilościowa [75](#)

Karta Wykres lejkowy

Meta-analiza ilościowa [79](#)
Metaanaliza binarna [79](#)
Metaanaliza binarna wielkości efektu [79](#)
Metaanaliza ciągłej wielkości efektu [79](#)

kategoria odniesienia

w OML [125](#), [126](#)

klasyfikacja

w analizie ROC [245](#)
w krzywej ROC [248](#)

kolumna ogółem

w raportach [226](#)

kontrasty

w jednoczynnikowej analizie wariancji ANOVA [119](#)
w OML [125](#), [126](#)

kontrasty Helmerta

w OML [125](#), [126](#)

kontrasty odchylenia

w OML [125](#), [126](#)

kontrasty powtarzane

w OML [125](#), [126](#)

kontrasty różnicy

w OML [125](#), [126](#)

kontrasty wielomianowe

w jednoczynnikowej analizie wariancji ANOVA [119](#)
w OML [125](#), [126](#)

korelacja Pearsona

w tabelach krzyżowych [94](#)

Korelacja Pearsona

w korelacjach parami [134](#)

korelacje

podczas symulacji [258](#)
rzędu zerowego [137](#)
w korelacjach parami [134](#)
w procedurze Korelacje cząstkowe [137](#)
w tabelach krzyżowych [94](#)

Korelacje cząstkowe

brakujące wartości [137](#)
dodatkowe właściwości komendy [138](#)
korelacje rzędu zerowego [137](#)
opcje [137](#)
statystyki [137](#)
w procedurze Regresja liniowa [150](#)

Korelacje parami

brakujące wartości [135](#)
dodatkowe właściwości komendy [136](#)
opcje [135](#)
poziom istotności [134](#)
przedział ufności [134](#)
przedziały ufności [136](#)
statystyki [135](#)
współczynniki korelacji [134](#)

korelacje rzędu zerowego

w procedurze Korelacje cząstkowe [137](#)

Kostki OLAP

statystyki [103](#)
tytuły [104](#)

KR20

w analizie rzetelności [229](#)

Kreator symulacji [252](#)

krokowa postępująca

w modelach liniowych [141](#)

kryteria informacyjne

w modelach liniowych [141](#)

kryteria wyboru

w procedurze Regresja liniowa [150](#)

kryterium dot. zabezpieczenia przed przeuczeniem

w modelach liniowych [141](#)

kryterium informacyjne Akaike

w modelach liniowych [141](#)

Krzywa ROC

statystyki i wykresy [248](#)

książka kodowa

output [82](#)
statystyki [84](#)

kurtoza

w częstościach [86](#)
w eksploracji [91](#)
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)

kurtoza (*kontynuacja*)
w raporcie z podsumowaniami w wierszach [224](#)
w Raport: Podsumowania w kolumnach [226](#)
w statystykach opisowych [88](#)
kwadrat odległości euklidesowej
w odległościach [139](#)
kwartyle
w częstościach [86](#)

L

lambda
w tabelach krzyżowych [94](#)
lambda Goodmana i Kruskala
w tabelach krzyżowych [94](#)
Lambda Wilksa
w analizie dyskryminacyjnej [179](#)
liczba obserwacji
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)
liczebność obserwowana
w tabelach krzyżowych [96](#)
liczebność oczekiwana
w tabelach krzyżowych [96](#)
Logita
w procedurze Proporcje dla jednej próby [106](#)

Ł

łączenie reguł
w modelach liniowych [142](#)

M

M-estymator Hampela
w eksploracji [91](#)
M-estymator Hubera
w eksploracji [91](#)
M-estymatory
w eksploracji [91](#)
macierz korelacji
w analizie czynnikowej [181](#), [182](#)
w analizie dyskryminacyjnej [178](#)
w regresji porządkowej [153](#)
macierz kowariancji
w analizie dyskryminacyjnej [178](#), [180](#)
w OML [131](#)
w procedurze Regresja liniowa [150](#)
w regresji porządkowej [153](#)
macierz transformacji
w analizie czynnikowej [181](#)
macierz wzorów
w analizie czynnikowej [181](#)
maksimum
porównywanie kolumn raportu [226](#)
w częstościach [86](#)
w eksploracji [91](#)
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)
w statystykach opisowych [88](#)
maksymalna liczba gałęzi

maksymalna liczba gałęzi (*kontynuacja*)
w dwustopniowej analizie skupień [187](#)
mapa kwadratowa
w analizie najbliższego sąsiedztwa [176](#)
maximum
w statystykach ilorazowych [237](#), [239](#)
McNemara
w procedurze Proporcje dla prób zależnych [109](#)
McNemara (ze skorygowaną ciągłością)
w procedurze Proporcje dla prób zależnych [109](#)
mediana
w częstościach [86](#)
w eksploracji [91](#)
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)
Mediana
w statystykach ilorazowych [237](#), [239](#)
mediana z danych pogrupowanych
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)
Meta-analiza ilościowa
Karta Wykres lasu [75](#)
Okno dialogowe drukowania [44](#)
Okno dialogowe Odchylenie [42](#)
Okno dialogowe wnioskowania [41](#)
Okno dialogowe wykresu [46](#)
Okno dialogowe Zapisz [45](#)
Okno kontrastu [42](#)
Skumulowany wykres lasu [77](#)
Metaanaliza
ciągłe wyniki [37](#)
regresja [37](#)
Wykresy [75](#)
wyniki binarne [37](#)
Metaanaliza (binarnie)
Karta Wykres bąbelkowy [78](#)
Karta Wykres L' Abba [81](#)
Okno dialogowe drukowania [61](#)
Zakładka Galbraith [80](#)
Metaanaliza binarna
Karta Wykres lasu [75](#)
Karta Wykres lejkowy [79](#)
Okno dialogowe Analiza [56](#)
Okno dialogowe Odchylenie [59](#)
Okno dialogowe Trim-and-Fill [60](#)
Okno dialogowe Wnioskowanie [57](#)
Okno dialogowe Wykres [63](#)
Okno dialogowe Zapisz [62](#)
Okno kontrastu [58](#)
Okno Kryteria [55](#)
Skumulowany wykres lasu [77](#)
Metaanaliza binarna wielkości efektu
Karta Wykres lasu [75](#)
Karta Wykres lejkowy [79](#)
Okno dialogowe Analiza [65](#)
Okno dialogowe drukowania [69](#)
Okno dialogowe Trim-and-Fill [68](#)
Okno dialogowe Wnioskowanie [66](#)
Skumulowany wykres lasu [77](#)
Zakładka Galbraith [80](#)
Metaanaliza ciągłej wielkości efektu
Karta Wykres bąbelkowy [78](#)

