

*IBM SPSS Advanced Statistics 29*

**IBM**

**Uwaga**

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 141.

**Informacje o produkcie**

Niniejsze wydanie publikacji dotyczy wersji 29, wydania 0, modyfikacji 2 produktu IBM® SPSS Statistics oraz wszystkich jego późniejszych wydań i modyfikacji, aż do odwołania w nowych wydaniach publikacji.

© Copyright International Business Machines Corporation .

---

# Spis treści

<b>Rozdział 1. Statystyki zaawansowane.....</b>	<b>1</b>
Wprowadzenie do modułu Advanced Statistics.....	1
Analiza OML wielu zmiennych.....	1
Model OML wielu zmiennych.....	3
Kontrasty OML wielu zmiennych.....	4
Wykresy profili OML wielu zmiennych.....	5
Porównania post hoc wielu zmiennych OML.....	5
Oszacowane średnie brzegowe OML.....	6
OML: Zapisz.....	7
Opcje analizy OML wielu zmiennych.....	7
Dodatkowe właściwości komendy GLM.....	9
OML powtarzanych pomiarów.....	9
OML powtarzanych pomiarów – definiowanie czynników.....	12
Model OML powtarzanych pomiarów.....	12
Kontrasty OML powtarzanych pomiarów.....	13
Wykresy profili OML powtarzanych pomiarów.....	14
OML powtarzanych pomiarów-porównania post hoc.....	15
Oszacowane średnie brzegowe OML.....	16
OLM powtarzanych pomiarów: Zapisz.....	16
Opcje OML powtarzanych pomiarów.....	17
Dodatkowe właściwości komendy GLM.....	18
Analiza składowych wariacji.....	18
Składowe wariacji: Model.....	19
Składowe wariacji: Opcje.....	20
Składowe wariacji: Zapisz w nowym pliku.....	21
Dodatkowe właściwości komendy VARCOMP.....	22
Liniowe modele mieszane.....	22
Liniowe modele mieszane: Zmienne obiektowe i powtórzone.....	23
Liniowe modele mieszane: Efekty stałe.....	24
Liniowe modele mieszane: Efekty losowe.....	25
Oszacowanie liniowych modeli mieszanych.....	26
Liniowe modele mieszane: Statystyki.....	27
Liniowe modele mieszane: Szacowane średnie brzegowe.....	28
Liniowe modele mieszane: Zapisz.....	28
Liniowe modele mieszane-Eksport.....	29
Dodatkowe właściwości komendy MIXED.....	29
Uogólnione modele liniowe.....	29
Uogólnione modele liniowe: Odpowiedź.....	32
Uogólnione modele liniowe: Predyktory.....	33
Uogólnione modele liniowe: model.....	33
Uogólnione modele liniowe: Oszacowanie.....	34
Uogólnione modele liniowe: Statystyki.....	35
Uogólnione modele liniowe: Średnie EM.....	37
Uogólnione modele liniowe: Zapisz.....	38
Uogólnione modele liniowe: Eksportuj.....	39
Dodatkowe właściwości komendy GENLIN.....	40
Ogólne równania estymujące.....	40
Ogólne równania estymujące: Typ modelu.....	42
Ogólne równania estymujące: Odpowiedź.....	44
Ogólne równania estymujące: Predyktory.....	45
Ogólne równania estymujące: Model.....	45

Ogólne równania estymujące: Oszacowanie.....	46
Ogólne równania estymujące: Statystyki.....	48
Ogólne równania estymujące: Średnie EM.....	49
Ogólne równania estymujące: Zapisz.....	50
Ogólne równania estymujące: Eksportuj.....	50
Dodatkowe właściwości komendy GENLIN.....	51
Uogólnione liniowe modele mieszane.....	51
Otrzymywanie uogólnionego liniowego modelu mieszanego.....	54
Zmienna przewidywana .....	54
Efekty stałe .....	56
Efekty losowe .....	57
Waga i przesunięcie .....	59
Ogólne opcje budowania .....	59
Oszacowanie .....	60
Oszacowane średnie .....	60
Zapisz .....	61
Eksportuj .....	62
Widok modelu .....	62
Analiza logliniowa: Wybór modelu.....	66
Analiza logliniowa: Definiuj zakres.....	67
Analiza logliniowa: Model.....	67
Opcje analizy logliniowej; wybór modelu.....	68
Dodatkowe właściwości komendy HILOGLINEAR.....	68
Ogólna analiza log-liniowa.....	68
Ogólna analiza logliniowa: Model.....	69
Ogólna analiza logliniowa: Opcje.....	70
Ogólna analiza logliniowa: Zapisz.....	70
Dodatkowe właściwości komendy GENLOG.....	71
Analiza logliniowa Logit.....	71
Analiza logliniowa Logit: Model.....	72
Analiza logliniowa Logit: Opcje.....	73
Analiza logliniowa Logit: Zapisz.....	73
Dodatkowe właściwości komendy GENLOG.....	74
Tabele trwania życia.....	74
Tabele trwania życia: Definiuj zakres.....	75
Opcja tabel trwania życia.....	75
Dodatkowe właściwości komendy SURVIVAL.....	76
Modele parametrycznego czasu przyspieszonego niepowodzenia.....	76
Modele parametrycznego czasu przyspieszonego niepowodzenia: kryteria.....	77
Modele parametrycznego czasu przyspieszonego niepowodzenia: Model.....	77
Modele parametrycznego czasu przyspieszonego niepowodzenia: Oszacowanie.....	78
Modele parametrycznego czasu niepowodzenia akceleracji: Drukuj.....	79
Modele parametrycznego przyspieszonego czasu awarii: przewidywanie.....	80
Parametryczne modele przyspieszonego czasu awarii: wykres.....	81
Parametryczne modele czasu przyspieszonego niepowodzenia: Eksportuj.....	81
Przeżycie AFT Definiuj zdarzenia dla zmiennych statusu.....	82
Parametryczne modele przyspieszonego czasu awarii: wybierz kategorię.....	82
Parametryczne współużytkowane modele Frailty.....	83
Parametryczne współużytkowane modele Frailty: Kryteria.....	83
Parametryczne współużytkowane modele frailty: Model.....	84
Parametryczne współużytkowane modele Frailty: Oszacowanie.....	85
Parametryczne współużytkowane modele frailty: Drukuj.....	86
Parametryczne współużytkowane modele frailty: przewidywanie.....	87
Parametryczne współużytkowane modele frailty: Wykres.....	88
Parametryczne współużytkowane modele Frailty: Eksportuj.....	88
Parametryczne współużytkowane modele frailty: Definiuj zdarzenia.....	89
Parametryczne współużytkowane modele frailty-przykłady.....	89
Parametryczne współużytkowane modele frailty-studium przypadku dla danych okresowych.....	91

Analiza przeżycia Kaplana-Meiera.....	93
Kaplan-Meier: Definiuj zdarzenie dla zmiennej statusu.....	94
Kaplan-Meier: Porównaj poziomy czynnika.....	94
Kaplan-Meier: Zapisz zmienne wynikowe.....	95
Kaplan-Meier: Opcje.....	95
Dodatkowe właściwości komendy KM.....	95
Analiza metodą regresji Coxa.....	95
Regresja Coxa: Definiuj współzmiennie jakościowe.....	96
Wykresy regresji Coxa.....	97
Regresja Coxa: Zapisz zmienne wynikowe.....	97
Regresja Coxa: Opcje.....	98
Regresja Coxa: Definiuj zdarzenie dla zmiennej statusu.....	98
Dodatkowe właściwości komendy COXREG.....	98
Obliczanie współzmiennych zależnych od czasu.....	99
Obliczanie współzmiennych zależnych od czasu.....	99
Schematy kodowania zmiennych jakościowych.....	100
Odchylenie.....	100
Proste.....	100
Helmerta.....	101
Różnica.....	101
Wielomianowe.....	101
Powtórzone.....	102
Specjalne.....	102
Wskaźnik.....	103
Struktury kowariancji.....	103
Statystyki Bayesa.....	107
Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Normalny.....	108
Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Dwumianowy.....	112
Wnioskowanie bayesowskie dla jednej próby: rozkład Poissona.....	114
Wnioskowanie bayesowskie dotyczące testów dla prób zależnych: Normalny.....	116
Bayesowskie testy dla prób niezależnych.....	118
Wnioskowanie bayesowskie dotyczące korelacji Pearsona.....	122
Wnioskowanie bayesowskie dotyczące modeli regresji liniowej.....	124
Bayesowska jednoczynnikowa ANOVA.....	129
Bayesowskie modele logliniowe.....	132
Bayesowska jednoczynnikowa ANOVA z powtarzаныmi pomiarami: Modele.....	135
Regresja grzbietowa jądra.....	137
Parametry jądra.....	139
Regresja grzbietowa jądra: Opcje.....	139
<b>Uwagi.....</b>	<b>141</b>
Znaki towarowe.....	142
<b>Indeks.....</b>	<b>145</b>



---

# Rozdział 1. Statystyki zaawansowane

Następujące funkcje dotyczące statystyk zaawansowanych są dostępne w module SPSS Statistics Standard Edition lub module Statystyki zaawansowane.

## Wprowadzenie do modułu Advanced Statistics

---

Opcja Statystyki zaawansowane obejmuje procedury, które oferują bardziej zaawansowane opcje modelowania, niż są dostępne za pośrednictwem opcji Statistics Base.

- Procedura OML Wielu zmiennych stanowi rozszerzenie ogólnego modelu liniowego procedury OML Jednej zmiennej, dopuszczając stosowanie więcej niż jednej zmiennej zależnej. Kolejne rozszerzenie, OML Powtarzanych pomiarów, umożliwia uwzględnienie powtarzanych pomiarów wielu zmiennych zależnych.
- Analiza składowych wariancji to specjalistyczne narzędzie do dekompozycji zmienności zmiennej zależnej na składowe stałe i losowe.
- Procedura Liniowe modele mieszane rozszerza uogólniony model liniowy w taki sposób, że dane mogą wykazywać skorelowaną i niestałą zmienność. Liniowy model mieszany zapewnia zatem elastyczność modelowania nie tylko średnich danych, lecz także lub wariancji i kowariancji.
- Procedura Uogólnione modele liniowe (GZLM) rozluźnia założenie o normalnym rozkładzie składnika błędu i wymaga jedynie, by zmienna zależna była związana zależnością liniową z predyktorami poprzez transformację lub funkcję łączenia. Procedura Uogólnione równania estymujące (GEE) rozszerza procedurę GZLM o możliwość uwzględnienia powtarzanych pomiarów.
- Ogólna analiza logliniowa umożliwia dopasowywanie modeli do danych o liczebności klasyfikowanych krzyżowo, a funkcja Analiza logliniowa: Wybór modelu pomaga w wyborze między modelami.
- Analiza logliniowa Logit umożliwia dopasowywanie modeli logliniowych w celu analizy relacji między jakościową zmienną zależną a jednym lub większą liczbą predyktorów jakościowych.
- Do analizy przeżycia można wykorzystać procedurę Tabele trwania życia, która pozwala na badanie rozkładu zmiennych odzwierciedlających czas do zajścia zdarzenia, potencjalnie na różnych poziomach zmiennej czynnika; procedurę Analiza przeżycia Kaplana-Meiera która pozwala na badanie rozkładu zmiennych odzwierciedlających czas do zajścia zdarzenia, potencjalnie na różnych poziomach zmiennej czynnika lub wygenerowanie osobnych analiz na różnych poziomach zmiennej warstwującej; i procedurę Regresja Coxa, która służy do modelowania czasu do zajścia określonego zdarzenia na podstawie wartości danych współzmiennych.

## Analiza OML wielu zmiennych

---

Procedura OML wielu zmiennych umożliwia analizę regresji oraz analizy wariancji wielu zmiennych zależnych względem dowolnej liczby czynników i/lub współzmiennych. Czynniki dzielą populację na grupy. Przy użyciu tej procedury ogólnego modelu liniowego można sprawdzić hipotezę zerową dotyczącą wpływu zmiennych-czynników na średnie różnych grup połączonego rozkładu łącznego zmiennych zależnych. Sprawdzać można interakcje zachodzące między poszczególnymi czynnikami, jak również wpływy poszczególnych czynników. 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. W modelu wielu zmiennych sumy kwadratów wynikające z wpływów w modelu i sumy kwadratów błędów mają postać macierzy, a nie postać skalarną, taką jak w analizie jednej zmiennej. Są to tzw. macierze SSCP (sums-of-squares and cross-products), czyli macierze sum kwadratów i iloczynów wektorowych. Jeśli określona jest więcej niż jedna zmienna zależna, dostępna jest analiza wariancji wielu zmiennych przy użyciu śladu Pillai, lambdy Wilksa, śladu Hotellinga i kryterium największego pierwiastka Roya z przybliżoną statystyką  $F$ , a także analiza wariancji

jednej zmiennej dla każdej ze zmiennych zależnych. Analiza OML wielu zmiennych 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. Testy wielokrotnych porównań post hoc są wykonywane oddzielnie dla każdej zmiennej zależnej.

Reszty, wartości przewidywane, Odległość Cooka i wartości wpływu można zapisać jako nowe zmienne w pliku danych dla celów sprawdzenia założeń. Dostępna jest także macierz SSCP reszt, która jest macierzą kwadratową sumy kwadratów i iloczynów wektorowych reszt, macierz kowariancji reszt, która jest macierzą SSCP podzieloną przez stopnie swobody reszt, a także macierz korelacji reszt, która jest zestandaryzowaną postacią macierzy kowariancji reszt.

W polu Waga WNK można określić zmienną używaną w celu nadania obserwacjom różnych wag podczas analizy metodą ważonych najmniejszych kwadratów (WNK), co kompensuje różne poziomy dokładności pomiarów.

**Przykład.** Producent wyrobów z tworzyw sztucznych mierzy trzy właściwości folii plastikowej: odporność na rozdarcie, połysk i przezroczystość. Próbuje trzech szybkości wytłaczania i dwóch różnych ilości dodatku do tworzywa, a wspomniane trzy właściwości są mierzone przy każdej kombinacji szybkości wytłaczania i ilości dodatku. Producent stwierdza, że zarówno szybkość wytłaczania, jak i ilość dodatku ma istotny wpływ na właściwości, ale interakcja między tymi dwoma czynnikami nie jest istotna.

**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, test wielokrotnego *F*Ryana-Einota-Gabriela-Welscha, test wielozakresowy Ryan-Einot-Gabriel-Welsch, test Studenta-Newmana-Keulsa, test uczciwie istotnej różnicy Tukey'a, test *b* Tukey'a, test Duncana, test GT2 Hochberga, test Gabriela, test *t* Wallera Duncana, test Dunnetta (jedno- i dwustronny), test T2 Tamhane'a, test T3 Dunnett'a, test Gamesa-Howella i test *C* Dunnetta. Statystyki opisowe: obserwowane średnie, odchylenia standardowe i liczebności wszystkich zmiennych zależnych we wszystkich komórkach; test jednorodności wariancji Levene'a; test *M* Boxa jednorodności macierzy kowariancji zmiennych zależnych; test sferyczności Bartletta.

**Wykresy.** Wykresy rozrzut-poziom, reszt i profili (interakcja).

Wymagania dotyczące danych dla OML wielu zmiennych

**Dane.** Zmienne zależne powinny być zmiennymi ilościowymi. Czynniki są jakościowe i mogą mieć wartości liczbowe lub tańcuchowe. Współzmiennie są zmiennymi ilościowymi powiązаныmi ze zmienną zależną.

**Założenia.** Dane zmiennych zależnych są losową próbą wektorów z normalnej populacji wielu zmiennych; w populacji macierze wariancji-kowariancji wszystkich komórek są takie same. Analiza wariancji jest odporna na odstępstwa od rozkładu normalnego, ale dane powinny być symetryczne. Do sprawdzenia założeń wykorzystać można testy jednorodności wariancji (w tym test *M* Boxa) oraz wykresy rozrzut-poziom. Można również analizować reszty i wykresy reszt.

**Procedury pokrewne.** Przed rozpoczęciem analizy wariancji należy użyć procedury eksploracji w celu zbadania danych. W przypadku jednej zmiennej zależnej należy użyć procedury OML jednej zmiennej. Jeśli te same zmienne zależne zmierzono kilkakrotnie dla każdego obiektu, należy użyć procedury OML powtarzanych pomiarów.

Uzyskiwanie tabel OML wielu zmiennych

1. Wybierz z menu następującą opcję:

**Analiza > Ogólny model liniowy > OML Wielu zmiennych...**

2. Wybierz co najmniej dwie zmienne zależne.

Opcjonalnie można określić Czynniki stałe, Współzmiennie oraz wartość WNK Waga.



## Model OML wielu zmiennych

**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 zagnieżdżany jest w drugim, ten z kolei w efekcie określonym 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.

## Kontrasty OML wielu zmiennych

Kontrasty służą do badania, czy poziomy efektu znacznie się od siebie różnią. Dla każdego czynnika w modelu można określić kontrast. Kontrasty przedstawiają liniowe kombinacje parametrów.

Testowanie hipotez jest oparte na hipotezie zerowej  $\mathbf{LBM} = 0$ , gdzie  $\mathbf{L}$  jest macierzą współczynników kontrastów,  $\mathbf{M}$  jest macierzą tożsamą, której wymiar jest równy liczbie zmiennych zależnych, a  $\mathbf{B}$  jest wektorem parametrów. Po określeniu kontrastu tworzona jest macierz  $\mathbf{L}$  w taki sposób, że kolumny macierzy reprezentujące czynniki odpowiadają kontrastowi. Pozostałe kolumny dopasowane są w sposób umożliwiający szacowanie macierzy  $\mathbf{L}$ .

Oprócz testu jednej zmiennej przy użyciu statystyk  $F$  oraz równoczesnych przedziałów ufności typu Bonferroniego opartych na rozkładzie  $t$  Studenta dla różnic kontrastów wszystkich zmiennych zależnych, dostępny jest również test wielowymiarowy, wykorzystujący ślad Pillai, lambdę Wilksa, ślad Hotellinga oraz kryterium największego pierwiastka Roya.

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.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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.

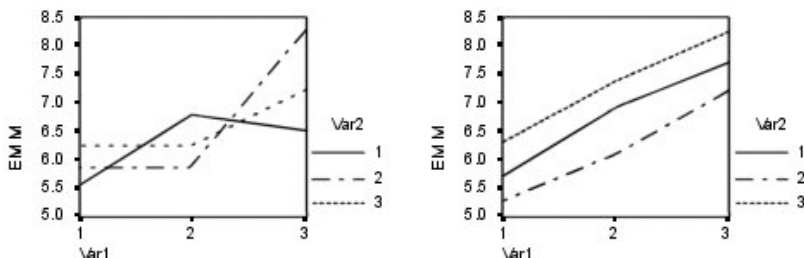
**Powtórzony.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz ostatniego) ze średnią poziomów następných.

**Wielomianowy.** Wybranie tego typu kontrastu powoduje porównanie efektu liniowego, efektu kwadratowego, sześciennego 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.

## Wykresy profili OML wielu zmiennych

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 są dostępne dla wykresów. Wykresy profili są tworzone dla każdej zmiennej zależnej.

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 1. 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.