- Metaanaliza ciągłej wielkości efektu (*kontynuacja*)
 - Karta Wykres L' Abba [81](#)
 - Karta Wykres lasu [75](#)
 - Karta Wykres lejkowy [79](#)
 - Okno dialogowe Analiza [48](#)
 - Okno dialogowe drukowania [52](#)
 - Okno dialogowe Obciążenie [50](#)
 - Okno dialogowe Trim-and-Fill [51](#)
 - Okno dialogowe wnioski [49](#)
 - Okno dialogowe wykresu [54](#)
 - Okno dialogowe Zapisz [53](#)
 - Okno kontrastu [50](#)
 - Skumulowany wykres lasu [77](#)
 - Zakładka Galbraith [80](#)
- Metaanaliza ilościowa
 - Karta Wykres bąbelkowy [78](#)
 - Karta Wykres L' Abba [81](#)
 - Karta Wykres lejkowy [79](#)
 - Okno dialogowe Analiza [40](#)
 - Okno dialogowe Trim-and-Fill [43](#)
 - Okno Kryteria [39](#), [47](#)
 - Zakładka Galbraith [80](#)
- Metaanaliza wielkości efektu (binarnie)
 - Karta Wykres bąbelkowy [78](#)
 - Karta Wykres L' Abba [81](#)
 - Okno dialogowe Odchylenie [67](#)
 - Okno dialogowe Zapisz [70](#)
 - Okno Kontrast [67](#)
 - Okno Kryteria [64](#)
- Metaanaliza wielkości efektu binarnego
 - Okno dialogowe wykresu [71](#)
- Metaanaliza: regresja
 - Okno dialogowe Zapisz [74](#)
 - Wykres bąbelkowy [78](#)
- metoda alfa [183](#)
- metoda maksymalnej wiarygodności
 - w analizie czynnikowej [183](#)
- metoda nieważonych najmniejszych kwadratów
 - w analizie czynnikowej [183](#)
- metoda obrazu [183](#)
- metoda osi głównych [183](#)
- metoda uogólnionych najmniejszych kwadratów
 - w analizie czynnikowej [183](#)
- metoda ważonych najmniejszych kwadratów
 - w regresji liniowej [146](#)
- miara niepodobieństwa Lance'a i Williamsa
 - w odległościach [139](#)
- miara odległości phi-kwadrat
 - w odległościach [139](#)
- miara różnica wielkości
 - w odległościach [139](#)
- miara różnica wzoru
 - w odległościach [139](#)
- miary odległości
 - w analizie najbliższego sąsiedztwa [171](#)
 - w hierarchicznej analizie skupień [194](#)
 - w odległościach [139](#)
- miary podobieństwa
 - w hierarchicznej analizie skupień [194](#)
 - w odległościach [139](#)
- miary rozkładu
 - w częstościach [86](#)
 - w statystykach opisowych [88](#)
- miary rozproszenia (*kontynuacja*)
 - w częstościach [86](#)
 - w eksploracji [91](#)
 - w statystykach ilorazowych [237](#), [239](#)
 - w statystykach opisowych [88](#)
- miary tendencji centralnej
 - w częstościach [86](#)
 - w eksploracji [91](#)
 - w statystykach ilorazowych [237](#), [239](#)
- minimum
 - porównywanie kolumn raportu [226](#)
 - w częstościach [86](#)
 - w eksploracji [91](#)
 - w kostkach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
 - w statystykach ilorazowych [237](#), [239](#)
 - w statystykach opisowych [88](#)
- mnożenie
 - iloczyn kolumn raportu [226](#)
- model Guttmana
 - w analizie rzetelności [228](#), [229](#)
- model kwadratowy
 - w estymacji krzywej [165](#)
- model liniowy
 - w estymacji krzywej [165](#)
- model logarytmiczny
 - w estymacji krzywej [165](#)
- model logistyczny
 - w estymacji krzywej [165](#)
- model odwrotny
 - w estymacji krzywej [165](#)
- model położenia
 - w regresji porządkowej [154](#)
- model potęgowy
 - w estymacji krzywej [165](#)
- model równoległy
 - w analizie rzetelności [228](#), [229](#)
- model S
 - w estymacji krzywej [165](#)
- model skali
 - w regresji porządkowej [154](#)
- model sześcienny
 - w estymacji krzywej [165](#)
- model ściśle równoległy
 - w analizie rzetelności [228](#), [229](#)
- model wykładniczy
 - w estymacji krzywej [165](#)
- model wzrostu
 - w estymacji krzywej [165](#)
- model złożony
 - w estymacji krzywej [165](#)
- modele liniowe
 - automatyczne przygotowanie danych [141](#), [143](#)
 - cele [140](#)
 - informacji statystycznej [143](#)
 - kryterium R kwadrat [143](#)
 - łączenie reguł [142](#)
 - opcje modelu [143](#)
 - oszacowane średnie [145](#)
 - podsumowanie modelu [143](#)
 - podsumowanie tworzenia modelu [146](#)
 - powielenie wyników [143](#)
 - poziom ufności [141](#)

modele liniowe (*kontynuacja*)
przewidywane przez obserwowane [144](#)
reszty [144](#)
Tabela ANOVA [144](#)
wartości odstające [144](#)
ważność predyktorów [144](#)
współczynniki [145](#)
wybór modelu [141](#)
zestawy [142](#)
modele użytkownika
w OML [124](#)
modelowanie geoprzestrzenne [267–277](#)
modelowanie przestrzenne [267](#)

N

najlepsze podzbiory
w modelach liniowych [141](#)
najmniejsza istotna różnica
w jednoczynnikowej analizie wariancji ANOVA [120](#)
w OML [128](#)
Newcomb'a (ze skorygowaną ciągłością)
w procedurze Proporcje dla prób niezależnych [112](#)
Newcomba
w procedurze Proporcje dla prób niezależnych [112](#)
w procedurze Proporcje dla prób zależnych [109](#)
Newman-Keuls
w OML [128](#)
NIR Fishera
w OML [128](#)
Normalizacja Kaisera
w analizie czynnikowej [184](#)
normalne wykresy prawdopodobieństwa
K-K [242](#)
P-P [240](#)
w eksploracji [91](#)
w regresji liniowej [148](#)
numerowanie stron
w raportach w wierszach [225](#)
w raportach z podsumowaniami w kolumnach [227](#)