## Porównania post hoc wielu zmiennych OML

**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 post hoc są wykonywane oddzielnie dla każdej zmiennej zależnej.

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 błędów doświadczenia na poziomie błędów 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).

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.

## 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**.

## 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.

- *Niestandaryzowane.* Wartość zmiennej zależnej przewidywana przez model.
- *Ważone.* Ważone niestandaryzowane 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 wpływu. Względny wpływ każdej obserwacji na dopasowanie modelu.

**Reszty.** Reszta niestandaryzowana 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 niestandaryzowane.

- *Niestandaryzowane.* Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- *Ważone.* Niestandaryzowane reszty ważone. Dostępne tylko wtedy, gdy wcześniej została wybrana zmienna WNK.
- *Standaryzowane.* Reszta podzielona przez oszacowanie jej odchylenia standardowego. Standaryzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Studentyzowane.* Reszta podzielona przez oszacowanie jej odchylenia standardowego, która zmienia się w zależności od obserwacji, w zależności od odległości wartości każdej obserwacji od zmiennych niezależnych od ich średnich. Czasem określane jako reszty studentyzowane wewnętrznie.
- *Usunięte.* Reszta dla obserwacji, gdy ta obserwacja jest wykluczona z obliczeń 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.

## Opcje analizy OML wielu zmiennych

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, odchyleń 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. Możliwe jest wyświetlanie **macierzy SSCP** oraz **macierzy SSCP reszt** hipotezy i błędów, a także wyświetlenie testu sferyczności Bartletta macierzy kowariancji reszt.

Zaznaczenie pola **Testowanie jednorodności** powoduje przeprowadzenie testu Levene'a na jednorodność wariancji dla każdej zmiennej zależnej, na wszystkich kombinacjach poziomów czynników międzyobiektowych, tylko dla takich czynników. Ponadto testy jednorodności obejmują test  $M$  Boxa jednorodności macierzy kowariancji zmiennych zależnych po wszystkich poziomach kombinacji 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.

## Wyświetl

### Statystyki opisowe

Powoduje wyświetlenie obserwowanych średnich, odchyłeń standardowych i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach.

### 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.

### Obserwowana moc

Powoduje wyświetlenie siły testu po ustaleniu hipotezy alternatywnej na podstawie obserwowanej wartości.

### Oszacowania parametrów

Powoduje wyświetlenie ocen parametrów, błędów standardowych, testów  $t$ , przedziałów ufności i obserwowanej siły każdego testu.

### Macierze SSCP

Wyświetla macierze SSCP hipotezy i błędu.

### Macierz SSCP reszt

Wyświetla macierz SSCP reszt hipotezy i błędu.

### Macierz transformacji

Wyświetla test sferyczności Bartletta macierzy kowariancji reszt.

### Testowanie jednorodności

Powoduje przeprowadzenie testu Levene'a na jednorodność wariancji dla każdej zmiennej zależnej, na wszystkich kombinacjach poziomów czynników międzyobiektowych, tylko dla takich czynników. Ponadto testy jednorodności obejmują test  $M$  Boxa jednorodności macierzy kowariancji zmiennych zależnych po wszystkich poziomach kombinacji czynników międzyobiektowych.

### Wykresy rozrzut-poziom

Wykresy te są przydatne podczas sprawdzania założeń równej wariancji. W przypadku braku czynników, opcja ta jest niedostępna.

### Wykres reszt

Powoduje utworzenie wykresu reszt rzeczywistych-szacowanych-standaryzowanych dla każdej zmiennej zależnej. Wykres jest przydatny podczas sprawdzania założeń równej wariancji.

### **Brak dopasowania**

Powoduje sprawdzenie, czy model nadaje się do prawidłowego opisu relacji zachodzącej między zmienną zależną a zmiennymi niezależnymi.

### **Ogólne funkcje estymowalne**

Pozwala na przeprowadzenie testów hipotez użytkownika dotyczących ogólnych funkcji estymowalnych. Wiersze w każdej macierzy współczynników kontrastu są liniowymi kombinacjami ogólnych funkcji estymowalnych.

### **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 GLM**

Te funkcje mogą dotyczyć jednej zmiennej, wielu zmiennych lub analizy z powtarzanymi pomiarami. 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 lub 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 powtarzanych pomiarów**

Procedura OML powtarzanych pomiarów służy do analizy wariancji w sytuacji, gdy ten sam pomiar każdej obserwacji lub każdego obiektu wykonywany jest kilkakrotnie. Czynniki międzyobiektowe, jeśli są określone, dzielą populację na grupy. Korzystając z tej procedury ogólnego modelu liniowego, można testować hipotezy zerowe dotyczące wpływów zarówno czynników międzyobiektowych, jak i wewnątrzobiektowych. Sprawdzać można interakcje zachodzące między poszczególnymi czynnikami, jak również wpływy poszczególnych czynników. Ponadto uwzględnić można wpływy stałych współzmiennych oraz ich interakcje z czynnikami międzyobiektowymi.

W przypadku pomiarów realizowanych jako test wielu zmiennych, zmienne zależne odpowiadają pomiarom więcej niż jednej zmiennej dla różnych poziomów czynników międzyobiektowych. Można na przykład u każdego z pacjentów zmierzyć zarówno tętno, jak i częstość oddechów w trzech różnych punktach czasowych.

Procedura OML powtarzanych pomiarów umożliwia analizę zarówno jednej, jak i wielu zmiennych w danych z powtarzanych pomiarów. Testować można zarówno modele zrównoważone, jak i niezrównoważone. Plan jest zrównoważony, gdy każda komórka modelu zawiera tę samą liczbę obserwacji. W modelu wielu zmiennych sumy kwadratów wynikające z wpływów w modelu i sumy kwadratów błędów mają postać macierzy, a nie postać skalarną, taką jak w analizie jednej zmiennej. Są to tzw. macierze SSCP (sums-of-squares and cross-products), czyli macierze sum kwadratów i iloczynów wektorowych. Analiza OML powtarzanych pomiarów umożliwia nie tylko testowanie hipotez, lecz również uzyskiwanie oszacowań parametrów.

Do testowania hipotez na czynnikach międzyobiektowych wykorzystuje się często 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. Dostępna jest także macierz SSCP reszt, która jest macierzą kwadratową sumy kwadratów i iloczynów wektorowych reszt, macierz kowariancji reszt, która jest macierzą SSCP podzieloną przez stopnie swobody reszt, a także macierz korelacji reszt, która jest zestandaryzowaną postacią macierzy kowariancji reszt.

W polu Waga WNK można określić zmienną używaną w celu nadania obserwacjom różnych wag podczas analizy metodą ważonych najmniejszych kwadratów (WNK), co kompensuje różne poziomy dokładności pomiarów.

**Przykład.** Dwanaścioro uczniów przypisano do grupy wysokiego i niskiego poziomu lęku na podstawie wyników testu oceniającego poziom lęku. Ocena lęku jest tutaj czynnikiem międzyobiektowym, ponieważ dzieli obiekty na grupy. Każdy z uczniów może czterokrotnie podchodzić do jednego zadania edukacyjnego, a w każdej próbie rejestrowana jest liczba błędów. Błędy popełnione w każdej próbie rejestrowane są w odrębnej zmiennej, a czynnik wewnątrzobiektowy (próba) jest zdefiniowany z użyciem czterech poziomów odpowiadających czterem próbom. Stwierdzono, że wpływ prób jest istotny, natomiast interakcja próba-lęk jest nieistotna.

**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 zakresu post hoc i wielokrotne porównania: najmniejsza istotna różnica, Bonferroniego, Sidaka, Scheffé'a, Ryana-Einota-Gabriela-Welscha wielokrotnych  $F$ , Ryana-Einota-Gabriela-Welscha wielu zakresów, Studenta-Newmana-Keulsa, rozsądna istotna różnica Tukey'a,  $b$  Tukey'a, Duncana, Hochberg'a GT2, Gabriela,  $t$  Wallera-Duncana, Dunnetta (jedno- i dwustronny), Tamhane'a T2, Dunnett'a T3, Gamesa-Howella i Dunnett'a C. Statystyki opisowe: obserwowane średnie, odchylenia standardowe i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach; test Levene'a jednorodności wariancji, test  $M$  Boxa jednorodności macierzy kowariancji zmiennych zależnych; i test sferyczności Bartletta.

**Wykresy.** Wykresy rozrzut-poziom, reszt i profili (interakcja).

Wymagania dotyczące danych dla OML powtarzanych pomiarów

**Dane.** Zmienne zależne powinny być zmiennymi ilościowymi. Czynniki międzyobektowe dzielą próbę na dyskretne podgrupy, np. mężczyzn i kobiety. Czynniki te są jakościowe i mogą mieć wartości liczbowe lub łańcuchowe. Czynniki wewnątrzobektowe definiuje się w oknie dialogowym OML Powtarzane pomiary: Definiuj czynniki. Współzmiennie są zmiennymi ilościowymi powiązаныmi ze zmienną zależną. W przypadku analizy z powtarzanymi pomiarami powinny one być stałe na każdym poziomie zmiennej wewnątrzobektowej.

Plik danych powinien zawierać zestaw zmiennych dla każdej grupy pomiarów obiektów. Ten zestaw powinien zawierać jedną zmienną dla każdego powtórzenia pomiaru w grupie. Czynniki wewnątrzobektowe powinny być zdefiniowane dla grupy i mieć liczbę poziomów równą liczbie powtórzeń. Na przykład pomiary wagi mogą być wykonywane w różnych dniach. Jeśli pomiary tej samej właściwości wykonywano w pięciu różnych dniach, to czynnik wewnątrzobektowy można zdefiniować jako *dzień* z pięcioma poziomami.



Gdy istnieje więcej niż jeden czynnik wewnątrzobiektywny, liczba pomiarów dla każdego obiektu jest równa iloczynowi liczby poziomów każdego czynnika. Na przykład, jeśli pomiary wykonywano trzy razy w ciągu dnia przez cztery dni, to łączna liczba pomiarów dla każdego obiektu wynosi 12. Czynniki wewnątrzobiektywne można zdefiniować jako *dzień*(4) i *godzina*(3).

**Założenia.** Do analizy z powtarzanymi pomiarami można podejść na dwa sposoby: jako do analizy jednej zmiennej albo wielu zmiennych.

W strategii jednej zmiennej (nazywanej także strategią wykresu podzielonego lub modelu mieszanego) zmienne zależne traktowane są jako odpowiedzi na poziomy czynników wewnątrzobiektywnych. Pomiary obiektu powinny być próbą z rozkładu normalnego wielu zmiennych, a macierze wariancji-kowariancji są takie same w komórkach utworzonych przez efekty (wpływy) międzyobiektywne. Obowiązują pewne założenia co do macierzy wariancji-kowariancji zmiennych zależnych. Ważność statystyki *F* stosowanej w strategii jednej zmiennej można zagwarantować, jeśli macierz wariancji-kowariancji ma postać okrągłą (Huynh i Mandeville, 1979).

Do przetestowania tego założenia można wykorzystać test sferyczności Mauchly'ego na macierzy wariancji-kowariancji ortonormalizowanej przekształconej zmiennej zależnej. Test Mauchly'ego jest wyświetlany automatycznie w przypadku analizy powtarzanych pomiarów. Dla małych prób siła tego testu jest niewielka. W przypadku dużych prób wynik testu może być istotny nawet wówczas, gdy wpływ odstępstw od sferyczności na wyniki jest mały. Duża istotność testu pozwala na przyjęcie hipotezy sferyczności. Jeśli jednak istotność testu jest mała, a założenie sferyczności wydaje się być naruszone, można skorygować licznik i mianownik stopni swobody w celu uzyskania ważności statystyki *F* jednej zmiennej. W procedurze OML powtarzanych pomiarów dostępne są trzy oszacowania tej korekty, tzw. **epsilon**. Liczba stopni swobody licznika i mianownika musi być pomnożona przez epsilon, a istotność statystyki *F* obliczona dla nowych stopni swobody.

W strategii wielu zmiennych przyjmuje się, że pomiary obiektu są próbą z rozkładu normalnego wielu zmiennych, a macierze wariancji-kowariancji są takie same w komórkach utworzonych przez efekty (wpływy) międzyobiektywne. Do sprawdzenia, czy macierze wariancji-kowariancji w komórkach są takie same, można zastosować test *M* Boxa.

**Procedury pokrewne.** Przed rozpoczęciem analizy wariancji należy użyć procedury eksploracji w celu zbadania danych. Jeśli *nie ma* pomiarów powtórzonych dla każdego obiektu, należy użyć procedury OML jednej zmiennej albo OML wielu zmiennych. Jeśli istnieją tylko dwa pomiary dla każdego obiektu (na przykład przed testem i po teście) i nie ma czynników międzyobiektywnych, można użyć procedury Test *t* dla prób zależnych.

Uzyskiwanie OML powtarzanych pomiarów

1. Wybierz z menu następującą opcję:

**Analiza > Ogólny model liniowy > Pomiary powtarzalne...**

2. Wpisz nazwę czynnika wewnątrzobiektywnego i jego liczbę poziomów.

3. Kliknij przycisk **Add**.

4. Powtórz te kroki dla każdego czynnika wewnątrzobiektywnego.

Aby zdefiniować czynniki pomiarów dla planu powtarzalnych pomiarów z podwójną wielowymiarowością:

5. Wpisz nazwę pomiaru.

6. Kliknij przycisk **Add**.

Po zdefiniowaniu wszystkich czynników i pomiarów:

7. Kliknij przycisk **Definiuj**.

8. Wybierz zmienną zależną odpowiadającą każdej kombinacji czynników wewnątrzobiektywnych (i opcjonalnie miar) na liście.

Do zmiany położenia zmiennych używaj strzałek w górę i w dół.

Aby zmodyfikować czynniki wewnątrzobiektywne, można ponownie otworzyć okno dialogowe OML. Powtarzane pomiary: Definiuj czynniki bez zamykania głównego okna dialogowego. Opcjonalnie można określić czynniki międzyobiektywne i współzmiennie.

## OML powtarzanych pomiarów – definiowanie czynników

Procedura OML: powtarzanych pomiarów analizuje grupy powiązanych zmiennych zależnych, które reprezentują różne pomiary tego samego atrybutu. W tym oknie dialogowym można zdefiniować co najmniej jeden czynnik wewnątrzobiektywne do użycia w OML powtarzanych pomiarów. Należy zwrócić uwagę na to, że istotna jest kolejność określania czynników wewnątrzobiektywnych. Każdy czynnik stanowi poziom wewnątrz poprzedniego czynnika.

Aby użyć powtarzanych pomiarów, należy poprawnie skonfigurować dane. W tym oknie dialogowym należy zdefiniować czynniki wewnątrzobiektywne. Te czynniki nie są zmiennymi istniejącymi w danych – są raczej czynnikami, które należy zdefiniować w tym miejscu.

**Przykład.** Załóżmy, że w badaniu utraty masy ciała masy kilku osób są mierzone każdego tygodnia przez pięć tygodni. W pliku danych każda osoba jest obiektem lub obserwacją. Masy z poszczególnych tygodni są rejestrowane w zmiennych *weight1*, *weight2* itd. Płeć każdej osoby jest rejestrowana w innej zmiennej. Masy – mierzone dla każdego obiektu wielokrotnie – mogą zostać pogrupowane poprzez zdefiniowanie czynnika wewnątrzobiektywnego. Czynniki ten może mieć nazwę *week* i można zdefiniować, że będzie miał pięć poziomów. W głównym oknie dialogowym zmienne *weight1*, ..., *weight5* są używane do przypisywania pięciu poziomów *tygodnia*. Zmienna w pliku danych, która grupuje mężczyzn i kobiety (*gender*), może zostać określona jako czynnik międzyobiektywne w celu sprawdzenia różnic między mężczyznami a kobietami.

**Miary.** Jeśli u obiektów każdorazowo testowana jest więcej niż jedna miara, należy zdefiniować miary. Na przykład tętno i szybkość oddychania powinny być mierzone u każdego obiektu każdego dnia przez tydzień. Te miary nie istnieją jako zmienne w pliku danych, ale są definiowane w tym miejscu. Model z więcej niż jedną miarą jest czasami nazywany modelem powtarzalnych pomiarów z podwójną wielowymiarowością.

## Model OML powtarzanych pomiarów

**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 międzyobiektywne.** Lista zawiera czynniki międzyobiektywne i współzmiennie.

**Model.** Model zależy od charakteru danych. Po wybraniu opcji **Niestandardowe**, można wybrać efekty i interakcje wewnątrzobiektywne oraz efekty i interakcje międzyobiektywne będące przedmiotem zainteresowania w czasie analizy.

**Suma kwadratów.** Metoda obliczania sum kwadratów dla modelu międzyobiektywnego. Dla międzyobiektywnych modeli zrównoważonych lub niezrównoważonych bez komórek z brakami danych najczęściej wykorzystywana jest metoda bazująca na sumie kwadratów typu III.

## 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.

## Kontrasty OML powtarzanych pomiarów

Kontrasty wykorzystywane są do sprawdzania różnic między poziomami czynnika międzyobiekowego. Dla każdego czynnika międzyobiekowego w modelu można określić kontrast. Kontrasty przedstawiają liniowe kombinacje parametrów.

Testowanie hipotez oparte jest na hipotezie zerowej  $\mathbf{LBM}=\mathbf{0}$ , gdzie  $\mathbf{L}$  jest macierzą współczynników kontrastów,  $\mathbf{B}$  jest wektorem parametrów, a  $\mathbf{M}$  – macierzą uśrednioną, odpowiadającą średniej transformacji dla zmiennej zależnej. Taką macierz transformacji można wyświetlić, wybierając opcję **Macierz transformacji** w oknie dialogowym Powtarzane pomiary: Opcje. Na przykład jeśli mamy cztery zmienne zależne, czteropoziomowy czynnik wewnątrzobiekowy oraz dla czynników

wewnątrzobiektywnych używane są kontrasty wielomianowe (ustawienie domyślne), to macierz **M** będzie miała postać (0,5 0,5 0,5 0,5)'. Po określeniu kontrastu tworzona jest macierz **L** w taki sposób, że kolumny macierzy reprezentujące czynniki międzyobiektywne odpowiadają kontrastowi. Pozostałe kolumny dopasowane są w sposób umożliwiający szacowanie macierzy **L**.

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.

Dla czynników wewnątrzobiektywnych należy wybrać inny kontrast niż **Brak**.

## Typy kontrastów

**Odchylenie.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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.

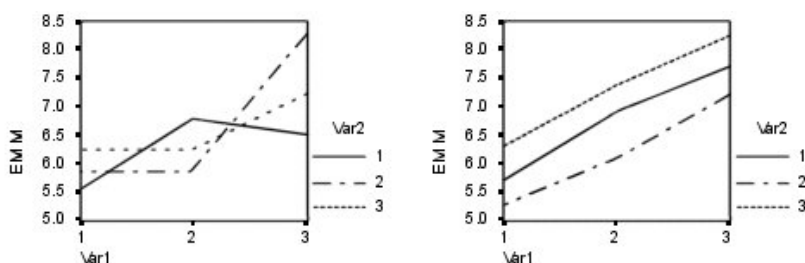
**Powtórzony.** Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz ostatniego) ze średnią poziomów następných.