O

obsługa szumu
w dwustopniowej analizie skupień [187](#)
Ocena (ze skorygowaną ciągłością) [107](#)
Ocena Wilsona
w procedurze Proporcje dla jednej próby [106](#)
ocena Wilsona (ze skorygowaną ciągłością)
w procedurze Proporcje dla jednej próby [106](#)
oceny czynnikowe [184](#)
oceny czynnikowe Andersona-Rubina [184](#)
oceny czynnikowe Bartletta [184](#)
odchylenie standardowe
w częstościach [86](#)
w eksploracji [91](#)
w kostkach OLAP [103](#)
w podsumowaniach obserwacji [98](#)
w procedurze Średnie [100](#)
w raporcie z podsumowaniami w wierszach [224](#)
w Raport: Podsumowania w kolumnach [226](#)
w statystykach ilorazowych [237, 239](#)
w statystykach opisowych [88](#)

Odległości
dodatkowe właściwości komendy [139](#)
miary niepodobieństwa [139](#)
miary podobieństwa [139](#)
obliczanie odległości między obserwacjami [138](#)
obliczanie odległości między zmiennymi [138](#)
przekształcanie wartości [139](#)
przykład [138](#)
statystyki [138](#)
transformacja miar [139](#)
odległości najbliższego sąsiedztwa
w analizie najbliższego sąsiedztwa [176](#)
odległość chi-kwadrat
w odległościach [139](#)
Odległość Cooka
w OML [131](#)
w procedurze Regresja liniowa [148](#)
odległość Czebyszewa
w odległościach [139](#)
Odległość euklidesowa
w analizie najbliższego sąsiedztwa [171](#)
w odległościach [139](#)
Odległość Mahalanobisa
w analizie dyskryminacyjnej [179](#)
w procedurze Regresja liniowa [148](#)
odległość manhattańska
w analizie najbliższego sąsiedztwa [171](#)
odległość miejska
w analizie najbliższego sąsiedztwa [171](#)
w odległościach [139](#)
odległość Minkowskiego
w odległościach [139](#)
odsyłacz
w regresji porządkowej [152](#)
Okno dialogowe Analiza
Meta-analiza ilościowa [40](#)
Metaanaliza binarna [56](#)
Metaanaliza binarna wielkości efektu [65](#)
Metaanaliza ciągłej wielkości efektu [48](#)
Okno dialogowe drukowania
Meta-analiza ilościowa [44](#)
Metaanaliza (binarnie) [61](#)
Metaanaliza binarna wielkości efektu [69](#)
Metaanaliza ciągłej wielkości efektu [52](#)
Regresja metaanalizy [74](#)
Okno dialogowe Obciążenie
Metaanaliza ciągłej wielkości efektu [50](#)
Okno dialogowe Odchylenie
Meta-analiza ilościowa [42](#)
Metaanaliza binarna [59](#)
Metaanaliza wielkości efektu (binarnie) [67](#)
Okno dialogowe Trim-and-Fill
Metaanaliza binarna [60](#)
Metaanaliza binarna wielkości efektu [68](#)
Metaanaliza ciągłej wielkości efektu [51](#)
Metaanaliza ilościowa [43](#)
Okno dialogowe wnioskowania
Meta-analiza ilościowa [41](#)
Metaanaliza ciągłej wielkości efektu [49](#)
Regresja metaanalizy [73](#)
Okno dialogowe Wnioskowanie
Metaanaliza binarna [57](#)
Metaanaliza binarna wielkości efektu [66](#)
Okno dialogowe Wykres

- Okno dialogowe Wykres (*kontynuacja*)
 - Metaanaliza binarna [63](#)
- Okno dialogowe wykresu
 - Meta-analiza ilościowa [46](#)
 - Metaanaliza ciągłej wielkości efektu [54](#)
 - Metaanaliza wielkości efektu binarnego [71](#)
 - Regresja metaanalizy [75](#)
- Okno dialogowe Zapisz
 - Meta-analiza ilościowa [45](#)
 - Metaanaliza binarna [62](#)
 - Metaanaliza ciągłej wielkości efektu [53](#)
 - Metaanaliza wielkości efektu (binarnie) [70](#)
 - Metaanaliza: regresja [74](#)
- Okno Kontrast
 - Metaanaliza wielkości efektu (binarnie) [67](#)
- Okno kontrastu
 - Metaanaliza binarna [58](#)
 - Metaanaliza ciągłej wielkości efektu [50](#)
 - Metaanaliza ilościowa [42](#)
- Okno Kryteria
 - Meta-analiza ilościowa [47](#)
 - Metaanaliza binarna [55](#)
 - Metaanaliza binarna wielkości efektu [64](#)
 - Metaanaliza ilościowa [39](#)
 - Regresja metaanalizy [72](#)
- Omega McDonalda
 - w analizie rzetelności [228](#), [229](#)
- OML jednej zmiennej [128](#), [130](#), [134](#)
- ostatni
 - w kostkach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
- oszacowania parametrów
 - w regresji porządkowej [153](#)

- P**
- pełne modele czynnikowe
 - w OML [124](#)
- percentyle
 - podczas symulacji [261](#)
 - w częstościach [86](#)
 - w eksploracji [91](#)
- Percentyle [89](#)
- phi
 - w tabelach krzyżowych [94](#)
- pierwsza
 - w kostkach OLAP [103](#)
 - w procedurze Średnie [100](#)
- pierwszy
 - w podsumowaniach obserwacji [98](#)
- plany testów heteroskedastyczności
 - w OML [133](#)
- PLUM
 - w regresji porządkowej [151](#)
- Podsumowania
 - opcje [98](#)
 - statystyki [98](#)
- Podsumowanie [97](#)
- podsumowanie błędów
 - w analizie najbliższego sąsiedztwa [177](#)
- podsumowanie całości
 - w raportach z podsumowaniami w kolumnach [227](#)
- podsumy (*kontynuacja*)
 - w raportach z podsumowaniami w kolumnach [227](#)
- pomocniczy model regresji
 - w OML [133](#)
- poprawka Yatesa w kierunku ciągłości
 - w tabelach krzyżowych [94](#)
- porównywanie grup
 - w kostkach OLAP [104](#)
- porównywanie zmiennych
 - w kostkach OLAP [104](#)
- powiązanie liniowe
 - w tabelach krzyżowych [94](#)
- procent całości
 - w tabelach krzyżowych [96](#)
- procent w kolumnie
 - w tabelach krzyżowych [96](#)
- procent w wierszu
 - w tabelach krzyżowych [96](#)
- procenty
 - w tabelach krzyżowych [96](#)
- prognoza
 - w estymacji krzywej [166](#)
- proporcje dla jednej próby [109](#)
- Proporcje dla jednej próby [105](#), [106](#)
- Proporcje dla prób niezależnych [110](#), [112](#)
- Proporcje dla prób zależnych [108](#)
- proste kontrasty
 - w OML [125](#), [126](#)
- próba szkoleniowa
 - w analizie najbliższego sąsiedztwa [172](#)
- próba wstrzymana
 - w analizie najbliższego sąsiedztwa [172](#)
- próby zależne [215](#), [218](#)
- próg początkowy
 - w dwustopniowej analizie skupień [187](#)
- przebieg iteracji
 - w regresji porządkowej [153](#)
- przeciętne bezwzględne odchylenie (AAD)
 - w statystykach ilorazowych [237](#), [239](#)
- przedział
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
- przedział ufności
 - w korelacjach parami [134](#)
- przedziały ilorazu wiarygodności
 - Testy nieparametryczne dla jednej próby [199](#)
- przedziały Jeffreya
 - Testy nieparametryczne dla jednej próby [199](#)
- przedziały predykcji
 - zapisywanie w estymacji krzywej [166](#)
 - zapisywanie w procedurze Regresja liniowa [148](#)
- przedziały ufności
 - w analizie ROC [246](#)
 - w eksploracji [91](#)
 - w jednoczynnikowej analizie wariancji ANOVA [121](#)
 - w korelacjach parami [136](#)
 - w krzywej ROC [248](#)
 - w OML [125](#)
 - w procedurze Regresja liniowa [150](#)
 - w teście t dla jednej próby [117](#)
 - w teście t dla prób niezależnych [115](#)
 - w teście t dla prób zależnych [116](#)
 - zapisywanie w procedurze Regresja liniowa [148](#)
- przegłdarka skupień

przeglądarka skupień (*kontynuacja*)
filtrowanie rekordów [193](#)
korzystanie [192](#)
o modelach skupień [189](#)
obrót skupienia i zmienne [190](#)
podsumowanie modelu [189](#)
porównanie skupień [192](#)
porządek wyświetlania skupień [191](#)
przegląd [189](#)
rozkład komórek [192](#)
rozmiar skupień [192](#)
sortowanie skupień [191](#)
sortowanie wyświetlania zmiennych [191](#)
sortowanie zawartości komórki [191](#)
sortowanie zmiennych [191](#)
transponuj skupienia i zmienne [190](#)
ważność predyktorów [191](#)
widok centrów skupień [190](#)
widok podstawowy [191](#)
widok podsumowania [189](#)
widok porównania skupień [192](#)
widok rozkładu komórek [192](#)
widok rozmiarów skupień [192](#)
widok skupień [190](#)
widok ważności predyktora skupień [191](#)
wyświetlanie zawartości komórki [191](#)
przydział pamięci
w dwustopniowej analizie skupień [187](#)

Q

Q Cochrana
w testach dla kilku prób zależnych [218](#)

R

R 2
w procedurze Regresja liniowa [150](#)
w procedurze Średnie [100](#)
zmiana R 2 [150](#)
R kwadrat
w modelach liniowych [143](#)
R- kwadrat Nagelkerke'a
w regresji porządkowej [153](#)
R-E-G-W F
w jednoczynnikowej analizie wariancji ANOVA [120](#)
w OML [128](#)
R-E-G-W Q
w jednoczynnikowej analizie wariancji ANOVA [120](#)
w OML [128](#)
R-kwadrat McFadden
w regresji porządkowej [153](#)
R2 Coxa i Snella
w regresji porządkowej [153](#)
range
w częstościach [86](#)
w statystykach ilorazowych [237](#), [239](#)
Raport — podsumowania w kolumnach
braki danych [227](#)
dodatkowe właściwości komendy [227](#)
format kolumny [224](#)
numerowanie stron [227](#)
podsumowanie całości [227](#)

Raport — podsumowania w kolumnach (*kontynuacja*)
podsumy [227](#)
suma dla kolumn [226](#)
układ strony [225](#)
ustawienia strony [227](#)
Raport — podsumowania w wierszach
braki danych [225](#)
dodatkowe właściwości komendy [227](#)
format kolumny [224](#)
kolumny danych [223](#)
kolumny grupujące [223](#)
numerowanie stron [225](#)
puste wiersze przed każdą grupą [224](#)
sortowanie [223](#)
stopki [225](#)
tytuły [225](#)
układ strony [225](#)
ustawienia strony [224](#)
zmienne w tytułach [225](#)
raporty
iloczyn wartości kolumn [226](#)
iloraz wartości kolumn [226](#)
porównywanie kolumn [226](#)
raporty w kolumnach [225](#)
raporty w wierszach [223](#)
suma dla kolumn [226](#)
złożenia ogółem [226](#)
raporty w kolumnach [225](#)
regresja
Metaanaliza [37](#)
Regresja liniowa [146](#)
regresja wielokrotna [146](#)
wykresy [148](#)
regresja liniowa
braki danych [150](#)
metody wyboru zmiennych [147](#), [150](#)
statystyki [150](#)
Regresja liniowa
bloki [146](#)
dodatkowe właściwości komendy [151](#)
eksportowanie informacji o modelu [148](#)
reszty [148](#)
wagi [146](#)
wykresy [148](#)
zapisywanie zmiennych wyników [148](#)
zmienna filtrująca [148](#)
Regresja liniowa grzbietowa
opcje. [162](#)
partycja [161](#)
zmienne [161](#)
Regresja liniowa lasso
opcje. [159](#)
partycja [158](#)
zmienne [158](#)
Regresja liniowa sieci elastycznej
opcje. [156](#)
partycja [155](#)
zmienne [155](#)
Regresja metaanalizy
Okno dialogowe drukowania [74](#)
Okno dialogowe wnioskowania [73](#)
Okno dialogowe wykresu [75](#)
Okno Kryteria [72](#)
Regresja metodą cząstkowych najmniejszych kwadratów