**Wielomianowy.** Wybranie tego typu kontrastu powoduje porównanie efektu liniowego, efektu kwadratowego, sześciennego 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.

## Wykresy profili OML powtarzanych pomiarów

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 są dostępne dla wykresów. Wykresy profili są tworzone dla każdej zmiennej zależnej. Na wykresach profili mogą być używane zarówno czynniki międzyobiektywne, jak i wewnątrzobiektywne.

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.

## OML powtarzanych pomiarów-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 nie są dostępne, jeśli nie ma czynników międzyobiektowych, a testy wielokrotnych porównań post hoc są wykonywane dla średniej dla poziomów czynników wewnątrzobiektowych.

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).

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 harmoniczej 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.

## 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**.

## OLM powtarzanych pomiarów: 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.
- *Błąd standardowy.* Oszacowanie odchylenia standardowego średniej wartości zmiennej zależnej dla obserwacji, które mają takie same wartości zmiennych niezależnych.

**Diagnostyka.** Narzędzia do identyfikowania obserwacji z niezwykle kombinacją wartości zmiennych niezależnych oraz obserwacji, które mogą w znacznym stopniu wpływać na model. Dostępne są wartości odległości Cooka oraz niecentrowanych wartości wpływu.

- *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 wpływu. 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.

- *Niestandardyzowane*. Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- *Standardyzowane*. Reszta podzielona przez oszacowanie jej odchylenia standardowego. Standardyzowane reszty, znane także jako reszty Pearsona, mają średnią arytmetyczną 0 oraz odchylenie standardowe 1.
- *Studentyzowane*. Reszta podzielona przez oszacowanie jej odchylenia standardowego, która zmienia się w zależności od obserwacji, w zależności od odległości wartości każdej obserwacji od zmiennych niezależnych od ich średnich. Czasem określane jako reszty studentyzowane wewnętrznie.
- *Usunięte*. Reszta dla obserwacji, gdy ta obserwacja jest wykluczona z obliczeń współczynników regresji. Jest to różnica pomiędzy wartością zmiennej zależnej, a jej skorygowaną wartością przewidywaną.

**Statystyki współczynników.** Zapisuje macierz wariancji-kowariancji estymacji parametrów w zestawie danych albo pliku danych. Dla każdej zmiennej zależnej tworzony jest również wiersz ocen parametrów, wiersz wartości istotności dla statystyk  $t$  odpowiadających ocenom parametrów i wiersz stopni swobody reszt. W modelach o wielu zmiennych tego typu wiersze tworzone są dla każdej zmiennej zależnej. Dane w tej macierzy można użyć w innych procedurach, które odczytują pliki macierzy. 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. .

## Opcje OML powtarzanych pomiarów

W tym oknie dialogowym dostępne są statystyki opcjonalne. Statystyki są obliczane przy użyciu modelu efektów statych.

### Wyświetl

#### Statystyki opisowe

Powoduje wyświetlenie obserwowanych średnich, odchyłeń standardowych i liczebności dla wszystkich zmiennych zależnych we wszystkich komórkach.

#### 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.

#### Obserwowana moc

Powoduje wyświetlenie siły testu po ustaleniu hipotezy alternatywnej na podstawie obserwowanej wartości.

#### Oszacowania parametrów

Powoduje wyświetlenie ocen parametrów, błędów standardowych, testów  $t$ , przedziałów ufności i obserwowanej siły każdego testu.

#### Macierze SSCP

Wyświetla macierze SSCP hipotezy i błędu.

#### Macierz SSCP reszt

Wyświetla macierz SSCP reszt.

#### Macierz transformacji

Wyświetla test sferyczności Bartletta macierzy kowariancji reszt.

#### Testowanie jednorodności

Powoduje przeprowadzenie testu Levene'a na jednorodność wariancji dla każdej zmiennej zależnej, na wszystkich kombinacjach poziomów czynników międzyobiektowych, tylko dla takich czynników. Ponadto testy jednorodności obejmują test  $M$  Boxa jednorodności macierzy kowariancji zmiennych zależnych po wszystkich poziomach kombinacji czynników międzyobiektowych.

#### Wykresy rozrzut-poziom

Opcję tę można wykorzystać do sprawdzenia założeń dotyczących danych. W przypadku braku czynników, opcja ta jest niedostępna.

### Wykres reszt

Powoduje utworzenie wykresu reszt rzeczywistych-szacowanych-standaryzowanych dla każdej zmiennej zależnej. Wykresy te są przydatne podczas sprawdzania założeń równej wariancji. W przypadku braku czynników, opcja ta jest niedostępna.

### Brak dopasowania

Powoduje sprawdzenie, czy model nadaje się do prawidłowego opisu relacji zachodzącej między zmienną zależną a zmiennymi niezależnymi.

### Ogólne funkcje estymowalne

Pozwala na przeprowadzenie testów hipotez użytkownika dotyczących ogólnych funkcji estymowalnych. Wiersze w każdej macierzy współczynników kontrastu są liniowymi kombinacjami ogólnych funkcji estymowalnych.

### 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 GLM

Te funkcje mogą dotyczyć jednej zmiennej, wielu zmiennych lub analizy z powtarzanymi pomiarami. 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*.

## Analiza składowych wariancji

Procedura Składowe wariancji przeznaczona do modeli efektów mieszanych szacuje udział każdego efektu losowego w wariancji zmiennej zależnej. Procedura ta jest szczególnie interesująca przy analizie modeli mieszanych, takich jak wykres podzielony, pomiary powtarzane jednej zmiennej i plany z blokiem mieszanym. Obliczając składowe wariancji, można określić, na których aspektach należy skupić uwagę, aby wariancję zmniejszyć.

Dostępne są cztery różne metody szacowania składowych wariancji: estymator MINQUE (minimum norm quadratic unbiased estimator), analiza wariancji (ANOVA), największa wiarygodność (ML) i warunkowa największa wiarygodność (REML). Dla każdej metody dostępne są różne opcje specyfikacji.



Domyślnie wyniki każdej metody zawierają oszacowania składowych wariancji. Wyniki metody ML i REML zawierają także tabelę asymptotycznej macierzy kowariancji. Do innych dostępnych rodzajów wyników należy tabela ANOVA i oczekiwane średnie kwadraty w przypadku metody ANOVA oraz historia iteracji w przypadku metod ML i REML. Procedura Składowe wariancji jest w pełni kompatybilna z procedurą OML Jednej zmiennej.

W polu WNK Waga można określić zmienną używaną w celu nadania obserwacjom różnych wag podczas analizy metodą ważonych najmniejszych kwadratów, co kompensuje różne poziomy precyzji pomiarów.

**Przykład.** Akademia rolnicza zmierzyła miesięczny przyrost masy ciała świń w sześciu różnych miotach. Zmienna miot jest czynnikiem losowym o sześciu poziomach. (Sześć badanych miotów stanowi losową próbę z dużej populacji miotów świń). Badacz stwierdza, że zmienność przyrostu masy ciała można przypisać w większym stopniu różnicom między miotami, a w mniejszym stopniu różnicom między osobnikami w jednym miocie.

Wymagania dotyczące danych dla składowych wariancji

**Dane.** Zmienna zależna jest ilościowa. Czynniki są typu jakościowego. Mogą mieć wartości liczbowe lub łańcuchowe składające się z nie więcej niż ośmiu bajtów. Co najmniej jeden czynnik musi być losowy. Oznacza to, że poziomy czynniki muszą być losową próbą spośród wszystkich możliwych poziomów. Współzmiennie są zmiennymi ilościowymi powiązаныmi ze zmienną zależną.

**Założenia.** We wszystkich metodach zakłada się, że parametry modelu efektu losowego mają średnie zerowe i skończone stałe wariancje oraz są wzajemnie nieskorelowane. Parametry modelu z różnych efektów losowych także są nieskorelowane.

Składnik reszty również ma średnią zero i skończoną stałą wariancję. Jest nieskorelowany z parametrami modelu jakiegokolwiek efektu losowego. Zakłada się, że składniki reszt z różnych obserwacji są nieskorelowane.

Na podstawie tych założeń obserwacje z tego samego poziomu czynnika losowego są skorelowane. Ten fakt odróżnia model składowych wariancji od uogólnionego modelu liniowego.

W metodach ANOVA i MINQUE założenie o normalności rozkładu nie jest wymagane. Obie te metody są odporne na umiarkowane odstępstwa od założenia o normalności rozkładu.

W metodach ML i REML parametr modelu i składnik reszt muszą mieć rozkłady normalne.

**Procedury pokrewne.** Przed rozpoczęciem analizy składowych wariancji należy użyć procedury eksploracji w celu zbadania danych. Do testowania hipotez można używać procedur OML Jednej zmiennej, OML Wielu zmiennych i OML Powtarzanych pomiarów.

Uzyskiwanie tabel składowych wariancji

1. Z menu wybierz:

**Analiza > Ogólny model liniowy > Składowe wariancji...**

2. Wybierz zmienną zależną.

3. Odpowiednio do swoich danych wybierz zmienne, które mają być Czynnikiami stałymi, Czynnikiami losowymi oraz Współzmiennymi. Do określania zmiennej ważącej należy użyć pola WNK Waga.

## Składowe wariancji: Model

**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. Model musi zawierać czynnik losowy.

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.

**Uwzględnij wyraz wolny w modelu.** Zwykle w modelu uwzględniony jest wyraz wolny. 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:

## Składowe wariacji: Opcje

**Metoda.** Można wybrać jedną z czterech metod estymacji składowych wariacji.

- Metoda **MINQUE** (minimum norm quadratic unbiased estimator) generuje estymacje niezmiennie względem efektów stałych. Jeśli dane mają rozkład normalny, a oszacowania są prawidłowe, ta metoda generuje najmniejszą wariację spośród wszystkich estymatorów nieobciążonych. Można wybrać metodę ważenia efektów losowych a priori.
- Metoda analizy wariacji **ANOVA** (analysis of variance) oblicza nieobciążone oszacowania przy użyciu sum kwadratów Typu I lub Typu III dla każdego efektu. Metoda ANOVA niekiedy generuje ujemne oszacowania wariacji, które mogą wskazywać na nieprawidłowość modelu, użycie nieodpowiedniej metody estymacji lub niewystarczającą ilość danych.
- Metoda **największej wiarygodności ML** (maximum likelihood) przy użyciu iteracji generuje oszacowania potencjalnie najbardziej spójne z faktycznie zaobserwowanymi danymi. Te oszacowania mogą być obciążone. Ta metoda asymptotycznie zbliża się do rozkładu normalnego. Translacja nie wpływa na oszacowania ML i REML. W tej metodzie nie jest brana pod uwagę liczba stopni swobody, przy której szacowano efekty stałe.
- Metoda **warunkowej największej wiarygodności REML** (Restricted maximum likelihood) redukuje oszacowania ANOVA dla wielu (a nawet wszystkich) obserwacji danych zrównoważonych. Ponieważ metoda ta jest skorygowana dla efektów stałych, powinna charakteryzować się mniejszymi błędami standardowymi niż metoda ML. W metodzie tej jest brana pod uwagę liczba stopni swobody, przy której szacowano efekty stałe.

**Effekt losowy a priori.** Opcja **Jednostajny** oznacza założenie, że wszystkie efekty losowe i składnik reszty mają taki sam wpływ na obserwacje. Schemat **Zero** jest równoważny założeniu zerowych wariacji efektów losowych. Opcja dostępna tylko w przypadku metody MINQUE.

**Suma kwadratów.** Sumy kwadratów **Typu I** używane są w przypadku modelu hierarchicznego, o którym często mowa jest w piśmiennictwie dotyczącym składowych wariacji. W przypadku wybrania **Typu III**, domyślnego w analizie OML, oszacowania wariacji można wykorzystać w procedurze OML Jednej

zmiennej do testowania hipotez z sumami kwadratów Typu III. Opcja dostępna tylko w przypadku metody ANOVA.

**Kryteria.** Można określić kryterium zbieżności i maksymalną liczbę iteracji. Opcja dostępna tylko w przypadku metody ML lub REML.

**Pokaż.** W metodzie ANOVA można nakazać przedstawienie sum kwadratów i oczekiwanych średnich kwadratów. Jeśli wybrano opcję **Największa wiarygodność** lub **Warunkowa największa wiarygodność**, można przedstawić historię iteracji.

## Suma kwadratów (składowe wariancji)

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. Metoda bazująca na sumie 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 III.** Wartość domyślna: W tej metodzie suma kwadratów efektu jest obliczana jako suma kwadratów uwzględniająca wszelkie inne efekty nie zawierające tego efektu i ortogonalne względem wszelkich efektów, które go zawierają. Sumy kwadratów typu III mają tę zaletę, że są niezmiennicze ze względu na częstości w komórkach, jeśli ogólna forma szacowania jest stała. Dlatego ten typ uważany jest często za przydatny 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.
- Dowolne, zrównoważone lub niezrównoważone, modele niezawierające pustych komórek.

## Składowe wariancji: Zapisz w nowym pliku

Niektóre wyniki tej procedury można zapisać w nowym pliku danych IBM SPSS Statistics.

**Oceny składowych wariancji.** Zapisuje oszacowania składowych wariancji i etykiety oszacowań w pliku danych lub zbiorze danych. Można je następnie wykorzystać do obliczania kolejnych statystyk i prowadzenia dalszych analiz w procedurach OML. Na przykład można użyć ich do obliczenia przedziałów ufności lub testowania hipotez.

**Kowariancja składowych.** Zapisuje macierz wariancji-kowariancji lub macierz korelacji w pliku danych lub zbiorze danych. Opcja dostępna tylko wtedy, gdy wybrano metodę **Największa wiarygodność** lub **Warunkowa największa wiarygodność**.

**Lokalizacja dla tworzonych zmiennych.** Umożliwia określenie nazwy zbioru danych lub nazwy pliku zewnętrznego zawierającego oszacowania składowych wariancji i/lub macierz. 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. .

Za pomocą komendy MATRIX można wyodrębnić potrzebne dane z pliku danych, a następnie obliczyć przedziały ufności lub wykonać testy.

## Dodatkowe właściwości komendy VARCAMP

Język składni komend umożliwia również:

- Określanie zagnieżdżonych efektów w planie (za pomocą opcji DESIGN).
- Uwzględnianie braków danych zdefiniowanych przez użytkownika (za pomocą opcji komendy MISSING).
- Określanie kryteriów EPS (za pomocą opcji komendy CRITERIA).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Liniove modele mieszane

---

Procedura Liniove modele mieszane rozszerza uogólniony model liniowy w taki sposób, że dane mogą wykazywać skorelowaną i niestałą zmienność. Liniowy model mieszany zapewnia zatem elastyczność modelowania nie tylko średnich danych, lecz także ich wariancji i kowariancji.

Procedura Liniove modele mieszane jest także elastycznym narzędziem do dopasowywania innych modeli, które można sformułować jako mieszane modele liniowe. Do modeli takich należą modele wielopoziomowe, hierarchiczne modele liniowe i modele współczynników losowych.

### Przykład

Sieć sklepów spożywczych jest zainteresowana wpływem różnych kuponów promocyjnych na wydatki klientów. Wybiera losową próbę spośród swoich stałych klientów i obserwuje wydatki każdego klienta z tej próby przez 10 tygodni. Co tydzień do klientów wysyłany jest pocztą inny kupon promocyjny. Liniowe modele mieszane są następnie używane do oszacowania wpływu różnych kuponów na wydatki, przy jednoczesnej korekcie korelacji ze względu na powtarzanie obserwacji każdego z obiektów w okresie 10 dni.

### Metody

Oszacowanie największej wiarygodności i warunkowej największej wiarygodności.

### Statystyki

Statystyki opisowe: rozmiary próby, średnie i odchylenia standardowe zmiennych zależnych i współzmiennych dla każdej odrębnej kombinacji poziomów czynników. Informacje na poziomie czynnika: posortowane wartości poziomów każdego z czynników oraz ich liczebności (częstości). Ponadto oszacowania parametrów i przedziały ufności dla efektów stałych i testów Walda oraz przedziały ufności dla parametrów macierzy kowariancji. Do oceny różnych hipotez można wykorzystać sumy kwadratów Typu I i Typu III. Typem domyślnym jest typ III.

## Wymagania dotyczące danych dla liniowych modeli mieszanych

### Dane

Zmienna zależna powinna być ilościowa. Czynniki powinny być jakościowe i mogą mieć wartości liczbowe lub łańcuchowe. Współzmiennie i zmienne ważące powinny być ilościowe. Obiekty i zmienne powtórzonych pomiarów mogą być dowolnego typu.

### Założenia

Zakłada się, że zmienna zależna jest liniowo powiązana z czynnikami stałymi, czynnikami losowymi i współzmiennymi. Efekty stałe modelują średnią zmiennej zależnej. Efekty losowe modelują strukturę kowariancji zmiennej zależnej. Efekty losowe uznaje się za niezależne od siebie nawzajem i dla każdego z nich obliczona zostanie odrębna macierz kowariancji; jednak składniki modelu określone dla tego samego efektu losowego mogą być skorelowane. Powtarzane pomiary modelują strukturę kowariancji reszt. Przyjmuje się także, że zmienna zależna pochodzi z rozkładu normalnego.

### Procedury pokrewne

Przed rozpoczęciem analizy należy użyć procedury eksploracji w celu zbadania danych. Jeśli nie podejrzewa się istnienia zmienności skorelowanej lub niestałej, można użyć procedury OML Jednej zmiennej lub OML Powtarzane pomiary. Ewentualnie można użyć procedury Analiza składowych wariancji, jeśli efekty losowe mają strukturę kowariancji składników wariancji i nie ma pomiarów powtórzonych.

## Uzyskiwanie analizy liniowych modeli mieszanych

1. Wybierz z menu następującą opcję:

**Analiza > Modele mieszane > Liniowy...**

2. Opcjonalnie wybierz jedną lub więcej zmiennych obiektowych.
3. Opcjonalnie wybierz jedną lub więcej zmiennych pomiarów powtórzonych. Jeśli zostały zdefiniowane zmienne powtórzone, wybierz z listy rozwijanej opcję **Typ kowariancji powtórzonej**.
4. Opcjonalnie wybierz jedną lub więcej zmiennych miar Kroneckera.
5. Opcjonalnie wybierz strukturę kowariancji reszt.
6. Kliknij przycisk **Kontynuuj**.
7. Wybierz zmienną zależną.
8. Wybierz co najmniej jeden czynnik lub współzmienną.
9. Kliknij opcję **Ustalona** lub **Losowe** i określ co najmniej model efektów stałych lub model efektów losowych.

Opcjonalnie wybierz zmienną ważącą.

## Liniowe modele mieszane: Zmienne obiektowe i powtórzone

To okno dialogowe umożliwia wybranie zmiennych, które definiują obiekty i powtórzone obserwacje, miar Kroneckera oraz struktury kowariancji dla reszt.

### Obiekty

Obiekt jest jednostką obserwacyjną, która może być rozważana niezależnie od innych obiektów. Na przykład odczyty ciśnienia krwi pacjenta w badaniu medycznym mogą być traktowane jako niezależne od odczytów innych pacjentów. Definiowanie obiektów staje się szczególnie istotne, gdy istnieją wielokrotne pomiary na każdy obiekt i gdy wymagane jest modelowanie korelacji między tymi obserwacjami. Na przykład można oczekiwać, że pomiary ciśnienia krwi u jednego pacjenta podczas kolejnych wizyt u lekarza będą skorelowane.