- Regresja metodą cząstkowych najmniejszych kwadratów (*kontynuacja*)
 - eksportowanie zmiennych [168](#)
 - model [168](#)
 - regresja porządkowa
 - dodatkowe właściwości komendy [155](#)
 - model potężenia [154](#)
 - model skali [154](#)
 - Regresja porządkowa
 - odsyłacz [152](#)
 - opcje [152](#)
 - statystyki [151](#)
 - regresja wielokrotna
 - w regresji liniowej [146](#)
 - Reszta Pearsona
 - w regresji porządkowej [153](#)
 - reszty
 - w tabelach krzyżowych [96](#)
 - zapisywanie w estymacji krzywej [166](#)
 - zapisywanie w procedurze Regresja liniowa [148](#)
 - reszty niestandardyzowane
 - w OML [131](#)
 - reszty standaryzowane
 - w OML [131](#)
 - w procedurze Regresja liniowa [148](#)
 - Reszty studentyzowane
 - w procedurze Regresja liniowa [148](#)
 - reszty usuniętych
 - w OML [131](#)
 - w procedurze Regresja liniowa [148](#)
 - rho
 - w korelacjach parami [134](#)
 - w tabelach krzyżowych [94](#)
 - rotacja Equamax
 - w analizie czynnikowej [184](#)
 - rotacja prosta Oblimin
 - w analizie czynnikowej [184](#)
 - rotacja Quartimax
 - w analizie czynnikowej [184](#)
 - rotacja Varimax
 - w analizie czynnikowej [184](#)
 - różnice między grupami obserwacji
 - w kostkach OLAP [104](#)
 - różnice pomiędzy zmiennymi
 - w kostkach OLAP [104](#)
 - ryzyko
 - w tabelach krzyżowych [94](#)
 - ryzyko względne
 - w tabelach krzyżowych [94](#)
 - rzeczywiście istotna różnica Tukey'a
 - w jednoczynnikowej analizie wariancji ANOVA [120](#)
 - w OML [128](#)
 - rzetelność półwkowa
 - w analizie rzetelności [228](#), [229](#)
 - rzetelność Spearmana-Browna
 - w analizie rzetelności [229](#)
- S**
- s-stress
 - w skalowaniu wielowymiarowym [234](#)
 - selekcja krokowa
 - w procedurze Regresja liniowa [147](#)
 - selekcja postępująca
 - w analizie najbliższego sąsiedztwa [172](#)
 - selekcja postępująca (*kontynuacja*)
 - w procedurze Regresja liniowa [147](#)
 - Seria Walda-Wolfowitza
 - w testach dla dwóch prób niezależnych [214](#)
 - skala
 - w analizie rzetelności [228](#)
 - w metodzie Ważona Kappa [232](#)
 - w skalowaniu wielowymiarowym [234](#)
 - Skalowanie wielowymiarowe
 - definiowanie kształtu danych [235](#)
 - dodatkowe właściwości komendy [236](#)
 - kryteria [236](#)
 - miary odległości [235](#)
 - model skalowania [235](#)
 - opcje wyświetlania [236](#)
 - poziom pomiaru [235](#)
 - przekształcanie wartości [235](#)
 - przykład [234](#)
 - statystyki [234](#)
 - tworzenie macierzy odległości [235](#)
 - warunki porównań [235](#)
 - wymiary [235](#)
 - składniki interakcji [124](#), [154](#), [155](#)
 - skorygowane R²
 - w procedurze Regresja liniowa [150](#)
 - Skorygowany dwumianowy mid-p
 - w procedurze Proporcje dla prób zależnych [109](#)
 - skorygowany R kwadrat
 - w modelach liniowych [141](#)
 - skośność
 - w częstościach [86](#)
 - w eksploracji [91](#)
 - w kostkach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
 - w raporcie z podsumowaniami w wierszach [224](#)
 - w Raport: Podsumowania w kolumnach [226](#)
 - w statystykach opisowych [88](#)
 - Skumulowany wykres lasu
 - Meta-analiza ilościowa [77](#)
 - Metaanaliza binarna [77](#)
 - Metaanaliza binarna wielkości efektu [77](#)
 - Metaanaliza ciągłej wielkości efektu [77](#)
 - skupienia
 - przeglądanie skupień [189](#)
 - wybijanie procedury [185](#)
 - słownik
 - książka kodowa [82](#)
 - standaryzacja
 - w dwustopniowej analizie skupień [187](#)
 - statystyka Browna-Forsythe'a
 - w jednoczynnikowej analizie wariancji ANOVA [121](#)
 - statystyka Cochrańa
 - w tabelach krzyżowych [94](#)
 - statystyka Durбина-Watsona
 - w procedurze Regresja liniowa [150](#)
 - statystyka Mantela-Haenszela
 - w tabelach krzyżowych [94](#)
 - statystyka R
 - w procedurze Regresja liniowa [150](#)
 - w procedurze Średnie [100](#)
 - statystyka Welcha
 - w jednoczynnikowej analizie wariancji ANOVA [121](#)
 - statystyki dla proporcji kolumnowych

- statystyki dla proporcji kolumnowych (*kontynuacja*)
 - w tabelach krzyżowych [96](#)
- statystyki F
 - w modelach liniowych [141](#)
- Statystyki ilorazowe
 - statystyki [237](#), [239](#)
- statystyki opisowe
 - porządek wyświetlania [88](#)
 - statystyki [88](#)
 - w częstościach [86](#)
 - w dwustopniowej analizie skupień [188](#)
 - w eksploracji [91](#)
 - w podsumowaniach obserwacji [98](#)
 - w statystykach ilorazowych [237](#), [239](#)
 - w statystykach opisowych [87](#)
 - zapisywanie statystyk z [87](#)
- Statystyki opisowe
 - dotatkowe właściwości komendy [89](#)
- statystyki z
 - w statystykach opisowych [87](#)
 - zapisywanie jako zmienne [87](#)
- suma
 - w częstościach [86](#)
 - w kostkach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
 - w statystykach opisowych [88](#)
- suma kwadratów
 - w OML [124](#)
- symulacja
 - analiza czułości [258](#)
 - analiza what-if [258](#)
 - dopasowywanie rozkładu [255](#)
 - dostosowywanie dopasowania rozkładu [257](#)
 - edytor równań [253](#)
 - funkcja gęstości prawdopodobieństwa [260](#)
 - funkcja skumulowanego rozkładu [260](#)
 - korelacje między danymi wejściowymi [258](#)
 - Kreator symulacji [252](#)
 - kryteria zatrzymywania [259](#)
 - obsługiwane modele [252](#)
 - określanie modelu [252](#)
 - opcje wykresów [266](#)
 - percentyle rozkładów zmiennych przewidywanych [261](#)
 - ponowne dopasowanie rozkładów do nowych danych [263](#)
 - próbkowanie wartości krańcowych [259](#)
 - tworzenie nowych danych wejściowych [254](#)
 - tworzenie planu symulacji [249–251](#)
 - uruchamianie planu symulacji [251](#), [263](#)
 - wykresy interaktywne [266](#)
 - wykresy rozrzutu [261](#)
 - wykresy skrzynkowe [261](#)
 - wykresy tornado [261](#)
 - wyniki [260](#), [261](#)
 - wyniki dopasowywania rozkładu [257](#)
 - wyświetlaj formaty dla zmiennych przewidywanych i danych wejściowych [261](#)
 - zapisz plan symulacji [263](#)
 - zapisz symulowane dane [263](#)
- symulacja Monte Carlo [249](#)