Obiekty mogą być również definiowane przez kombinację wielu zmiennych na poziomie czynnika; na przykład można określić *Płeć* i *Kategorię wiekową* jako zmienne obiektowe, aby zamodelować założenie, że *mężczyźni w wieku powyżej 65 lat* są podobni do siebie nawzajem, ale niezależni od *mężczyzn w wieku poniżej 65 lat* i *kobiet*.

Wszystkie zmienne określone na liście **Obiekty** są używane do zdefiniowania obiektów dla struktury kowariancji reszt. Można wykorzystać niektóre lub wszystkie zmienne do zdefiniowania obiektów dla struktury kowariancji obiektów losowych.

### Powtórzone

Zmienne określone na tej liście służą do identyfikowania obserwacji powtórzonych. Na przykład pojedyncza zmienna *Tydzień* może identyfikować 10 tygodni obserwacji w badaniu medycznym, a w celu identyfikacji obserwacji codziennych w ciągu roku można używać razem zmiennych *Miesiąc* i *Dzień*.

### Typ powtórzonej kowariancji

Określa strukturę kowariancji dla reszt. Dostępne są następujące struktury:

- Przed-zależność: Pierwszego rzędu
- AR(1)
- Iloczyn prosty AR1 (UN\_AR1)
- Iloczyn prosty nieustrukturyzowany (UN\_UN)
- Iloczyn prosty - symetria złożona (UN\_CS)
- AR(1): Heterogeniczna
- ARMA(1,1)
- Symetria złożona

- Symetria złożona: Metryka korelacji
- Symetria złożona: Heterogeniczna
- Przekątna
- Analityczna czynnika: Pierwszego rzędu
- Analityczna czynnika: Pierwszego rzędu, Heterogeniczna
- Huynh-Feldt
- Skalowana tożsamość
- Toeplitz
- Toeplitz: Heterogeniczna
- Nieustrukturalizowana
- Nieustrukturalizowana: Metryka korelacji
- Przestrzenna: potęgowa
- Przestrzenna: wykładnicza
- Przestrzenna: Gaussa
- Przestrzenna: liniowa
- Przestrzenna: liniowo-logarytmiczna
- Przestrzenna: sferyczna

### **Miary Kroneckera**

Wybierz zmienne określające strukturę obiektów dla pomiarów kowariancji Kroneckera i określ, jak skorelowane są błędy pomiaru. Pole jest dostępne tylko wtedy, gdy jako **Typ kowariancji powtórzonej** wybrany jest jeden z następujących typów:

- Iloczyn prosty AR1 (UN\_AR1)
- Iloczyn prosty nieustrukturyzowany (UN\_UN)
- Iloczyn prosty - symetria złożona (UN\_CS)

### **Współrzędne kowariancji przestrzennej**

Zmienne na tej liście określają współrzędne powtórzonych obserwacji, gdy dla typu struktury kowariancji wybrano jeden z typów kowariancji przestrzennej.

Więcej informacji można znaleźć w temacie [“Struktury kowariancji”](#) na stronie 103.

## **Liniowe modele mieszane: Efekty stałe**

**Efekty stałe.** Nie istnieje model domyślny, zatem należy jawnie określić efekty stałe. Zamiast tego można utworzyć składniki zagnieżdżone lub niezagnieżdżone.

**Uwzględnij wyraz wolny.** Wyraz wolny jest zwykle uwzględniany w modelu. Jeśli można założyć, że dane przechodzą przez początek układu współrzędnych, wyraz wolny można wyłączyć z modelu.

**Suma kwadratów.** Metoda obliczania sum kwadratów. W przypadku modeli bez brakujących komórek często używa się metody Typu III.

## **Tworzenie składników niezagnieżdżonych**

Dla wybranych czynników i współzmiennych:

**Czynnikowy.** Tworzy wszystkie możliwe interakcje i efekty główne wybranych zmiennych. Jest to wartość domyślna.

**Interakcja.** Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu.

**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.

## Tworzenie składników zagnieżdżonych

W tej procedurze można zbudować składniki zagnieżdżone dla modelu. Składniki zagnieżdżone są przydatne do modelowania efektu czynników lub współzmiennych, których wartości nie wchodzi w interakcje z poziomami innego czynnika. Na miarę ich sklepów spożywczych może analizować zwyczaj zakupowe swoich klientów w kilku sklepach. Ponieważ każdy klient bywa regularnie tylko w jednym z tych sklepów, efekt *Klient* jest **zagnieżdżony w** efekcie *Lokalizacja sklepu*.

Ponadto można uwzględnić efekty interakcji lub dodać wiele poziomów zagnieżdżania do zagnieżdżonego składnika.

**Ograniczenia.** W odniesieniu do składników zagnieżdżonych obowiązują następujące ograniczenia:

- Wszystkie czynniki w interakcji muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A^*A$ .
- Wszystkie czynniki w efekcie zagnieżdżonym muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A(A)$ .
- Efekt nie może być zagnieżdżony w obrębie współzmiennnej. A zatem, jeśli  $A$  jest czynnikiem, a  $X$  jest współzmienną, to określenie  $A(X)$  jest niedozwolone.

## 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 zagnieżdżany jest w drugim, ten z kolei w efekcie określonym jako trzeci itd. (ten sposób zagnieżdżania można określić jedynie za pomocą składni).

**Typ III.** Wartość domyślna: W tej metodzie suma kwadratów efektu jest obliczana jako suma kwadratów uwzględniająca wszelkie inne efekty nie zawierające tego efektu i ortogonalne względem wszelkich efektów, które go 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.
- Dowolne, zrównoważone lub niezrównoważone, modele niezawierające pustych komórek.

## Linijowe modele mieszane: Efekty losowe

**Typ struktury kowariancji.** Umożliwia określenie struktury kowariancji dla modelu efektów losowych. Dla każdego efektu losowego szacowana jest odrębna macierz kowariancji. Dostępne są następujące struktury:

- Przed-zależność: Pierwszego rzędu
- AR(1)
- AR(1): Heterogeniczna
- ARMA(1,1)
- Symetria złożona
- Symetria złożona: Metryka korelacji
- Symetria złożona: Heterogeniczna
- Przekątna
- Analityczna czynnika: Pierwszego rzędu
- Analityczna czynnika: Pierwszego rzędu, Heterogeniczna
- Huynh-Feldt
- Skalowana tożsamość
- Toeplitz
- Toeplitz: Heterogeniczna
- Bez struktury
- Nieustrukturalizowana: Metryka korelacji
- Składowe wariancji

Więcej informacji można znaleźć w temacie “Struktury kowariancji” na stronie 103.

**Efekty losowe.** Nie istnieje model domyślny, zatem należy jawnie określić efekty losowe. Zamiast tego można utworzyć składniki zagnieżdżone lub niezagnieżdżone. Można także nakazać uwzględnienie składnika stałej w modelu efektów losowych.

Można określić więcej niż jeden model efektów losowych. Po utworzeniu pierwszego modelu należy kliknąć przycisk **Dalej**, aby utworzyć następny. Przycisk **Poprzedni** umożliwia cofanie się do istniejących modeli. Zakłada się, że każdy model efektów losowych jest niezależny od każdego innego modelu efektów losowych, tj. dla każdego modelu obliczona zostanie odrębna macierz kowariancji. Składniki określone w tym samym modelu efektów losowych mogą być skorelowane.

**Grupowanie obiektów.** Na liście wymienione są zmienne wybrane jako zmienne obiektowe w oknie dialogowym wyboru obiektów/zmiennych powtarzanych pomiarów. Wybierz niektóre lub wszystkie wymienione zmienne, aby zdefiniować obiekty dla modelu efektów losowych.

**Wyświetl predykcje parametrów dla tego zbioru efektów losowych.** Pozwala wyświetlać szacowania parametrów według estymatora efektów losowych.

## Oszacowanie liniowych modeli mieszanych

### Metoda

Wybierz oszacowanie największej wiarygodności lub warunkowej największej wiarygodności.

### Stopnie swobody

Umożliwia określenie liczby stopni swobody dla wszystkich testów.

### Metoda reszt

Metoda reszt umożliwia określenie stałej liczby stopni swobody dla wszystkich testów. Jest przydatna, jeśli próba jest wystarczająco duża lub dane są zrównoważone, albo model używa kowariancji prostszego typu – na przykład przekątna lub tożsamość skalowana.

### Przybliżenie Satterthwaite'a

W metodzie Satterthwaite'a stosuje się liczbę stopni swobody zmiennej we wszystkich testach. Jest przydatna, jeśli próba jest niewielka, dane są niezrównoważone lub w modelu używany jest skomplikowany typ kowariancji – na przykład nieustrukturyzowana.



### **Przybliżenie Kenwarda-Rogera**

Metoda Kenwarda-Rogera udostępnia bardziej precyzyjny estymator małych prób dla wariacji-kowariancji dala parametrów efektów stałych i przybliżoną liczbę stopni swobody mianownika w testach  $t$  i  $F$ . Metoda wprowadza współczynnik skalowania dla statystyki  $F$  i estymuje ją oraz liczbę stopni swobody mianownika przy użyciu rozwinięcia szeregu Taylora dla estymowanej struktury losowej w danych.

**Uwaga:** Metoda Kenwarda-Rogera stosowana jest w przypadku kowariancji opartej na modelu (a nie kowariancji odpornej). W przypadku wybrania zarówno metody Kenwarda-Rogera, jak i kowariancji odpornej, metoda Kenwarda-Rogera stosowana jest względem kowariancji opartej na modelu i wyświetlane jest następujące ostrzeżenie: „Ponieważ wybrano metodę Kenwarda-Rogera, metoda kowariancji odpornej została zamieniona na metodę kowariancji opartej na modelu”.

### **Iteracje**

Dostępne są następujące opcje:

#### **Maksymalna liczba iteracji**

Podaj nieujemną liczbę całkowitą.

#### **Maksimum kroków połowienia**

Przy każdej iteracji rozmiar kroku jest zmniejszany o 0,5, aż do momentu wzrostu logarytmu wiarygodności lub osiągnięcia maksymalnej wartości połowienia kroków. Podaj dodatnią liczbę całkowitą.

#### **Pokaż przebieg iteracji co każdy $n$ kroków**

Ta opcja powoduje wyświetlenie tabeli zawierającej oszacowania wartości i parametru funkcji logarytmu wiarygodności co każde  $n$  iteracji, począwszy od iteracji 0 (wstępne estymacje). Jeśli zostanie wybrana opcja pokazywania przebiegu iteracji, wówczas zawsze przedstawiana będzie ostatnia iteracja bez względu na wartość  $n$ .

### **Zbieżność logarytmu wiarygodności**

Zbieżność jest zakładana, jeśli zmiana bezwzględna lub względna w funkcji logarytmu wiarygodności jest mniejsza niż podana wartość, która musi być nieujemna. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### **Zbieżność parametru:**

Zbieżność jest zakładana, jeśli maksymalna zmiana bezwzględna lub względna w oszacowaniach parametru jest mniejsza niż podana wartość, która musi być nieujemna. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### **Zbieżność Hessego**

W przypadku specyfikacji **Wartości bezwzględne** zakładana jest zbieżność, jeśli statystyka oparta na macierzy Hessego jest mniejsza niż określona wartość. W przypadku specyfikacji **Względne** zbieżność jest zakładana, jeśli statystyka jest mniejsza niż iloczyn wartości określonej i wartości bezwzględnej logarytmu wiarygodności. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### **Maksymalna liczba kroków oceny**

Ta opcja żąda użycia algorytmu oceny Fishera do iteracji numer  $n$ . Podaj nieujemną liczbę całkowitą.

### **Tolerancja osobliwości**

Ta wartość jest stosowana jako tolerancja podczas kontroli osobliwości. Podaj wartość dodatnią.

## **Linijowe modele mieszane: Statystyki**

**Statystyki podsumowujące.** Generuje tabele do:

- **Statystyki opisowe.** Przedstawia rozmiary prób, średnie i odchylenia standardowe zmiennej zależnej oraz współzmiennych (jeśli są określone). Statystyki te są prezentowane dla każdej odrębnej kombinacji poziomów czynników.
- **Podsumowanie przetwarzania obserwacji.** Przedstawia posortowane wartości czynników, zmienne pomiarów powtórzonych, obiekty pomiarów powtórzonych i obiekty efektów losowych oraz ich częstości.

**Statystyki modelu.** Generuje tabele do:

- **Oszacowania parametrów dla efektów stałych.** Przedstawia oszacowania parametrów efektów stałych i ich przybliżone błędy standardowe.
- **Testy dla parametrów kowariancji.** Przedstawia asymptotyczne błędy standardowe i testy Walda parametrów kowariancji.
- **Korelacje oszacowań parametrów.** Przedstawia asymptotyczną macierz korelacji oszacowań parametrów efektów stałych.
- **Kowariancje oszacowań parametrów.** Przedstawia asymptotyczną macierz kowariancji oszacowań parametrów efektów stałych.
- **Kowariancje efektów losowych.** Przedstawia szacowaną macierz kowariancji efektów losowych. Ta opcja jest dostępna tylko wtedy, gdy określono co najmniej jeden efekt losowy. Jeśli dla efektu losowego określona jest zmienna obiektowa, to wyświetlany jest wspólny blok.
- **Kowariancje reszt.** Przedstawia szacowaną macierz kowariancji reszt. Ta opcja jest dostępna tylko wtedy, gdy określono co najmniej jedną zmienną pomiarów powtarzanych. Jeśli określona jest zmienna obiektowa, to wyświetlany jest wspólny blok.
- **Macierz współczynników kontrastów.** Ta opcja powoduje wyświetlenie funkcji estymowalnych używanych do testowania efektów stałych i hipotez użytkownika.

**Przedział ufności.** Ta wartość jest używana każdorazowo przy wyznaczaniu przedziału ufności. Podaj wartość większą lub równą 0 i mniejszą od 100. Wartość domyślna wynosi 95.

## Liniowe modele mieszane: Szacowane średnie brzegowe

**Szacowane średnie brzegowe dopasowanych modeli.** W tej grupie można zażądać wyznaczenia średnich brzegowych zmiennych zależnych w komórkach szacowanych przez model oraz ich błędów standardowych dla określonych czynników. Ponadto można nakazać porównanie poziomów czynników efektów głównych.

- **Czynniki i interakcje czynników.** Ta lista zawiera czynniki i interakcje czynników określone w oknie dialogowym Ustalone oraz składnik ŚREDNIA OGÓLNA. Na liście nie są ujęte składniki modelu utworzone na podstawie współzmiennych.
- **Pokaż średnie dla.** Procedura obliczy szacowane średnie brzegowe dla czynników i interakcji czynników uwzględnionych na tej liście. Jeśli wybrany jest składnik ŚREDNIA OGÓLNA, to przedstawione zostaną szacowane średnie brzegowe zmiennej zależnej zwinięte dla wszystkich czynników. Należy zauważyć, że wszelkie wybrane czynniki lub interakcje czynników pozostają wybrane, chyba że odpowiednia zmienna zostanie usunięta z listy Czynniki w głównym oknie dialogowym.
- **Porównaj efekty główne.** Ta opcja nakazuje porównanie parami poziomów wybranych efektów głównych. Opcja Korekta przedziału ufności umożliwia skorygowanie przedziałów ufności i wartości istotności z uwzględnieniem wielokrotnych porównań. Dostępne są metody LSD (brak), Bonferroni i Sidak. Wreszcie, dla każdego czynnika można wybrać kategorię odniesienia, z której będą dokonywane porównania. Jeśli nie zostanie wybrana kategoria odniesienia, to utworzone zostaną wszystkie porównania parami. Dostępne opcje kategorii odniesienia to: pierwsza, ostatnia albo użytkownika (w tym przypadku należy wprowadzić wartość określającą kategorię odniesienia).

## Liniowe modele mieszane: Zapisz

To okno dialogowe umożliwia zapisanie różnych wyników modelu w pliku roboczym.

**Przewidywane wartości stałe.** Zapisuje zmienne związane ze średnimi regresji bez efektów.

- **Wartości przewidywane.** Średnie regresji bez efektów losowych.
- **Błędy standardowe.** Błędy standardowe oszacowań.
- **Stopnie swobody.** Stopnie swobody powiązane z oszacowaniami.

**Wartości przewidywane i reszty.** Zapisuje zmienne związane z wartością dopasowaną przez model.

- **Wartości przewidywane.** Wartość dopasowana przez model.

- **Błędy standardowe.** Błędy standardowe oszacowań.
- **Stopnie swobody.** Stopnie swobody powiązane z oszacowaniami.
- **Reszty.** Wartość danych pomniejszona o wartość przewidywaną.

## Liniowe modele mieszane-Eksport

To okno dialogowe umożliwia eksportowanie zawartości tabel wynikowych EBLUP (empirycznych najlepszych liniowych predykcji nieobciążonych) do zbiorów danych lub plików .sav . Przycisk Eksportuj w oknie dialogowym Liniowe modele mieszane jest aktywny, jeśli w oknie dialogowym Losowe określono co najmniej jeden efekt losowy i zaznaczono pole wyboru "Wyświetl predykcje parametrów dla tego bloku".

### Eksportuj EBLUPS

Wybierz miejsce docelowe: **Zestaw danych** lub **Plik danych**

Opcjonalnie podaj nazwę.

Jeśli określono wiele efektów losowych z wygenerowanymi EBLUP, usuń zaznaczenie pola wyboru, aby każda tabela wyników była wyświetlana w osobnym zbiorze danych lub pliku.

## Dodatkowe właściwości komendy MIXED

Język składni komend umożliwia również:

- Określanie testów efektów w odniesieniu do liniowej kombinacji efektów lub wartości (za pomocą opcji komendy TEST).
- Uwzględnianie braków danych zdefiniowanych przez użytkownika (za pomocą opcji komendy MISSING).
- Oblicz szacowane średnie brzegowe dla określonych wartości współzmiennych (przy użyciu słowa kluczowego WITH opcji komendy EMMEANS).
- Porównaj proste efekty główne interakcji (przy użyciu opcji komendy EMMEANS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Uogólnione modele liniowe

Uogólniony model liniowy rozszerza ogólny model liniowy w taki sposób, że zmienna zależna jest liniowo powiązana z czynnikami i współzmiennymi za pośrednictwem określonej funkcji łączenia. Model pozwala ponadto, aby zmienna zależna nie miała rozkładu normalnego. Dzięki bardzo ogólnej postaci wzoru modelu obejmuje on wiele modeli statystycznych, takich jak regresja liniowa dla odpowiedzi o rozkładzie normalnym, modele logistyczne dla danych binarnych, modele logarytmiczno-liniowe dla danych o liczebności i wiele innych modeli statystycznych.

**Przykłady.** Uogólnionych modeli liniowych może używać firma transportowa, aby dopasować regresję Poissona w celu obliczenia szkód poniesionych na skutek eksploatacji kilku typów statków wyprodukowanych w różnych okresach, a model wynikowy może ułatwić ustalenie, które typy statków są najbardziej podatne na uszkodzenia.