Ś

- średnia
 - podgrupa [100](#), [102](#)
 - w częstościach [86](#)
 - w eksploracji [91](#)
 - w jednoczynnikowej analizie wariancji ANOVA [121](#)
 - w kosztach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
 - w raporcie z podsumowaniami w wierszach [224](#)
 - w Raport: Podsumowania w kolumnach [226](#)
 - w statystykach opisowych [88](#)
 - z wielu kolumn raportu [226](#)
- Średnia
 - w statystykach ilorazowych [237](#), [239](#)
- średnia geometryczna
 - w kosztach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
- średnia harmoniczna
 - w kosztach OLAP [103](#)
 - w podsumowaniach obserwacji [98](#)
 - w procedurze Średnie [100](#)
- Średnia przycięta
 - w eksploracji [91](#)
- średnia ważona
 - w statystykach ilorazowych [237](#), [239](#)
- Średnie
 - opcje [100](#)
 - statystyki [100](#)
- średnie grupowe [100](#), [102](#)
- średnie w podgrupach [100](#), [102](#)

T

- t-kwadrat Hotellinga
 - w analizie rzetelności [228](#), [229](#)
- T2 Tamhane'a
 - w jednoczynnikowej analizie wariancji ANOVA [120](#)
 - w OML [128](#)
- T3 Dunnetta
 - w jednoczynnikowej analizie wariancji ANOVA [120](#)
 - w OML [128](#)
- tabela klasyfikacji
 - w analizie najbliższego sąsiedztwa [177](#)
- tabela krzyżowa
 - w tabelach krzyżowych [93](#)
 - wielokrotne odpowiedzi [221](#)
- tabele częstości
 - w częstościach [85](#)
 - w eksploracji [91](#)
- tabele kontyngencji [93](#)
- Tabele krzyżowe
 - warstwy [94](#)
 - zawartość komórek [96](#)
 - zgrupowane wykresy słupkowe [94](#)
 - zmiennie sterujące [94](#)
- tabele krzyżowe wielokrotnych odpowiedzi [221](#)
- Tabele krzyżowe wielokrotnych odpowiedzi
 - brakujące wartości [222](#)
 - definiowanie zakresów wartości [222](#)
 - łączenie zmiennych poprzez zestawy odpowiedzi [222](#)
 - odsetka komórek [222](#)

Tabele krzyżowe wielokrotnych odpowiedzi (*kontynuacja*)
 procent na podstawie obserwacji [222](#)
 procent na podstawie odpowiedzi [222](#)

Tabele przestawne
 formaty [97](#)
 statystyki [94](#)
 ukrywanie tabel [93](#)

tau Goodmana i Kruskala
 w tabelach krzyżowych [94](#)

tau Kruskala
 w tabelach krzyżowych [94](#)

tau-b
 w tabelach krzyżowych [94](#)

tau-b Kendalla
 w korelacjach parami [134](#)
 w tabelach krzyżowych [94](#)

tau-c
 w tabelach krzyżowych [94](#)

tau-c Kendalla
 w tabelach krzyżowych [94](#)

Test addytywności Tukey'a
 w analizie rzetelności [228](#), [229](#)

test b Tukey'a
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test chi-kwadrat
 Testy nieparametryczne dla jednej próby [199](#), [200](#)

Test dla dwóch prób niezależnych
 definiowanie grup [215](#)
 typy testów [214](#)
 zmienne grupujące [215](#)

Test dokładny Fishera
 w tabelach krzyżowych [94](#)

test Dunnetta
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test dwumianowy
 dodatkowe właściwości komendy [210](#)
 dychotomie [209](#)
 Testy nieparametryczne dla jednej próby [199](#)

Test dwumianowy
 brakujące wartości [210](#)
 opcje [210](#)
 statystyki [210](#)

test Friedmana
 testy nieparametryczne dla prób zależnych [205](#)
 w testach dla kilku prób zależnych [218](#)

test Gabriela porównań parami
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test Gamesa i Howella porównań parami
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test jednorodności brzegowej
 testy nieparametryczne dla prób zależnych [205](#)
 w testach dla dwóch prób zależnych [215](#)

test Kołmogorowa-Smirnowa
 Testy nieparametryczne dla jednej próby [199](#), [200](#)

Test Kołmogorowa-Smirnowa
 Smirnowa
 test Lillieforsa [200](#)

Test Kołmogorowa-Smirnowa dla jednej próby
 brakujące wartości [213](#)

Test Kołmogorowa-Smirnowa dla jednej próby (*kontynuacja*)
 dodatkowe właściwości komendy [214](#)
 opcje [213](#)
 rozkład testowy [212](#)
 statystyki [213](#)
 test Lillieforsa [213](#)
 Test Lillieforsa [212](#)

Test Levene'a
 w eksploracji [91](#)

test Levene'a
 w jednoczynnikowej analizie wariancji ANOVA [121](#)

test Lillieforsa
 w eksploracji [91](#)

Test Lillieforsa [212](#)

Test linii równoległych
 w regresji porządkowej [153](#)

test M Boxa
 w analizie dyskryminacyjnej [178](#)

test McNemara
 testy nieparametryczne dla prób zależnych [205](#)
 Testy nieparametryczne dla prób zależnych [206](#)
 w tabelach krzyżowych [94](#)
 w testach dla dwóch prób zależnych [215](#)

test mediany
 w testach dla dwóch prób niezależnych [216](#)

Test przebiegu
 dodatkowe właściwości komendy [211](#)

test Q Cochrańa
 testy nieparametryczne dla prób zależnych [205](#)
 Testy nieparametryczne dla prób zależnych [207](#)

test Scheffégo
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test serii
 Testy nieparametryczne dla jednej próby [199](#), [201](#)

Test serii
 brakujące wartości [211](#)
 opcje [211](#)
 punkty podziału [210](#), [211](#)
 statystyki [211](#)

test sferyczności Bartletta
 w analizie czynnikowej [182](#)

test Shapiro-Wilka
 w eksploracji [91](#)

test skrajnych reakcji Moseśa
 w testach dla dwóch prób niezależnych [214](#)

Test Studenta-Newmana-Keulsa
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test t
 w teście t dla jednej próby [117](#)
 w teście t dla prób niezależnych [114](#)
 w teście t dla prób zależnych [115](#)

test t dla dwóch prób
 w teście t dla prób niezależnych [114](#)

Test t dla jednej próby
 dodatkowe właściwości komendy [116](#), [118](#)

Test T dla jednej próby
 braki danych [117](#)
 opcje [117](#)
 przedziały ufności [117](#)

Test t dla prób niezależnych
 brakujące wartości [115](#)
 definiowanie grup [115](#)

Test t dla prób niezależnych (*kontynuacja*)
 opcje [115](#)
 przedziały ufności [115](#)
 zmienne grupujące [115](#)
 zmienne łańcuchowe [115](#)

Test t dla prób zależnych
 brakujące wartości [116](#)
 opcje [116](#)
 wybór zmiennych zależnych [115](#)

test t Sidaka
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test t Studenta [114](#)

test t Wallera-Duncana
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test wielokrotnych rozstępów Duncana
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)

test znaków
 testy nieparametryczne dla prób zależnych [205](#)
 w testach dla dwóch prób zależnych [215](#)

test znaków rangowanych Wilcoxon
 Testy nieparametryczne dla jednej próby [199](#)
 testy nieparametryczne dla prób zależnych [205](#)
 w testach dla dwóch prób zależnych [215](#)

testy dla dwóch prób niezależnych
 brakujące wartości [215](#)
 dodatkowe właściwości komendy [215](#)
 opcje [215](#)
 statystyki [215](#)

Testy dla dwóch prób zależnych
 braki danych [216](#)
 dodatkowe właściwości komendy [216](#)
 opcje [216](#)
 statystyki [216](#)
 typy testów [216](#)

testy dla kilku prób niezależnych
 braki danych [217](#)
 definiowanie zakresu [217](#)
 dodatkowe właściwości komendy [218](#)
 opcje [217](#)
 statystyki [217](#)
 typy testów [217](#)
 zmienne grupujące [217](#)

testy dla kilku prób zależnych
 dodatkowe właściwości komendy [218](#)
 statystyki [218](#)
 typy testów [218](#)

testy jednorodności wariancji
 w jednoczynnikowej analizie wariancji ANOVA [121](#)

testy liniowości
 w procedurze Średnie [100](#)

testy nieparametryczne
 chi-kwadrat [208](#)
 Test dla dwóch prób niezależnych [214](#)
 Test Kołmogorowa-Smirnowa dla jednej próby [212](#)
 Test serii [210](#)
 Testy dla dwóch prób zależnych [215](#)
 testy dla kilku prób niezależnych [216](#)
 testy dla kilku prób zależnych [218](#)

Testy nieparametryczne dla jednej próby
 test chi-kwadrat [200](#)

Testy nieparametryczne dla jednej próby (*kontynuacja*)
 test dwumianowy [199](#)
 test Kołmogorowa-Smirnowa [200](#)
 test serii [201](#)
 zmienne [198](#)

Testy nieparametryczne dla prób niezależnych
 Karta Zmienne [203](#)

Testy nieparametryczne dla prób zależnych
 pola [205](#)
 test McNemara [206](#)
 test Q Cochrańa [207](#)

testy niezależności
 chi-kwadrat [94](#)

testy normalności
 w eksploracji [91](#)

testy t dla zmiennych zależnych
 w teście t dla prób zależnych [115](#)

tolerancja
 w procedurze Regresja liniowa [150](#)

Tolerancja
 regresja liniowa [150](#)

tytuły
 w kostkach OLAP [104](#)

U

U Mann-Whitney'a
 w testach dla dwóch prób niezależnych [214](#)

Uogólniony model liniowy (OLM)
 model [124](#)
 suma kwadratów [124](#)
 testy post hoc [128](#)
 wykresy profili [126](#)
 zapisywanie macierzy [131](#)
 zapisywanie zmiennych [131](#)

Uogólniony model liniowy jednej zmiennej
 kontrasty [125](#), [126](#)

ustawienia strony
 w raportach w wierszach [225](#)
 w raportach z podsumowaniami w kolumnach [227](#)

V

V
 w tabelach krzyżowych [94](#)

V Craméra
 w tabelach krzyżowych [94](#)

V Rao
 w analizie dyskryminacyjnej [179](#)

W

W Kendalla
 w testach dla kilku prób zależnych [218](#)

Wald
 w procedurze Proporcje dla prób zależnych [109](#)

Walda
 w procedurze Proporcje dla jednej próby [106](#)
 w procedurze Proporcje dla prób niezależnych [112](#)
 w procedurze Proporcje dla prób zależnych [109](#)

Walda (ze skorygowaną ciągłością)
 w procedurze Proporcje dla jednej próby [106](#)
 w procedurze Proporcje dla prób niezależnych [112](#)