Uogólnionych modeli liniowych może używać firma sprzedająca ubezpieczenia komunikacyjne, aby dopasować regresję gamma do żądań odszkodowań za samochody, a model wynikowy może ułatwić określenie czynników, które przyczyniają się najbardziej do wartości roszczenia.

Uogólnionych modeli liniowych mogą używać badacze medyczni, aby dopasować komplementarną regresję log-log do przedziałowo cenzurowanych danych dotyczących przeżycia, aby oszacować czas do nawrotu choroby.

Wymagania dotyczące danych dla uogólnionych modeli liniowych

**Dane.** Odpowiedź może być zmienną ilościową, liczebnościową, binarną lub określającą zdarzenia w próbach. Zakłada się, że czynniki są jakościowe. Przyjmuje się, że współzmiennie, waga parametru skali i przesunięcie są ilościowe.

**Założenia.** Zakłada się, że wszystkie obserwacje są niezależne.

Otrzymywanie uogólnionego modelu liniowego

Wybierz z menu następującą opcję:

**Analiza > Uogólnione modele liniowe > Uogólnione modele liniowe...**

1. Określ rozkład i funkcję łączenia (szczegółowe informacje na temat różnych opcji są dostępne poniżej).
2. Na karcie Odpowiedź wybierz zmienną zależną.
3. Na karcie Predyktory wybierz czynniki i współzmiennie, które mają być używane do predykcji zmiennej zależnej.
4. Na karcie Model określ efekty modelu przy użyciu wybranych czynników i współzmiennych.

Na karcie Typ modelu można określić rozkład i funkcję łączenia dla modelu, przy wykorzystaniu skrótów do kilku typowych modeli, które są sklasyfikowane według typu odpowiedzi.

Typy modeli

**Ilościowa zależna.** Dostępne są następujące opcje:

- **Liniowy.** Określa Normalny jako rozkład i Tożsamość jako funkcję łączenia.
- **Gamma wiązany logarytmicznie.** Określa Gamma jako rozkład i Log jako funkcję łączenia.

**Porządkowa zależna.** Dostępne są następujące opcje:

- **Porządkowy logistyczny.** Określa Wielomianowy (porządkowy) jako rozkład i Logit Skumulowane jako funkcję łączenia.
- **Porządkowy probit.** Określa Wielomianowy (porządkowy) jako rozkład i Probit Skumulowane jako funkcję łączenia.

**Liczebności.** Dostępne są następujące opcje:

- **Poissona logliniowy.** Określa Poissona jako rozkład i Log jako funkcję łączenia.
- **Ujemny dwumianowy wiązany logarytmicznie.** Określa Ujemny dwumianowy (z wartością 1 dla parametru dodatkowego) jako rozkład i Log jako funkcję łączenia. Aby procedura oszacowała wartość parametru dodatkowego, należy podać model niestandardowy z rozkładem ujemnym dwumianowym, a następnie wybrać opcję **Szacuj wartość** w grupie Parametr.

**Binarna zależna lub dane ze zdarzeń/prób.** Dostępne są następujące opcje:

- **Binarny logistyczny.** Określa Dwumianowy jako rozkład i Logit jako funkcję łączenia.
- **Binarny probit.** Określa Dwumianowy jako rozkład i Probit jako funkcję łączenia.
- **Przeżycia obcięte przedziałowe.** Określa Dwumianowy jako rozkład i Komplementarny Log-log jako funkcję łączenia.

**Mieszany.** Dostępne są następujące opcje:

- **Tweediego wiązany logarytmicznie.** Określa Tweediego jako rozkład i Log jako funkcję łączenia.
- **Tweediego wiązany tożsamościowo.** Określa Tweediego jako rozkład i Tożsamość jako funkcję łączenia.

**Użytkownika.** Umożliwia określenie własnej kombinacji funkcji rozkładu i łączenia.

Dystrybucja

Ten wybór określa rozkład zmiennej zależnej. Możliwość określenia rozkładu innego niż normalny i nietożsamościowej funkcji łączenia jest istotnym ulepszeniem uogólnionego modelu liniowego w porównaniu do ogólnego modelu liniowego. Istnieje wiele możliwych kombinacji rozkład-funkcja łączenia, a kilka z nich może być odpowiednich dla dowolnego zbioru danych, dlatego wybór może być zależny od rozważań teoretycznych a priori lub tego, która kombinacja wydaje się zapewniać najlepsze dopasowanie.

- **Dwumianowy.** Ten rozkład jest odpowiedni tylko dla zmiennych, które reprezentują zmienne dychotomiczne lub liczbę zdarzeń.

- **Gamma.** Ten rozkład jest odpowiedni dla zmiennych z dodatnimi wartościami skali, które są skośne w kierunku większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Odwrócony Gaussa.** Ten rozkład jest odpowiedni dla zmiennych z dodatnimi wartościami skali, które są skośne w kierunku większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Ujemny dwumianowy.** Ten rozkład może być traktowany jako seria prób wymaganych do zaobserwowania  $k$  sukcesów i jest odpowiedni dla zmiennych z nieujemnymi liczbami całkowitymi. Jeśli wartość danych nie jest liczbą całkowitą, jest mniejsza od 0 lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie. Wartość parametru dodatkowego ujemnego rozkładu dwumianowego może być dowolną liczbą większą niż lub równą 0; można ustawić ją jako wartość stałą lub zezwolić na jej oszacowanie w czasie trwania procedury. Jeśli parametr dodatkowy jest ustawiony na 0, użycie tego rozkładu da takie same efekty, jak użycie rozkładu Poissona.
- **Normalny.** Ten rozkład jest odpowiedni dla zmiennych ilościowych, których wartości rozkładają się symetrycznie, w kształcie dzwonu, wokół wartości centralnej (średniej). Zmienna zależna musi być typu liczbowego.
- **Poissona.** Ten rozkład można traktować jako liczbę wystąpień zdarzenia badanego w ustalonym okresie i jest odpowiedni dla zmiennych o nieujemnych wartościach całkowitych. Jeśli wartość danych nie jest liczbą całkowitą, jest mniejsza od 0 lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Tweedie'go.** Ten rozkład jest odpowiedni dla zmiennych, które mogą być reprezentowane przez poissonowsko mieszane rozkłady gamma; rozkład ten jest mieszany, to znaczy że łączy właściwości rozkładu ciągłego (nieujemne wartości rzeczywiste) i dyskretnego (prawdopodobieństwo dodatnie dla pojedynczej wartości, 0). Zmienna zależna musi być liczbowa, z wartościami danych większymi niż lub równymi zero. Jeśli wartość danych jest mniejsza niż zero lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie. Wartość stała parametru rozkładu Tweedie'go może być dowolną liczbą większą niż jeden i mniejszą niż dwa.
- **Wielomianowy.** Ten rozkład jest odpowiedni dla zmiennych, które reprezentują odpowiedzi porządkowe. Zmienna zależna może być liniowa lub łańcuchowa i musi zawierać co najmniej dwie różniące się poprawne wartości danych.

#### Funkcje łączenia

Funkcja łączenia to przekształcenie zmiennej zależnej, które umożliwi estymację modelu. Dostępne są następujące funkcje:

- **Tożsamość.**  $f(x)=x$ . Zmienna zależna nie jest przekształcana. To połączenie może być używane dla dowolnego rozkładu.
- **Komplementarny logarytmiczno-logarytmiczny.**  $f(x)=\log(-\log(1-x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Skumulowany Cauchit.**  $f(x) = \tan(\pi(x-0.5))$ , stosowany do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany logarytmiczno-logarytmiczny dopełnienia.**  $f(x)=\ln(-\ln(1-x))$  ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany logit.**  $f(x)=\ln(x/(1-x))$ ; ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany ujemny logarytmiczno-logarytmiczny.**  $f(x) = -\ln(-\ln(x))$ ; ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany probit.**  $f(x) = \Phi^{-1}(x)$  ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi, gdzie  $\Phi^{-1}$  jest odwrotnością funkcji skumulowanego rozkładu standardowego normalnego. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Logarytm.**  $f(x)=\log(x)$ . To połączenie może być używane dla dowolnego rozkładu.

- **Logarytmiczny dopełnienia.**  $f(x)=\log(1-x)$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Ujemny dwumianowy.**  $f(x)=\log(x / (x+k^{-1}))$ , gdzie  $k$  to parametr dodatkowy ujemnego rozkładu dwumianowego. Ma zastosowanie tylko w przypadku ujemnego rozkładu dwumianowego.
- **Ujemny logarytmiczno-logarytmiczny.**  $f(x)=-\log(-\log(x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Nieparzysty potęgowy.**  $f(x)=[(x/(1-x))^\alpha-1]/\alpha$ , jeśli  $\alpha \neq 0$ .  $f(x)=\log(x)$ , jeśli  $\alpha=0$ .  $\alpha$  jest wymaganą specyfikacją liczbową i musi być liczbą rzeczywistą. Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , gdzie  $\Phi^{-1}$  jest odwrotnością funkcji skumulowanego rozkładu standardowego normalnego. Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Potęgowy.**  $f(x)=x^\alpha$ , jeśli  $\alpha \neq 0$ .  $f(x)=\log(x)$ , jeśli  $\alpha=0$ .  $\alpha$  jest wymaganą specyfikacją liczbową i musi być liczbą rzeczywistą. To połączenie może być używane dla dowolnego rozkładu.

## Uogólnione modele liniowe: Odpowiedź

W wielu przypadkach można po prostu określić zmienną zależną; jednak zmienne, które przyjmują tylko dwie wartości, i odpowiedzi, które rejestrują zdarzenia w próbach, wymagają dodatkowej uwagi.

- **Odpowiedź binarna.** Gdy zmienna binarna przyjmuje tylko dwie wartości, można określić kategorię odniesienia dla oszacowania parametru. Zmienna odpowiedzi binarnej może być łańcuchowa lub liczbowa.
- **Liczba zdarzeń, które wystąpiły w zbiorze prób.** Gdy odpowiedź jest liczbą zdarzeń występujących w zbiorze prób, zmienna zależna zawiera liczbę zdarzeń. Można wybrać dodatkową zmienną zawierającą liczbę prób. Jeśli natomiast liczba prób jest taka sama we wszystkich obiektach, oznacza to, że próby mogły być określone za pomocą wartości stałej. Liczba prób powinna być większa niż lub równa liczbie zdarzeń w każdej obserwacji. Zdarzenia powinny być nieujemnymi liczbami całkowitymi, a próby powinny być dodatnimi liczbami całkowitymi.

W przypadku porządkowych modeli wielomianowych można określić kolejność kategorii odpowiedzi: rosnącą, malejącą lub według danych (co oznacza, że pierwsza wartość napotkana w danych definiuje pierwszą kategorię, a ostatnia — ostatnią kategorię).

**Waga parametru skali.** Parametr skali to oszacowanie parametru modelu w odniesieniu do wariancji odpowiedzi. Wagi parametru skali są wartościami „znanymi”, które mogą się różnić między obserwacjami. Jeśli określona jest zmienna wagi parametru skali, wówczas parametr skali, który jest powiązany z wariancją odpowiedzi, jest dzielony przez jej wartości dla każdej obserwacji. W analizie nie są używane obserwacje z wartościami wagi parametru skali mniejszymi od zera lub równymi zeru ani obserwacje brakujące.

## Uogólnione modele liniowe: Kategoria odniesienia

W przypadku zmiennych dychotomicznych można wybrać kategorię odniesienia dla zmiennej zależnej. Może to wpływać na określone wyniki, takie jak oszacowania parametrów i zapisane wartości, ale nie powinno zmienić dopasowania modelu. Na przykład, jeśli zmienna dychotomiczna przyjmuje wartości 0 i 1:

- Domyślnie procedura wybierze jako kategorię odniesienia ostatnią (o najwyższej wartości) kategorię lub 1. W tej sytuacji prawdopodobieństwa zapisane w modelu oszacują szansę, że dana obserwacja będzie miała wartość 0, a oszacowania parametrów powinny być interpretowane w odniesieniu do wiarygodności kategorii 0.
- Jeśli jako kategoria odniesienia wybrana zostanie pierwsza (o najniższej wartości) kategoria lub 0, wówczas prawdopodobieństwa zapisane w modelu oszacują szansę, że dana obserwacja będzie miała wartość 1.

- Jeśli wybrana zostanie kategoria użytkownika, a zmienna zawiera zdefiniowane etykiety, kategorię odniesienia można ustawić, wybierając wartość z listy. Może to być wygodne, kiedy w trakcie określania modelu, użytkownik nie pamięta dokładnie, jak była zakodowana konkretna zmienna.

## Uogólnione modele liniowe: Predyktory

Karta Predyktory umożliwia określenie czynników i współzmiennych używanych do opracowania efektów modelu i określenia opcjonalnego przesunięcia.

**Czynniki.** Czynniki są predyktorami jakościowymi; mogą być liczbowe lub łańcuchowe.

**Współzmiennie.** Współzmiennie są predyktorami skalarnymi; muszą być liczbowe.

*Uwaga:* gdy odpowiedź jest dwumianowa i ma format binarny, ta procedura oblicza dewiancję i statystyki dobroci dopasowania chi-kwadrat na podstawie podpopulacji, które są oparte na klasyfikacji krzyżowej obserwowanych wartości wybranych czynników i współzmiennych. W wielu przebiegach tej procedury należy zachować ten sam zestaw predyktorów w celu zapewnienia spójnej liczby podpopulacji.

**Przesunięcie.** Składnik przesunięcia jest predyktorem „strukturalnym”. Jego wskaźnik nie jest szacowany przez model, ale przyjmuje się, że ma wartość 1; dlatego wartości przesunięcia są po prostu dodawane do predyktora liniowego zmiennej przewidywanej. Jest to szczególnie przydatne w modelach regresji Poissona, w których każda obserwacja może mieć inny poziom ekspozycji na badane zdarzenie.

Na przykład przy modelowaniu częstości wypadków wśród poszczególnych kierowców należy pamiętać o istotnej różnicy między kierowcą, który miał jeden wypadek w ciągu trzech lat, a kierowcą, który miał jeden wypadek w ciągu 25 lat! Liczba wypadków może być modelowana jako odpowiedź o rozkładzie Poissona lub odpowiedź o rozkładzie ujemnym dwumianowym z logarytmiczną funkcją łączenia, jeśli logarytm naturalny doświadczenia kierowcy jest uwzględniony w składniku przesunięcia.

Inne kombinacje typów rozkładu i funkcji łączenia będą wymagały przekształcenia zmiennej przesunięcia.

## Uogólnione modele liniowe: Opcje

Te opcje są stosowane względem wszystkich czynników określonych na karcie Predyktory.

**Braki danych zdefiniowane przez użytkownika.** Czynniki 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 czynnikowych.

**Porządek kategorii.** Jest istotny podczas określania ostatniego poziomu czynnika, który może być powiązany z parametrem nadmiarowym w algorytmie oszacowania. Zmiana porządku kategorii może spowodować zmianę wpływu poziomu czynnika, ponieważ te oszacowania parametrów są obliczane względem „ostatniego” poziomu. Czynniki mogą być sortowane w kolejności rosnącej od wartości najniższej do najwyższej, kolejności malejącej od wartości najwyższej do najniższej albo w kolejności porządku danych. Oznacza to, że pierwsza wartość napotkana w danych definiuje pierwszą kategorię, a ostatnia napotkana wartość unikalna definiuje ostatnią kategorię.

## Uogólnione modele liniowe: model

### Określ efekty modelu

Model domyślny obejmuje tylko wyraz wolny, dlatego należy jawnie określić inne efekty modelu. Zamiast tego można utworzyć składniki zagnieżdżone lub niezagnieżdżone.

### Składniki niezagnieżdżone

Dla wybranych czynników i współzmiennych:

#### Efekty główne

Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

#### Interakcje

Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu.

### **Czynnikowy**

Tworzy wszystkie możliwe interakcje i efekty główne wybranych zmiennych.

### **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.

## **Składniki zagnieżdżone**

W tej procedurze można zbudować składniki zagnieżdżone dla modelu. Składniki zagnieżdżone są przydatne do modelowania efektu czynników lub współzmiennych, których wartości nie wchodzą w interakcje z poziomami innego czynnika. Na przykład sieć sklepów spożywczych może analizować zwyczaję zakupowe swoich klientów w kilku sklepach. Ponieważ każdy klient bywa regularnie tylko w jednym z tych sklepów, efekt *Klient* jest **zagnieżdżony w** efekcie *Lokalizacja sklepu*.

Ponadto można uwzględnić efekty interakcji, takie jak składniki wielomianowe z tą samą współzmienną, lub dodać wiele poziomów zagnieżdżenia do składnika zagnieżdżonego.

**Ograniczenia:** W odniesieniu do składników zagnieżdżonych obowiązują następujące ograniczenia:

- Wszystkie czynniki w interakcji muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A^*A$ .
- Wszystkie czynniki w efekcie zagnieżdżonym muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A(A)$ .
- Efekt nie może być zagnieżdżony w obrębie współzmiennnej. A zatem, jeśli  $A$  jest czynnikiem, a  $X$  jest współzmienną, to określenie  $A(X)$  jest niedozwolone.

## **Uogólnione modele liniowe: Oszacowanie**

**Estymacja parametru.** Elementy sterujące w tej grupie pozwalają na określenie metod estymacji i udostępniają wartości początkowe dla oszacowania parametrów.

- **Metoda.** Można wybrać metodę estymacji parametru. Można wybrać metodę oceny Newtona-Raphsona, Fishera lub metodę hybrydową, w której interakcje oceny Fishera są wykonywane przed przejściem do metody Newtona-Raphsona. Jeśli zbieżność zostanie osiągnięta w fazie oceny Fishera metodą hybrydową przed wykonaniem maksymalnej liczby iteracji Fishera, algorytm będzie kontynuował działanie, stosując metodę Newtona-Raphsona.
- **Metoda parametru skali.** Można wybrać metodę parametru skali. Metoda maksymalnej wiarygodności umożliwia oszacowanie parametru skali wraz z efektami modelu; należy zauważyć, że ta opcja nie jest odpowiednia, jeśli odpowiedź jest w rozkładzie ujemnym dwumianowym, Poissona, dwumianowym lub w rozkładzie wielomianowym. Opcje odchylenia i chi-kwadratu Pearsona pozwalają oszacować parametr skali na podstawie wartości tych statystyk. Alternatywnie można określić wartość stałą parametru skali.
- **Wartości początkowe.** Procedura automatycznie wyliczy wartości początkowe parametrów. Alternatywnie, użytkownik może sam określić wartości początkowe dla oszacowania parametru.

**Macierz kowariancji.** Dla uogólnionej odwrotności macierzy Hessego estymator oparty na modelu ma wartość ujemną. Ten estymator odporny (nazywany również estymatorem Hubera/White/typu sandwich) jest „skorygowanym” estymatorem na podstawie modelu, który zapewnia spójne oszacowanie kowariancji, nawet w sytuacji, gdy specyfikacja wariancji i funkcji łączenia jest niepoprawna.

**Iteracje.** Dostępne są następujące opcje:



- **Maksymalna liczba iteracji.** Maksymalna liczba iteracji, jakie wykona algorytm. Podaj nieujemną liczbę całkowitą.
- **Maksimum kroków połowienia.** Przy każdej iteracji rozmiar kroku jest zmniejszany o 0,5, aż do momentu wzrostu logarytmu wiarygodności lub osiągnięcia maksymalnej wartości połowienia kroków. Podaj dodatnią liczbę całkowitą.
- **Sprawdź separację punktów danych.** Po wybraniu tej opcji algorytm przeprowadza testy, aby sprawdzić, czy oszacowania parametru mają unikalne wartości. Rozdzielenie następuje, kiedy procedura może stworzyć model, który poprawnie klasyfikuje wszystkie obserwacje. Ta opcja jest dostępna dla zmiennych wielomianowych i dwumianowych w układzie dychotomicznym.

**Regresja logistyczna: Kryteria zbieżności.** Dostępne są następujące opcje

- **Zbieżność parametru.** Po zaznaczeniu tej opcji algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana oszacowania parametru jest mniejsza od określonej wartości, która musi być dodatnia.
- **Zbieżność logarytmu wiarygodności.** Po zaznaczeniu tej opcji algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana funkcji logarytmu wiarygodności jest mniejsza od określonej wartości, która musi być dodatnia.
- **Zbieżność Hessego.** Dla specyfikacji bezwzględnej zakłada się zbieżność, jeśli statystyka w oparciu o zbieżność Hessego jest mniejsza od określonej wartości dodatniej. Dla specyfikacji względnej zakłada się zbieżność, jeśli statystyka jest mniejsza niż iloczyn określonej wartości dodatniej i wartości bezwzględnej logarytmu wiarygodności.

**Tolerancja osobliwości.** Macierze osobliwe (lub nieodwracalne) zawierają liniowo zależne kolumny, które mogą powodować problemy z algorytmem estymacji. Nawet macierze prawie osobliwe mogą powodować uzyskiwanie słabych wyników, dlatego procedura będzie traktować macierze, których wyznacznik jest mniejszy zakres tolerancji, jako osobliwe. Określ wartość dodatnią.

## Uogólnione modele liniowe: Wartości początkowe

Jeśli podane są wartości początkowe, muszą zostać podane dla wszystkich parametrów (w tym parametrów nadmiarowych) w modelu. W zbiorze danych kolejność zmiennych od lewej do prawej musi być następująca: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., gdzie *RowType\_* i *VarName\_* to zmienne łańcuchowe, a *P1*, *P2*, ... są zmiennymi liczbowymi, które odpowiadają uporządkowanej liście parametrów.

- Wartości początkowe są podane w rekordzie z wartością *EST* dla zmiennej *RowType\_*; rzeczywiste wartości początkowe są podane w zmiennych *P1*, *P2*, .... Ta procedura ignoruje wszystkie rekordy, dla których *RowType\_* ma wartość inną niż *EST*, a także wszelkie rekordy poza pierwszym wystąpieniem *RowType\_* równym *EST*.
- Pierwszymi wartościami początkowymi muszą być wyraz wolny, jeśli jest uwzględniony w modelu, albo wartości progowe, jeśli odpowiedź ma rozkład wielomianowy.
- Wartością początkową, którą należy określić jako ostatnią jest parametr skali, a jeśli odpowiedź ma rozkład ujemny dwumianowy — również ujemny parametr dwumianu.
- Jeśli obowiązuje opcja dzielenia pliku, zmienne muszą zaczynać się od zmiennej *split-file* albo od zmiennych w kolejności podanej podczas dzielenia pliku, po czym powinny następować zmienne *RowType\_*, *VarName\_*, *P1*, *P2*, ... jak powyżej. Podziały muszą następować w podanym zbiorze wartości w tej samej kolejności, co w oryginalnym zbiorze wartości. .

*Uwaga:* Nazwy zmiennych *P1*, *P2*, ... nie są wymagane; procedura zaakceptuje dowolne poprawne nazwy zmiennych dla parametrów, ponieważ mapowanie zmiennych na parametry jest oparte na pozycji zmiennej, a nie na jej nazwie. Wszelkie zmienne poza ostatnim parametrem są ignorowane.

Struktura pliku dla wartości początkowych jest taka sama, jak używana podczas eksportowania modelu w postaci danych; dzięki temu wartości końcowe z jednego przebiegu procedury można wykorzystać jako początkowe w kolejnym przebiegu.

## Uogólnione modele liniowe: Statystyki

**Efekty modelu.** Dostępne są następujące opcje:

- **Typ analizy.** Określa typ analizy, jaka ma zostać utworzona. Analiza typu I jest ogólnie odpowiednia w przypadku znanych przyczyn a priori dot. sortowania predyktorów w modelu; typ III ma szersze zastosowanie. Statystyki Walda lub ilorazu wiarygodności są obliczane na podstawie wyboru dokonanego w grupie statystyk chi-kwadrat.
- **Przedziały ufności.** Określa poziom ufności większy niż 50 i mniejszy niż 100. Przedziały Walda są tworzone przy założeniu, że parametry mają asymptotyczny rozkład normalny; przedziały wiarygodności profilu są bardziej dokładne, ale wymagające obliczeniowo. Poziom tolerancji dla przedziałów wiarygodności profilu stanowi kryterium zatrzymywania algorytmu iteracyjnego używanego do obliczenia interwałów.
- **Funkcja logarytmu wiarygodności.** Kontroluje format wyświetlania funkcji logarytmu wiarygodności. Pełna funkcja obejmuje dodatkowy składnik, który jest stały w odniesieniu do oszacowań parametrów; nie ma wpływu na oszacowanie parametrów i w niektórych produktach oprogramowania nie jest wyświetlany.

**Drukuj.** Dostępne są następujące wyniki:

- **Podsumowanie przetwarzania przypadku.** Wyświetla liczbę i wartość procentową obserwacji uwzględnionych w analizie i z niej wykluczonych oraz tabelę Podsumowanie danych skorelowanych.
- **Statystyki opisowe.** Wyświetla statystyki opisowe i informacje podsumowujące dotyczące zmiennej zależnej, współzmiennych i czynników.
- **Informacje o modelu.** Wyświetla nazwę zbioru danych, zmienną zależną lub zdarzenia i zmienne prób, zmienną przesunięcia, ilościową zmienną ważącą, rozkład prawdopodobieństwa i funkcję łączenia.
- **Statystyki dobroci dopasowania.** Wyświetla odchylenie i skalowane odchylenie, chi-kwadrat Pearsona i skalowany chi-kwadrat Pearsona, logarytm wiarygodności, AIC (kryterium informacyjne Akaike), AICC (skończone skorygowane AIC próby), BIC (Bayesowskie kryterium informacyjne) oraz CAIC (spójne AIC).
- **Statystyki podsumowujące modelu.** Wyświetla testy dopasowania modelu, z uwzględnieniem statystyki ilorazu wiarygodności dla testu typu omnibus dopasowania modelu oraz statystyki dla kontrastu typu I lub III dla każdego efektu.
- **Oceny parametrów.** Wyświetlane są oszacowania parametrów i odpowiednie statystyki testu oraz przedziały ufności. Opcjonalnie oprócz surowych oszacowań parametrów można wyświetlić wykładnicze oszacowania parametrów.
- **Macierz kowariancji szacowanych parametrów** Wyświetla macierz kowariancji szacowanych parametrów.
- **Macierz korelacji szacowanych parametrów.** Wyświetla macierz korelacji oszacowanych parametrów.
- **Macierz współczynników kontrastów (L).** Wyświetla współczynniki kontrastu dla domyślnych efektów i dla szacowanych średnich brzegowych, o ile zażądano ich na karcie Średnie EM.
- **Ogólne funkcje estymowalne.** Wyświetla macierze do generowania macierzy współczynnika kontrastu (L).
- **Przebieg iteracji.** Wyświetla przebieg iteracji dla oszacowań parametrów i logarytmu wiarygodności oraz drukuje ostatnią ewaluację wektora gradientu i macierzy Hessego. Tabela przebiegu iteracji wyświetla oszacowania parametrów dla co  $n$ -tej iteracji począwszy od iteracji 0 (oszacowanie początkowe), gdzie  $n$  jest wartością przedziału drukowania. Jeśli tworzone jest żądanie przebiegu iteracji, wówczas ostatnia iteracja jest zawsze wyświetlana, niezależnie od wartości  $n$ .
- **Test mnożnikiem Lagrange'a parametru skali lub parametru podległego dla rozkładu ujemnego dwumianowego** Wyświetla statystyki testu mnożnika Lagrange'a służące do oceny ważności parametru skali obliczanego na podstawie odchylenia lub chi-kwadrat Pearsona albo określonego jako stała w przypadku rozkładu normalnego, gamma, odwróconego Gaussa i Tweediego. W przypadku ujemnego rozkładu dwumianowego testuje ustalony parametr pomocniczy.

## Uogólnione modele liniowe: Średnie EM

Ta karta umożliwia wyświetlenie szacowanych średnich brzegowych dla poziomów współczynników i interakcji czynników. Można również zażądać, aby wyświetlona została oszacowana średnia ogólna. Szacowane średnie brzegowe są niedostępne w przypadku porządkowych modeli wielomianowych.

**Czynniki i interakcje.** Ta lista zawiera czynniki określone na karcie Predyktory oraz interakcje czynników określone na karcie Model. Współzmiennie są wykluczone z tej listy. Składniki można wybierać bezpośrednio z tej listy albo włączać do składnika interakcji za pomocą przycisku **Przez \***.

**Pokaż średnie dla.** Szacowane średnie są obliczane dla wybranych czynników i interakcji czynników. Kontrast określa konfigurację testów hipotezy w celu porównania oszacowanych średnich. Kontrast prosty wymaga kategorii odniesienia lub poziomu czynnika, z którym porównywane będą inne.

- **Parami.** Porównania parami są obliczane dla kombinacji wszystkich poziomów czynników podanych lub domniemanych. Jest to jedyny kontrast dostępny dla interakcji czynników.
- *Proste.* 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.
- **Odchylenie.** Każdy poziom czynnika jest porównywany ze średnią ogólną. Kontrasty odchylenia nie są ortogonalne.
- *Różnica.* Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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.
- *Wielomianowe.* Wybranie tego typu kontrastu powoduje porównanie efektu liniowego, efektu kwadratowego, sześciennego 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.

**Powięszczenie.** Szacowane średnie brzegowe mogą być obliczane dla odpowiedzi na podstawie oryginalnej skali zmiennej zależnej albo dla predyktora liniowego na podstawie zmiennej zależnej przekształconej przez funkcję łączenia.

**Korekta dla porównań wielokrotnych.** Podczas testowania hipotez z użyciem kontrastów wielokrotnych całościowy poziom istotności można skorygować na podstawie poziomów istotności dla uwzględnionych kontrastów. W tej grupie możliwy jest wybór metody korekty.

- **Najmniejsza istotna różnica.** Ta metoda nie kontroluje ogólnego prawdopodobieństwa odrzucenia hipotez, które stwierdzają, że niektóre kontrasty liniowe różnią się od wartości hipotezy zerowej.
- *Bonferroni.* Metoda ta dostosowuje obserwowany poziom istotności do faktu przeprowadzania wielu porównań.
- *Sekwencyjna Bonferroniego.* Jest to sekwencyjne zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.
- *Sidak.* Metoda ta zapewnia węższe granice niż test Bonferroniego.
- *Sekwencyjna Sidaka.* Jest to sekwencyjnie zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.

## Uogólnione modele liniowe: Zapisz

Zaznaczone elementy są zapisywane pod określoną nazwą; można nakazać nadpisywanie istniejących zmiennych o tej samej nazwie nowymi zmiennymi albo unikanie kolizji nazw poprzez dołączanie przyrostków, tak aby nazwy nowych zmiennych były unikalne.

### Przewidywana wartość średniej dla zależnej

Zapisuje przewidywane przez model wartości poszczególnych obserwacji w pierwotnej metryce odpowiedzi. Gdy rozkład odpowiedzi jest dwumianowy, a zmienna zależna jest binarna, procedura zapisuje przewidywane prawdopodobieństwa. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **Łączne przewidywane prawdopodobieństwo**, a procedura zapisuje łączne przewidywane prawdopodobieństwo każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

### Dolna granica przedziału ufności średniej dla zależnej

Zapisuje dolną granicę przedziału ufności średniej odpowiedzi. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **Dolna granica przedziału ufności dla skumulowanego przewidywanego prawdopodobieństwa**, a procedura zapisuje dolną granicę dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

### Górna granica przedziału ufności średniej dla zależnej

Zapisuje górną granicę przedziału ufności średniej odpowiedzi. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **Górna granica przedziału ufności dla skumulowanego przewidywanego prawdopodobieństwa**, a procedura zapisuje górną granicę dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

### Przewidywana kategoria

W przypadku modeli z rozkładem dwumianowym i binarnej zmiennej zależnej lub rozkładu wielomianowego zapisywane są przewidywane kategorie odpowiedzi dla każdej obserwacji. Opcja ta jest niedostępna w przypadku innych rozkładów odpowiedzi.

### Przewidywana wartość predyktora liniowego

Zapisuje przewidywane przez model wartości poszczególnych obserwacji w metryce predyktora liniowego (odpowiedzi przekształconej przez określoną funkcję łączenia). Gdy rozkład odpowiedzi jest wielomianowy, procedura zapisuje wartość przewidywaną dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

### Oszacowany błąd standardowy wartości przewidywanej predyktora liniowego

Gdy rozkład odpowiedzi jest wielomianowy, procedura zapisuje błąd standardowy dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

Następujące elementy są niedostępne, gdy odpowiedzi mają rozkład wielomianowy.

### 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.

### Wartość wpływu

Mierzy wpływ punktu na dopasowanie regresji. Wycentrowane wartości wpływu zawierają się w przedziale od 0 (brak wpływu na dopasowanie) do  $(N-1)/N$ .

### Reszta

Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.

### Reszta Pearsona

Pierwiastek kwadratowy udziału obserwacji w statystyce chi-kwadrat Pearsona, ze znakiem reszty surowej.

### Standaryzowana reszta Pearsona

Reszta Pearsona pomnożona przez pierwiastek kwadratowy odwrotności iloczynu parametru skali i 1-wpływ obserwacji.

### Reszta odchylenia

Pierwiastek kwadratowy udziału obserwacji w statystyce Odchylenie, ze znakiem reszty surowej.

### Standaryzowana reszta odchylenia

Reszta odchylenia pomnożona przez pierwiastek kwadratowy odwrotności iloczynu parametru skali i 1-wpływ obserwacji.

### Reszta wiarygodności

Pierwiastek kwadratowy ważonej średniej (opartej na wpływie obserwacji) kwadratów reszt standaryzowanych Pearsona i reszt standaryzowanych odchylenia, ze znakiem reszty surowej.

## Uogólnione modele liniowe: Eksportuj

**Eksportuj model jako dane.** Zapisuje zbiór danych w formacie IBM SPSS Statistics zawierającym korelację parametru lub macierz kowariancji z oszacowaniami parametrów, błędami standardowymi, wartościami istotności oraz stopniami swobody. Kolejność zmiennych w pliku macierzowym jest następująca.

- **Zmienne dzielące.** Jeśli są używane, wszelkie zmienne definiujące zostają podzielone.
- **RowType\_.** Przyjmuje wartości (i etykiety wartości) *COV* (kowariancje), *CORR* (korelacje), *EST* (oszacowania parametrów), *SE* (błędy standardowe), *SIG* (poziomy istotności) i *DF* (stopnie swobody planu losowania). Dla każdego parametru modelu istnieje osobna obserwacja z typem wiersza *COV* (lub *CORR*), a także osobna obserwacja dla każdego z pozostałych typów wierszy.
- **VarName\_.** Przyjmuje wartości *P1*, *P2*, ..., odpowiadające uporządkowanej liście wszystkich oszacowanych parametrów modelu (z wyjątkiem parametrów skali lub ujemnych parametrów dwumianu) — dla wierszy typów *COV* lub *CORR*, z wartościami etykiet odpowiadającymi ciągom znaków parametrów, które są pokazane w tabeli Oszacowania parametru. Te komórki są puste w przypadku wierszy innych typów.
- **P1, P2, ...** Te zmienne odpowiadają uporządkowanej liście wszystkich parametrów modelu (z wyjątkiem, odpowiednio, parametrów skali lub ujemnych parametrów dwumianu) z wartościami etykiet odpowiadającymi ciągom znaków parametrów, które są pokazane w tabeli Oszacowania parametru, a ponadto przyjmują wartości zgodnie z typem wiersza.

W odniesieniu do parametrów nadmiarowych wszystkie kowariancje są ustawione na zero, korelacje są ustawione na systemowy brak danych; wszystkie oszacowania parametrów są ustawione na zero, a wszystkie błędy standardowe, poziomy istotności i stopnie swobody reszt są ustawione na systemowy brak danych.

W przypadku parametru skali kowariancje, korelacje, poziom istotności oraz poziomy swobody są ustawione na systemowy brak danych. Jeśli estymacja parametru skali zostanie przeprowadzona metodą największej wiarygodności, błąd standardowy będzie podany; w przeciwnym wypadku zostanie ustawiony na systemowy brak danych.

W przypadku ujemnego parametru dwumianu kowariancje, korelacje, poziom istotności oraz stopnie swobody są ustawione na systemowy brak danych. Jeśli estymacja ujemnego parametru dwumianu zostanie przeprowadzona metodą największej wiarygodności, błąd standardowy będzie podany; w przeciwnym wypadku zostanie ustawiony na systemowy brak danych.

Jeśli występują podziały, lista parametrów musi zostać skumulowana po wszystkich podziałach. W konkretnym podziale niektóre parametry mogą być nieistotne; nie oznacza to jednak, że są nadmiarowe. W przypadku parametrów nieistotnych wszystkie kowariancje i korelacje, oszacowania parametrów, błędy standardowe, poziomy istotności oraz poziomy swobody są ustawione na systemowy brak danych.

Można użyć tego pliku macierzowego jako wartości początkowych do dalszej estymacji modelu; tego pliku nie można użyć bezpośrednio do dalszych analiz w innych procedurach, które odczytują plik macierzowy, chyba że te procedury akceptują wszystkie typy wierszy eksportowane w tym przypadku. Nawet wówczas należy zadbać o to, aby wszystkie parametry w tym pliku macierzowym miały to samo znaczenie dla procedury odczytującej plik.

**Eksportuj model jako XML.** Jeśli ta opcja jest wybrana, oszacowania parametrów oraz macierz kowariancji parametrów są zapisywane w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

## Dodatkowe właściwości komendy GENLIN

Język składni komend umożliwia również:

- Określanie wartości początkowych dla oszacowań parametru w postaci listy liczb (za pomocą opcji komendy CRITERIA).
- Ustalanie współzmiennych na wartościach innych niż ich średnie podczas obliczania szacowanych średnich brzegowych (za pomocą opcji komendy EMMEANS).
- Określanie niestandardowych kontrastów wielomianowych dla szacowanych średnich brzegowych (za pomocą opcji komendy EMMEANS).
- Określanie podzbioru czynników, dla których szacowane średnie brzegowe są wyświetlane w celu porównywania z użyciem podanego typu kontrastu (za pomocą słów kluczowych TABLES i COMPARE opcji komendy EMMEANS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Ogólne równania estymujące

Procedura ogólnych równań estymujących rozszerza ogólny model liniowy w taki sposób, że pozwala na analizę powtarzanych pomiarów lub innych skorelowanych obserwacji, takich jak dane zgrupowane.