Walda (ze skorygowaną ciągłością) (*kontynuacja*)
 w procedurze Proporcje dla prób zależnych [109](#)
 Walda H0
 w procedurze Proporcje dla prób niezależnych [112](#)
 Walda H0 (ze skorygowaną ciągłością)
 w procedurze Proporcje dla prób niezależnych [112](#)
 wariancja
 w częstościach [86](#)
 w eksploracji [91](#)
 w kostkach OLAP [103](#)
 w podsumowaniach obserwacji [98](#)
 w procedurze Średnie [100](#)
 w raporcie z podsumowaniami w wierszach [224](#)
 w Raport: Podsumowania w kolumnach [226](#)
 w statystykach opisowych [88](#)
 warstwy
 w tabelach krzyżowych [94](#)
 wartości odstające
 w dwustopniowej analizie skupień [187](#)
 w eksploracji [91](#)
 w regresji liniowej [148](#)
 wartości przewidywane
 zapisywanie w estymacji krzywej [166](#)
 zapisywanie w procedurze Regresja liniowa [148](#)
 wartości skrajne
 w eksploracji [91](#)
 wartości standaryzowane
 w statystykach opisowych [87](#)
 wartości własne
 w analizie czynnikowej [182](#), [183](#)
 w procedurze Regresja liniowa [150](#)
 wartości wpływu
 w OML [131](#)
 w procedurze Regresja liniowa [148](#)
 wartość STRESS
 w skalowaniu wielowymiarowym [234](#)
 ważność predyktorów
 modele liniowe [144](#)
 ważność zmiennych
 w analizie najbliższego sąsiedztwa [176](#)
 Ważona Kappa
 drukowanie [233](#)
 kryteria [233](#)
 przykład [232](#)
 statystyki [233](#)
 Statystyki [232](#)
 tabela krzyżowa [233](#)
 ważona kappa Cohena [233](#)
 Ważona kappa Cohena [232](#), [233](#)
 ważne wartości przewidywane
 w OML [131](#)
 widok modelu
 w analizie najbliższego sąsiedztwa [174](#)
 wielkość efektu
 w przypadku testu t dla prób zależnych [115](#)
 w teście t dla prób niezależnych [114](#)
 Wielokrotne F Ryana-Einota-Gabriela-Welscha
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)
 Wielokrotne odpowiedzi
 dodatkowe właściwości komendy [222](#)
 wielokrotne porównania
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 wielokrotne porównania post hoc [120](#)
 wielokrotne R
 w procedurze Regresja liniowa [150](#)
 wielokrotny rozstęp Ryana-Einota-Gabriela-Welscha
 w jednoczynnikowej analizie wariancji ANOVA [120](#)
 w OML [128](#)
 wizualizacja
 modele skupień [189](#)
 wskaźnik koncentracji
 w statystykach ilorazowych [237](#), [239](#)
 wskaźnik regresyjności (PRD)
 w statystykach ilorazowych [237](#), [239](#)
 współczynnik alfa
 w analizie rzetelności [228](#), [229](#)
 współczynnik dyspersji (COD)
 w statystykach ilorazowych [237](#), [239](#)
 współczynnik kontyngencji
 w tabelach krzyżowych [94](#)
 współczynnik korelacji r
 w korelacjach parami [134](#)
 w tabelach krzyżowych [94](#)
 współczynnik korelacji rang
 w korelacjach parami [134](#)
 współczynnik korelacji Spearmana
 w korelacjach parami [134](#)
 w tabelach krzyżowych [94](#)
 współczynnik korelacji wewnątrzklasowej (ICC)
 w analizie rzetelności [229](#)
 współczynnik Kuder-Richardsona 20 (KR20)
 w analizie rzetelności [229](#)
 współczynnik niepewności
 w tabelach krzyżowych [94](#)
 Współczynnik zgodności Kendalla (W)
 testy nieparametryczne dla prób zależnych [205](#)
 współczynnik zmienności (COV)
 w statystykach ilorazowych [237](#), [239](#)
 współczynniki beta
 w procedurze Regresja liniowa [150](#)
 współczynniki regresji
 w procedurze Regresja liniowa [150](#)
 wybór funkcji
 w analizie najbliższego sąsiedztwa [176](#)
 wybór k
 w analizie najbliższego sąsiedztwa [176](#)
 wybór k i funkcji
 w analizie najbliższego sąsiedztwa [177](#)
 wykaz obserwacji [97](#)
 Wykres bąbelkowy
 Metaanaliza: regresja [78](#)
 Wykres Galbraitha [75](#)
 Wykres L' Abb'a [75](#)
 Wykres lasu [75](#)
 Wykres lejkowy [75](#)
 wykres przestrzeni właściwości
 w analizie najbliższego sąsiedztwa [174](#)
 wykres rozrzutu
 podczas symulacji [261](#)
 wykresy
 etykiety obserwacji [164](#)
 w analizie ROC [245](#)
 w krzywej ROC [248](#)
 wykresy prawdopodobieństwa [240](#), [242](#)
 Wykresy
 Metaanaliza [75](#)
 wykresy cząstkowe

wykresy cząstkowe (*kontynuacja*)
w regresji liniowej [148](#)
wykresy kołowe
w częstościach [87](#)
wykresy ładunków
w analizie czynnikowej [184](#)
wykresy todyga-i-liście
w eksploracji [91](#)
wykresy normalne bez trendu
w eksploracji [91](#)
wykresy prawdopodobieństwa
K-K [242](#)
P-P [240](#)
wykresy prawdopodobieństwa K-K [242](#)
wykresy prawdopodobieństwa P-P [240](#)
wykresy profili
w OML [126](#)
wykresy rozrzut-poziom
w eksploracji [91](#)
wykresy rozrzutu
w regresji liniowej [148](#)
wykresy skrzynkowe
podczas symulacji [261](#)
porównywanie poziomów czynników [91](#)
porównywanie zmiennych [91](#)
w eksploracji [91](#)
wykresy słupkowe
w częstościach [87](#)
wykresy soplekowe
w hierarchicznej analizie skupień [195](#)
wykresy tornado
podczas symulacji [261](#)
Wynik [107](#)
wyniki binarne
surowe dane wielkości efektu [37](#)
Wstępnie obliczona wielkość efektu [37](#)
wzmocnienie
w modelach liniowych [140](#)

Z

Z Kołmogorowa-Smirnowa
w testach dla dwóch prób niezależnych [214](#)
w teście Kołmogorowa-Smirnowa dla jednej próby
[212](#)
Zakładka Galbraith
Metaanaliza (binarnie) [80](#)
Metaanaliza binarna wielkości efektu [80](#)
Metaanaliza ciągłej wielkości efektu [80](#)
Metaanaliza ilościowa [80](#)
zakres
w kostkach OLAP [103](#)
w statystykach opisowych [88](#)
zbieżność
w analizie czynnikowej [183](#), [184](#)
w analizie skupień metodą k-średnich [197](#)
Zdefiniuj zestawy wielokrotnych odpowiedzi
dychotomie [219](#)
kategorie [219](#)
ustaw etykiety [219](#)
ustaw nazwy [219](#)
zestawy
w modelach liniowych [142](#)
zestawy odpowiedzi wielokrotnych

zestawy odpowiedzi wielokrotnych (*kontynuacja*)
książka kodowa [82](#)
Zgodność między oceniającymi [229](#)
zmienna filtrująca
w regresji liniowej [148](#)
zmiennie sterujące
w tabelach krzyżowych [94](#)