**Przykład.** Instytucja zdrowia publicznego może użyć ogólnych równań estymujących w celu dopasowania regresji logistycznej z powtarzanymi pomiarami w ramach badania wpływu zanieczyszczenia atmosfery na zdrowie dzieci.

Uogólnione równania estymujące: Zagadnienia dotyczące danych

**Dane.** Odpowiedź może być zmienną ilościową, liczebnościową, binarną lub określającą zdarzenia w próbach. Zakłada się, że czynniki są jakościowe. Przyjmuje się, że współzmiennie, waga parametru skali i przesunięcie są ilościowe. Zmiennych definiujących obiekty lub pomiary powtarzane w ramach jednego obiektu nie można używać do definiowania odpowiedzi, ale mogą one pełnić inne role w modelu.

**Założenia.** Zakłada się, że obserwacje są zależne w ramach obiektów i niezależne między obiektami. Jako część modelu szacowana jest macierz korelacji odzwierciedlająca zależności w ramach obiektów.

Uzyskiwanie uogólnionych równań estymujących

Wybierz z menu następującą opcję:

**Analiza > Uogólnione modele liniowe > Ogólne równania estymujące...**

1. Wybierz jedną lub więcej zmiennych obiektowych (dodatkowe opcje opisano poniżej).

Kombinacja wartości wybranych zmiennych powinna jednoznacznie definiować **obiekty** w zbiorze danych. Na przykład jedna zmienna *ID pacjenta* powinna wystarczyć do definicji obiektów w jednym szpitalu. Jednak jeśli w różnych szpitalach pacjenci nie mają nadawanych różnych identyfikatorów, konieczne może być połączenie zmiennych *ID szpitala* i *ID pacjenta*. Przy wielokrotnych pomiarach dla każdego obiektu zapisywane są wielokrotne obserwacje. Z tego powodu jeden obiekt może być w zbiorze danych przedstawiany w wielu obserwacjach.

2. Na karcie Typ modelu określ rozkład i funkcję łączenia.
3. Na karcie Odpowiedź wybierz zmienną zależną.
4. Na karcie Predyktory wybierz czynniki i współzmiennie, które mają być używane do predykcji zmiennej zależnej.
5. Na karcie Model określ efekty modelu przy użyciu wybranych czynników i współzmiennych.

Opcjonalnie na karcie Powtórzone można określić:

**Zmienne wewnątrzobektowe.** Kombinacja wartości zmiennych wewnątrzobektowych definiuje kolejność pomiarów w obiektach. Zatem połączenie zmiennych obiektu i wewnątrzobektowych stanowi unikatową definicję każdego pomiaru. Na przykład połączenie zmiennych *Okres*, *ID szpitala* i *ID pacjenta* definiuje dla każdego przypadku określoną wizytę w gabinecie, złożoną przez określonego pacjenta w określonym szpitalu.

Nie jest konieczne określanie zmiennej wewnątrzobektowej, jeśli zbiór danych jest już posortowany i pomiary dotyczące każdego obiektu stanowią jeden ciągły, prawidłowo ułożony zbiór obserwacji; w takim przypadku można anulować wybór opcji **Sortuj obserwacje wg zmiennych obiektowych i wewnątrzobektowych** i zaoszczędzić czas, eliminując konieczność (tymczasowego) sortowania. Z reguły wskazane jest wykorzystanie zmiennych wewnątrzobektowych do zapewnienia prawidłowego uporządkowania pomiarów.

Obiektów w zmiennych wewnątrzobektowych nie można używać do definiowania odpowiedzi, ale mogą one pełnić inne funkcje w modelu. Na przykład zmiennej *ID szpitala* można użyć jako czynnika w modelu.

**Macierz kowariancji.** Dla uogólnionej odwrotności macierzy Hessego estymator oparty na modelu ma wartość ujemną. Estymator odporny (nazwany także estymatorem Hubera/White/typu sandwich) jest „skorygowanym” estymatorem na podstawie modelu, który zapewnia spójne oszacowanie kowariancji nawet w sytuacji, gdy robocza macierz korelacji jest nieprawidłowo określona. Ta specyfikacja ma zastosowanie do parametrów w części modelu liniowego ogólnych równań estymujących, natomiast specyfikacja na karcie Estymacja ma zastosowanie tylko do początkowego uogólnionego modelu liniowego.

**Robocza macierz korelacji.** Ta macierz korelacji przedstawia zależności wewnątrzobektowe. Jej rozmiar jest określany przez liczbę pomiarów, czyli przez liczbę połączeń wartości zmiennych wewnątrzobektowych. Można wybrać jedną z poniższych struktur:

- **Niezależna.** Powtarzane pomiary są nieskorelowane.
- **AR(1).** Powtarzane pomiary wykazują relację autoregresyjną pierwszego rzędu. Korelacja między dwoma elementami odpowiada  $\rho$  dla sąsiednich elementów,  $\rho^2$  dla elementów oddzielonych jednym elementem itd. jest ograniczone, tak że  $-1 < \rho < 1$ .
- **Zamienne.** Struktura ma jednorodne korelacje między elementami. Znana również jako struktura symetrii złożonej.
- **M-zależna.** Kolejno następujące po sobie pomiary mają wspólny współczynnik korelacji, pary pomiarów oddzielone trzecim pomiarem mają wspólny współczynnik korelacji itd., dla wszystkich par pomiarów oddzielonych  $m-1$  innymi pomiarami. Załóżmy na przykład, że co roku uczniowie w klasach od 3. do 7. piszą zestandaryzowany test. W takiej strukturze zakłada się, że oceny klasy 3. i 4., 4. i 5., 5. i 6. oraz 6. i 7. będą miały tę samą korelację; podobnie oceny klasy 3. i 5., 4. i 6. oraz 5. i 7. będą miały tę samą korelację; jak również oceny klasy 3. i 6. oraz 4. i 7. będą miały tę samą korelację. Zakłada się, że pomiary oddzielone więcej niż  $m$  innymi pomiarami są nieskorelowane. Wybierając tę strukturę, należy określić wartość  $m$  mniejszą od rzędu roboczej macierzy korelacji.
- **Nieustrukturalizowana.** Jest to w pełni ogólna macierz kowariancji.

Domyślnie procedura skoryguje oszacowania korelacji o liczbę nienadmiarowych parametrów. Usunięcie tej korekty może być pożądane, jeśli chcemy, by oszacowania były niezmiennie względem zmian replikacji na poziomie obiektów.

- **Maksymalna liczba iteracji.** Maksymalna liczba iteracji, jakie wykona algorytm ogólnych równań estymujących. Podaj nieujemną liczbę całkowitą. Ta specyfikacja ma zastosowanie do parametrów w części modelu liniowego ogólnych równań estymujących, natomiast specyfikacja na karcie Estymacja ma zastosowanie tylko do początkowego uogólnionego modelu liniowego.
- **Aktualizuj macierz.** Elementy w roboczej macierzy korelacji są szacowane na podstawie oszacowań parametrów, które są aktualizowane w każdej iteracji algorytmu. Jeśli robocza macierz korelacji w ogóle nie zostanie zaktualizowana, to w całym procesie estymacji używana jest początkowa robocza macierz korelacji. Jeśli macierz jest aktualizowana, można określić interwał iteracji, od którego będzie zależeć częstotliwość aktualizacji elementów roboczej macierzy korelacji. Określenie wartości większej od 1 może przyspieszyć przetwarzanie.

**Regresja logistyczna: Kryteria zbieżności.** Te specyfikacje mają zastosowanie do parametrów w części modelu liniowego ogólnych równań estymujących, natomiast specyfikacja na karcie Estymacja ma zastosowanie tylko do początkowego uogólnionego modelu liniowego.

- **Zbieżność parametru.** Po zaznaczeniu tej opcji algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana oszacowania parametru jest mniejsza od określonej wartości, która musi być dodatnia.

- **Zbieżność Hessego.** Zakłada się istnienie zbieżności, jeśli statystyka oparta na macierzy Hessego jest mniejsza niż określona wartość, która musi być dodatnia.

## Ogólne równania estymujące: Typ modelu

Karta Typ modelu umożliwia określenie rozkładu i funkcji łączenia dla modelu. Zawiera skróty odpowiednie dla kilku typowych modeli podzielonych na kategorie wg typu odpowiedzi.

Typy modeli

**Ilościowa zależna.** Dostępne są następujące opcje:

- **Liniowy.** Określa Normalny jako rozkład i Tożsamość jako funkcję łączenia.
- **Gamma wiązany logarytmicznie.** Określa Gamma jako rozkład i Log jako funkcję łączenia.

**Porządkowa zależna.** Dostępne są następujące opcje:

- **Porządkowy logistyczny.** Określa Wielomianowy (porządkowy) jako rozkład i Logit Skumulowane jako funkcję łączenia.
- **Porządkowy probit.** Określa Wielomianowy (porządkowy) jako rozkład i Probit Skumulowane jako funkcję łączenia.

**Liczebności.** Dostępne są następujące opcje:

- **Poissona logliniowy.** Określa Poissona jako rozkład i Log jako funkcję łączenia.
- **Ujemny dwumianowy wiązany logarytmicznie.** Określa Ujemny dwumianowy (z wartością 1 dla parametru dodatkowego) jako rozkład i Log jako funkcję łączenia. Aby procedura oszacowała wartość parametru dodatkowego, należy podać model niestandardowy z rozkładem ujemnym dwumianowym, a następnie wybrać opcję **Szacuj wartość** w grupie Parametr.

**Binarna zależna lub dane ze zdarzeń/prób.** Dostępne są następujące opcje:

- **Binarny logistyczny.** Określa Dwumianowy jako rozkład i Logit jako funkcję łączenia.
- **Binarny probit.** Określa Dwumianowy jako rozkład i Probit jako funkcję łączenia.
- **Przeżycia obcięte przedziałowe.** Określa Dwumianowy jako rozkład i Komplementarny Log-log jako funkcję łączenia.

**Mieszany.** Dostępne są następujące opcje:

- **Tweediego wiązany logarytmicznie.** Określa Tweediego jako rozkład i Log jako funkcję łączenia.
- **Tweediego wiązany tożsamościowo.** Określa Tweediego jako rozkład i Tożsamość jako funkcję łączenia.

**Użytkownika.** Umożliwia określenie własnej kombinacji funkcji rozkładu i łączenia.

Dystrybucja

Ten wybór określa rozkład zmiennej zależnej. Możliwość określenia rozkładu innego niż normalny i nietożsamościowej funkcji łączenia jest istotnym ulepszeniem uogólnionego modelu liniowego w porównaniu do ogólnego modelu liniowego. Istnieje wiele możliwych kombinacji rozkład-funkcja łączenia, a kilka z nich może być odpowiednich dla dowolnego zbioru danych, dlatego wybór może być zależny od rozważań teoretycznych a priori lub tego, która kombinacja wydaje się zapewniać najlepsze dopasowanie.

- **Dwumianowy.** Ten rozkład jest odpowiedni tylko dla zmiennych, które reprezentują zmienne dychotomiczne lub liczbę zdarzeń.
- **Gamma.** Ten rozkład jest odpowiedni dla zmiennych z dodatnimi wartościami skali, które są skośne w kierunku większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Odwrócony Gaussa.** Ten rozkład jest odpowiedni dla zmiennych z dodatnimi wartościami skali, które są skośne w kierunku większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.



- **Ujemny dwumianowy.** Ten rozkład może być traktowany jako seria prób wymaganych do zaobserwowania  $k$  sukcesów i jest odpowiedni dla zmiennych z nieujemnymi liczbami całkowitymi. Jeśli wartość danych nie jest liczbą całkowitą, jest mniejsza od 0 lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie. Wartość parametru dodatkowego ujemnego rozkładu dwumianowego może być dowolną liczbą większą niż lub równą 0; można ustawić ją jako wartość stałą lub zezwolić na jej oszacowanie w czasie trwania procedury. Jeśli parametr dodatkowy jest ustawiony na 0, użycie tego rozkładu da takie same efekty, jak użycie rozkładu Poissona.
- **Normalny.** Ten rozkład jest odpowiedni dla zmiennych ilościowych, których wartości rozkładają się symetrycznie, w kształcie dzwonu, wokół wartości centralnej (średniej). Zmienna zależna musi być typu liczbowego.
- **Poissona.** Ten rozkład można traktować jako liczbę wystąpień zdarzenia badanego w ustalonym okresie i jest odpowiedni dla zmiennych o nieujemnych wartościach całkowitych. Jeśli wartość danych nie jest liczbą całkowitą, jest mniejsza od 0 lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Tweedie'go.** Ten rozkład jest odpowiedni dla zmiennych, które mogą być reprezentowane przez poissonowsko mieszane rozkłady gamma; rozkład ten jest mieszany, to znaczy że łączy właściwości rozkładu ciągłego (nieujemne wartości rzeczywiste) i dyskretnego (prawdopodobieństwo dodatnie dla pojedynczej wartości, 0). Zmienna zależna musi być liczbowa, z wartościami danych większymi niż lub równymi zero. Jeśli wartość danych jest mniejsza niż zero lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie. Wartość stała parametru rozkładu Tweedie'go może być dowolną liczbą większą niż jeden i mniejszą niż dwa.
- **Wielomianowy.** Ten rozkład jest odpowiedni dla zmiennych, które reprezentują odpowiedzi porządkowe. Zmienna zależna może być liniowa lub łańcuchowa i musi zawierać co najmniej dwie różniące się poprawne wartości danych.

Funkcja łączenia

Funkcja łączenia to przekształcenie zmiennej zależnej, które umożliwia estymację modelu. Dostępne są następujące funkcje:

- **Tożsamość.**  $f(x)=x$ . Zmienna zależna nie jest przekształcana. To połączenie może być używane dla dowolnego rozkładu.
- **Komplementarny logarytmiczno-logarytmiczny.**  $f(x)=\log(-\log(1-x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Skumulowany Cauchit.**  $f(x) = \tan(\pi(x-0.5))$ , stosowany do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany logarytmiczno-logarytmiczny dopełnienia.**  $f(x)=\ln(-\ln(1-x))$  ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany logit.**  $f(x)=\ln(x/(1-x))$ ; ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany ujemny logarytmiczno-logarytmiczny.**  $f(x) = -\ln(-\ln(x))$ ; ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Skumulowany probit.**  $f(x) = \Phi^{-1}(x)$  ma zastosowanie do skumulowanego prawdopodobieństwa dla każdej kategorii odpowiedzi, gdzie  $\Phi^{-1}$  jest odwrotnością funkcji skumulowanego rozkładu standardowego normalnego. Ma zastosowanie tylko w przypadku rozkładu wielomianowego.
- **Logarytm.**  $f(x)=\log(x)$ . To połączenie może być używane dla dowolnego rozkładu.
- **Logarytmiczny dopełnienia.**  $f(x)=\log(1-x)$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Logit.**  $f(x)=\log(x/(1-x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Ujemny dwumianowy.**  $f(x)=\log(x/(x+k^{-1}))$ , gdzie  $k$  to parametr dodatkowy ujemnego rozkładu dwumianowego. Ma zastosowanie tylko w przypadku ujemnego rozkładu dwumianowego.

- **Ujemny logarytmiczno-logarytmiczny.**  $f(x) = -\log(-\log(x))$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Nieparzysty potęgowy.**  $f(x) = [(x/(1-x))^\alpha - 1]/\alpha$ , jeśli  $\alpha \neq 0$ .  $f(x) = \log(x)$ , jeśli  $\alpha = 0$ .  $\alpha$  jest wymaganą specyfikacją liczbową i musi być liczbą rzeczywistą. Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Probit.**  $f(x) = \Phi^{-1}(x)$ , gdzie  $\Phi^{-1}$  jest odwrotnością funkcji skumulowanego rozkładu standardowego normalnego. Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Potęgowy.**  $f(x) = x^\alpha$ , jeśli  $\alpha \neq 0$ .  $f(x) = \log(x)$ , jeśli  $\alpha = 0$ .  $\alpha$  jest wymaganą specyfikacją liczbową i musi być liczbą rzeczywistą. To połączenie może być używane dla dowolnego rozkładu.

## Ogólne równania estymujące: Odpowiedź

W wielu przypadkach można po prostu określić zmienną zależną; jednak zmienne, które przyjmują tylko dwie wartości, i odpowiedzi, które rejestrują zdarzenia w próbach, wymagają dodatkowej uwagi.

- **Odpowiedź binarna.** Gdy zmienna binarna przyjmuje tylko dwie wartości, można określić kategorię odniesienia dla oszacowania parametru. Zmienna odpowiedzi binarnej może być łańcuchowa lub liczbowa.
- **Liczba zdarzeń, które wystąpiły w zbiorze prób.** Gdy odpowiedź jest liczbą zdarzeń występujących w zbiorze prób, zmienna zależna zawiera liczbę zdarzeń. Można wybrać dodatkową zmienną zawierającą liczbę prób. Jeśli natomiast liczba prób jest taka sama we wszystkich obiektach, oznacza to, że próby mogły być określone za pomocą wartości stałej. Liczba prób powinna być większa niż lub równa liczbie zdarzeń w każdej obserwacji. Zdarzenia powinny być nieujemnymi liczbami całkowitymi, a próby powinny być dodatnimi liczbami całkowitymi.

W przypadku porządkowych modeli wielomianowych można określić kolejność kategorii odpowiedzi: rosnącą, malejącą lub według danych (co oznacza, że pierwsza wartość napotkana w danych definiuje pierwszą kategorię, a ostatnia — ostatnią kategorię).

**Waga parametru skali.** Parametr skali to oszacowanie parametru modelu w odniesieniu do wariancji odpowiedzi. Wagi parametru skali są wartościami „znanymi”, które mogą się różnić między obserwacjami. Jeśli określona jest zmienna wagi parametru skali, wówczas parametr skali, który jest powiązany z wariancją odpowiedzi, jest dzielony przez jej wartości dla każdej obserwacji. W analizie nie są używane obserwacje z wartościami wagi parametru skali mniejszymi od zera lub równymi zeru ani obserwacje brakujące.

## Uogólnione równania estymujące-kategoria odniesienia

W przypadku zmiennych dychotomicznych można wybrać kategorię odniesienia dla zmiennej zależnej. Może to wpływać na określone wyniki, takie jak oszacowania parametrów i zapisane wartości, ale nie powinno zmienić dopasowania modelu. Na przykład, jeśli zmienna dychotomiczna przyjmuje wartości 0 i 1:

- Domyślnie procedura wybierze jako kategorię odniesienia ostatnią (o najwyższej wartości) kategorię lub 1. W tej sytuacji prawdopodobieństwa zapisane w modelu oszacują szansę, że dana obserwacja będzie miała wartość 0, a oszacowania parametrów powinny być interpretowane w odniesieniu do wiarygodności kategorii 0.
- Jeśli jako kategoria odniesienia wybrana zostanie pierwsza (o najniższej wartości) kategoria lub 0, wówczas prawdopodobieństwa zapisane w modelu oszacują szansę, że dana obserwacja będzie miała wartość 1.
- Jeśli wybrana zostanie kategoria użytkownika, a zmienna zawiera zdefiniowane etykiety, kategorię odniesienia można ustawić, wybierając wartość z listy. Może to być wygodne, kiedy w trakcie określania modelu, użytkownik nie pamięta dokładnie, jak była zakodowana konkretna zmienna.

## Ogólne równania estymujące: Predyktory

Karta Predyktory umożliwia określenie czynników i współzmiennych używanych do opracowania efektów modelu i określenia opcjonalnego przesunięcia.

**Czynniki.** Czynniki są predyktorami jakościowymi; mogą być liczbowe lub łańcuchowe.

**Współzmiennie.** Współzmiennie są predyktorami skalarnymi; muszą być liczbowe.

*Uwaga:* gdy odpowiedź jest dwumianowa i ma format binarny, ta procedura oblicza dewiancję i statystyki dobroci dopasowania chi-kwadrat na podstawie podpopulacji, które są oparte na klasyfikacji krzyżowej obserwowanych wartości wybranych czynników i współzmiennych. W wielu przebiegach tej procedury należy zachować ten sam zestaw predyktorów w celu zapewnienia spójnej liczby podpopulacji.

**Przesunięcie.** Składnik przesunięcia jest predyktorem „strukturalnym”. Jego wskaźnik nie jest szacowany przez model, ale przyjmuje się, że ma wartość 1; dlatego wartości przesunięcia są po prostu dodawane do predyktora liniowego zmiennej przewidywanej. Jest to szczególnie przydatne w modelach regresji Poissona, w których każda obserwacja może mieć inny poziom ekspozycji na badane zdarzenie.

Na przykład przy modelowaniu częstości wypadków wśród poszczególnych kierowców należy pamiętać o istotnej różnicy między kierowcą, który miał jeden wypadek w ciągu trzech lat, a kierowcą, który miał jeden wypadek w ciągu 25 lat! Liczba wypadków może być modelowana jako odpowiedź o rozkładzie Poissona lub odpowiedź o rozkładzie ujemnym dwumianowym z logarytmiczną funkcją łączenia, jeśli logarytm naturalny doświadczenia kierowcy jest uwzględniony w składniku przesunięcia.

Inne kombinacje typów rozkładu i funkcji łączenia będą wymagały przekształcenia zmiennej przesunięcia.

## Opcje ogólnych równań estymujących

Te opcje są stosowane względem wszystkich czynników określonych na karcie Predyktory.

**Braki danych zdefiniowane przez użytkownika.** Czynniki 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 czynnikowych.

**Porządek kategorii.** Jest istotny podczas określania ostatniego poziomu czynnika, który może być powiązany z parametrem nadmiarowym w algorytmie oszacowania. Zmiana porządku kategorii może spowodować zmianę wpływu poziomu czynnika, ponieważ te oszacowania parametrów są obliczane względem „ostatniego” poziomu. Czynniki mogą być sortowane w kolejności rosnącej od wartości najniższej do najwyższej, kolejności malejącej od wartości najwyższej do najniższej albo w kolejności porządku danych. Oznacza to, że pierwsza wartość napotkana w danych definiuje pierwszą kategorię, a ostatnia napotkana wartość unikalna definiuje ostatnią kategorię.

## Ogólne równania estymujące: Model

**Określenie efektów modelu.** Model domyślny obejmuje tylko wyraz wolny, dlatego należy jawnie określić inne efekty modelu. Zamiast tego można utworzyć składniki zagnieżdżone lub niezagnieżdżone.

Składniki niezagnieżdżone

Dla wybranych czynników i współzmiennych:

**Efekty główne.** Dla każdej wybranej zmiennej tworzy składnik efektów głównych.

**Interakcja.** Dla wszystkich wybranych zmiennych tworzy składnik interakcji najwyższego rzędu.

**Czynnikowy.** Tworzy wszystkie możliwe interakcje i efekty główne wybranych zmiennych.

**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.

## Składniki zagnieżdżone

W tej procedurze można zbudować składniki zagnieżdżone dla modelu. Składniki zagnieżdżone są przydatne do modelowania efektu czynników lub współzmiennych, których wartości nie wchodzi w interakcje z poziomami innego czynnika. Na przykład sieć sklepów spożywczych może analizować zwyczaje zakupowe swoich klientów w kilku sklepach. Ponieważ każdy klient bywa regularnie tylko w jednym z tych sklepów, efekt *Klient* jest **zagnieżdżony w** efekcie *Lokalizacja sklepu*.

Ponadto można uwzględnić efekty interakcji lub dodać wiele poziomów zagnieżdżenia do zagnieżdżonego składnika.

**Ograniczenia.** W odniesieniu do składników zagnieżdżonych obowiązują następujące ograniczenia:

- Wszystkie czynniki w interakcji muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A^*A$ .
- Wszystkie czynniki w efekcie zagnieżdżonym muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A(A)$ .
- Efekt nie może być zagnieżdżony w obrębie współzmiennnej. A zatem, jeśli  $A$  jest czynnikiem, a  $X$  jest współzmienną, to określenie  $A(X)$  jest niedozwolone.

**Wyraz wolny.** Wyraz wolny jest zwykle uwzględniany w modelu. Jeśli można założyć, że dane przechodzą przez początek układu współrzędnych, wyraz wolny można wyłączyć z modelu.

Modele z wielomianowym rozkładem porządkowym nie mają jednego wyrazu wolnego; zamiast niego występują w nich parametry progowe, które określają miejsca przejść między sąsiednimi kategoriami. Progi są zawsze uwzględnione w modelu.

## Ogólne równania estymujące: Oszacowanie

**Estymacja parametru.** Elementy sterujące w tej grupie pozwalają na określenie metod estymacji i udostępniają wartości początkowe dla oszacowania parametrów.

- **Metoda.** Można wybrać metodę oszacowania parametru: Newtona-Raphsona, Fishera lub metodę hybrydową, w której interakcje oceny Fishera są wykonywane przed przejściem do metody Newtona-Raphsona. Jeśli zbieżność zostanie osiągnięta w fazie oceny Fishera metodą hybrydową przed wykonaniem maksymalnej liczby iteracji Fishera, algorytm będzie kontynuował działanie, stosując metodę Newtona-Raphsona.
- **Metoda parametru skali.** Można wybrać metodę parametru skali.

Metoda maksymalnej wiarygodności umożliwia oszacowanie parametru skali wraz z efektami modelu; należy zauważyć, że ta opcja nie jest odpowiednia, jeśli odpowiedź jest w rozkładzie ujemnym dwumianowym, Poissona lub rozkładzie dwumianowym. Ponieważ pojęcie wiarygodności nie jest uwzględnione w ogólnych równaniach estymujących, ta specyfikacja dotyczy tylko początkowego uogólnionego modelu liniowego; to oszacowanie parametru skali jest następnie przekazywane do procedury ogólnych równań estymujących, która modyfikuje parametr skali z uwzględnieniem wartości chi-kwadrat Pearsona podzielonej przez liczbę stopni swobody.

Opcje odchylenia i chi-kwadrat Pearsona służą do szacowania parametru skali na podstawie tych statystyk w początkowym ogólnym modelu liniowym; to oszacowanie parametru skali jest następnie przekazywane do procedury ogólnych równań estymujących, która traktuje je jako niezmiennie.

Alternatywnie można określić wartość stałą parametru skali. Będzie ona traktowana jako niezmienna podczas estymacji początkowego uogólnionego modelu liniowego oraz w ogólnych równaniach estymujących.

- **Wartości początkowe.** Procedura automatycznie wyliczy wartości początkowe parametrów. Alternatywnie, użytkownik może sam określić wartości początkowe dla oszacowania parametru.

Iteracje i kryteria zbieżności określone na tej karcie mają zastosowanie tylko do początkowego uogólnionego modelu liniowego. Opis kryteriów estymacji używanych przy dopasowywaniu ogólnych równań estymujących zawiera omówienie karty Powtórzone.

**Iteracje.** Dostępne są następujące opcje:

- **Maksymalna liczba iteracji.** Maksymalna liczba iteracji, jakie wykona algorytm. Podaj nieujemną liczbę całkowitą.
- **Maksimum kroków połowienia.** Przy każdej iteracji rozmiar kroku jest zmniejszany o 0,5, aż do momentu wzrostu logarytmu wiarygodności lub osiągnięcia maksymalnej wartości połowienia kroków. Podaj dodatnią liczbę całkowitą.
- **Sprawdź separację punktów danych.** Po wybraniu tej opcji algorytm przeprowadza testy, aby sprawdzić, czy oszacowania parametru mają unikalne wartości. Rozdzielenie następuje, kiedy procedura może stworzyć model, który poprawnie klasyfikuje wszystkie obserwacje. Ta opcja jest dostępna dla zmiennych wielomianowych i dwumianowych w układzie dychotomicznym.

**Regresja logistyczna: Kryteria zbieżności.** Dostępne są następujące opcje

- **Zbieżność parametru.** Po zaznaczeniu tej opcji algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana oszacowania parametru jest mniejsza od określonej wartości, która musi być dodatnia.
- **Zbieżność logarytmu wiarygodności.** Po zaznaczeniu tej opcji algorytm zatrzymuje się, gdy bezwzględna lub względna zmiana funkcji logarytmu wiarygodności jest mniejsza od określonej wartości, która musi być dodatnia.
- **Zbieżność Hessego.** Dla specyfikacji bezwzględnej zakłada się zbieżność, jeśli statystyka w oparciu o zbieżność Hessego jest mniejsza od określonej wartości dodatniej. Dla specyfikacji względnej zakłada się zbieżność, jeśli statystyka jest mniejsza niż iloczyn określonej wartości dodatniej i wartości bezwzględnej logarytmu wiarygodności.

**Tolerancja osobliwości.** Macierze osobliwe (lub nieodwracalne) zawierają liniowo zależne kolumny, które mogą powodować problemy z algorytmem estymacji. Nawet macierze prawie osobliwe mogą powodować uzyskiwanie słabych wyników, dlatego procedura będzie traktować macierze, których wyznacznik jest mniejszy zakres tolerancji, jako osobliwe. Określ wartość dodatnią.

## Uogólnione równania estymujące-wartości początkowe

Procedura szacuje początkowy uogólniony model liniowy, a oszacowania z tego modelu są używane jako wartości początkowe dla oszacowań parametrów w części modelu liniowego uogólnionych równań estymujących. Wartości początkowe nie są wymagane dla roboczej macierzy korelacji, ponieważ składniki macierzy są oparte na oszacowaniach parametrów. Wartości początkowe określone w tym oknie dialogowym są używane jako punkt początkowy dla początkowego uogólnionego modelu liniowego, a nie dla uogólnionych równań estymujących, chyba że maksymalna liczba iteracji na karcie [Estymacja](#) jest ustawiona na 0.

Jeśli podane są wartości początkowe, muszą zostać podane dla wszystkich parametrów (w tym parametrów nadmiarowych) w modelu. W zbiorze danych kolejność zmiennych od lewej do prawej musi być następująca: *RowType\_*, *VarName\_*, *P1*, *P2*,... , gdzie *RowType\_* i *VarName\_* to zmienne łańcuchowe, a *P1*, *P2*,... są zmiennymi liczbowymi, które odpowiadają uporządkowanej liście parametrów.

- Wartości początkowe są podane w rekordzie z wartością *EST* dla zmiennej *RowType\_*; rzeczywiste wartości początkowe są podane w zmiennych *P1*, *P2*, .... Ta procedura ignoruje wszystkie rekordy, dla których *RowType\_* ma wartość inną niż *EST*, a także wszelkie rekordy poza pierwszym wystąpieniem *RowType\_* równym *EST*.
- Pierwszymi wartościami początkowymi muszą być wyraz wolny, jeśli jest uwzględniony w modelu, albo wartości progowe, jeśli odpowiedź ma rozkład wielomianowy.
- Wartością początkową, którą należy określić jako ostatnią jest parametr skali, a jeśli odpowiedź ma rozkład ujemny dwumianowy — również ujemny parametr dwumianu.
- Jeśli obowiązuje opcja dzielenia pliku, zmienne muszą zaczynać się od zmiennej *split-file* albo od zmiennych w kolejności podanej podczas dzielenia pliku, po czym powinny następować zmienne *RowType\_*, *VarName\_*, *P1*, *P2*, ... jak powyżej. Podziały muszą następować w podanym zbiorze wartości w tej samej kolejności, co w oryginalnym zbiorze wartości. .

*Uwaga:* Nazwy zmiennych *P1*, *P2*,... nie są wymagane; procedura zaakceptuje dowolne poprawne nazwy zmiennych dla parametrów, ponieważ mapowanie zmiennych na parametry jest oparte na pozycji zmiennej, a nie na jej nazwie. Wszelkie zmienne poza ostatnim parametrem są ignorowane.

Struktura pliku dla wartości początkowych jest taka sama, jak używana podczas eksportowania modelu w postaci danych; dzięki temu wartości końcowe z jednego przebiegu procedury można wykorzystać jako początkowe w kolejnym przebiegu.

## Ogólne równania estymujące: Statystyki

**Efekty modelu.** Dostępne są następujące opcje:

- **Typ analizy.** Określ typ analizy, która ma być używana do wygenerowania efektów modelu testującego. Analiza typu I jest ogólnie odpowiednia w przypadku znanych przyczyn a priori dot. sortowania predyktorów w modelu; typ III ma szersze zastosowanie. Statystyki Walda lub oceny uogólnionej są obliczane na podstawie wyboru dokonanego w grupie statystyk chi-kwadrat.
- **Przedziały ufności.** Określa poziom ufności większy niż 50 i mniejszy niż 100. Przedziały Walda są zawsze generowane, niezależnie od typu wybranej statystyki chi-kwadrat, i są oparte na założeniu, że parametry mają rozkład asymptotycznie normalny.
- **Funkcja log quasi-wiarygodności.** Wpływa na format wyświetlania funkcji logarytmu quasi-wiarygodności. Pełna funkcja obejmuje dodatkowy składnik, który jest stały w odniesieniu do oszacowań parametrów; nie ma wpływu na oszacowanie parametrów i w niektórych produktach oprogramowania nie jest wyświetlany.

**Drukuj.** Dostępne są następujące wyniki.

- **Podsumowanie przetwarzania przypadku.** Wyświetla liczbę i wartość procentową obserwacji uwzględnionych w analizie i z niej wykluczonych oraz tabelę Podsumowanie danych skorelowanych.
- **Statystyki opisowe.** Wyświetla statystyki opisowe i informacje podsumowujące dotyczące zmiennej zależnej, współzmiennych i czynników.
- **Informacje o modelu.** Wyświetla nazwę zbioru danych, zmienną zależną lub zdarzenia i zmienne prób, zmienną przesunięcia, ilościową zmienną ważącą, rozkład prawdopodobieństwa i funkcję łączenia.
- **Statystyki dobroci dopasowania.** Przedstawia dwa rozszerzenia kryterium informacyjnego Akaike do wyboru modelu: quasi-wiarygodność podlegająca kryterium niezależności (QIC) do wyboru najlepszej struktury korelacji oraz inną miarę QIC do wyboru najlepszego podzbioru predyktorów.
- **Statystyki podsumowujące modelu.** Wyświetla testy dopasowania modelu, z uwzględnieniem statystyki ilorazu wiarygodności dla testu typu omnibus dopasowania modelu oraz statystyki dla kontrastu typu I lub III dla każdego efektu.
- **Oceny parametrów.** Wyświetlane są oszacowania parametrów i odpowiednie statystyki testu oraz przedziały ufności. Opcjonalnie oprócz surowych oszacowań parametrów można wyświetlić wykładnicze oszacowania parametrów.
- **Macierz kowariancji szacowanych parametrów** Wyświetla macierz kowariancji szacowanych parametrów.
- **Macierz korelacji szacowanych parametrów.** Wyświetla macierz korelacji oszacowanych parametrów.
- **Macierz współczynników kontrastów (L).** Wyświetla współczynniki kontrastu dla domyślnych efektów i dla szacowanych średnich brzegowych, o ile zażądano ich na karcie Średnie EM.
- **Ogólne funkcje estymowalne.** Wyświetla macierze do generowania macierzy współczynnika kontrastu (L).
- **Przebieg iteracji.** Wyświetla przebieg iteracji dla oszacowań parametrów i logarytmu wiarygodności oraz drukuje ostatnią ewaluację wektora gradientu i macierzy Hessego. Tabela przebiegu iteracji wyświetla oszacowania parametrów dla co  $n$ -tej iteracji począwszy od iteracji 0 (oszacowanie początkowe), gdzie  $n$  jest wartością przedziału drukowania. Jeśli tworzone jest żądanie przebiegu iteracji, wówczas ostatnia iteracja jest zawsze wyświetlana, niezależnie od wartości  $n$ .
- **Robocza macierz korelacji.** Przedstawia wartości macierzy reprezentujące zależności wewnątrzobiekto. Jej struktura zależy od specyfikacji określonej na karcie Powtórzone.

## Ogólne równania estymujące: Średnie EM

Ta karta umożliwia wyświetlenie szacowanych średnich brzegowych dla poziomów współczynników i interakcji czynników. Można również zażądać, aby wyświetlona została oszacowana średnia ogólna. Szacowane średnie brzegowe są niedostępne w przypadku porządkowych modeli wielomianowych.

**Czynniki i interakcje.** Ta lista zawiera czynniki określone na karcie Predyktory oraz interakcje czynników określone na karcie Model. Współzmiennie są wykluczone z tej listy. Składniki można wybierać bezpośrednio z tej listy albo włączać do składnika interakcji za pomocą przycisku **Przez \***.

**Pokaż średnie dla.** Szacowane średnie są obliczane dla wybranych czynników i interakcji czynników. Kontrast określa konfigurację testów hipotezy w celu porównania oszacowanych średnich. Kontrast prosty wymaga kategorii odniesienia lub poziomu czynnika, z którym porównywane będą inne.

- **Parami.** Porównania parami są obliczane dla kombinacji wszystkich poziomów czynników podanych lub domniemanych. Jest to jedyny kontrast dostępny dla interakcji czynników.
- *Proste.* 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.
- **Odchylenie.** Każdy poziom czynnika jest porównywany ze średnią ogólną. Kontrasty odchylenia nie są ortogonalne.
- *Różnica.* Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz 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 następujących.
- *Powtórzone.* Wybranie tego typu kontrastu powoduje porównanie średniej każdego poziomu (prócz ostatniego) ze średnią poziomów następujących.
- *Wielomianowe.* Wybranie tego typu kontrastu powoduje porównanie efektu liniowego, efektu kwadratowego, sześciennego 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.

**Powięszczenie.** Szacowane średnie brzegowe mogą być obliczane dla odpowiedzi na podstawie oryginalnej skali zmiennej zależnej albo dla predyktora liniowego na podstawie zmiennej zależnej przekształconej przez funkcję łączenia.

**Korekta dla porównań wielokrotnych.** Podczas testowania hipotez z użyciem kontrastów wielokrotnych całościowy poziom istotności można skorygować na podstawie poziomów istotności dla uwzględnionych kontrastów. W tej grupie możliwy jest wybór metody korekty.

- **Najmniejsza istotna różnica.** Ta metoda nie kontroluje ogólnego prawdopodobieństwa odrzucenia hipotez, które stwierdzają, że niektóre kontrasty liniowe różnią się od wartości hipotezy zerowej.
- *Bonferroni.* Metoda ta dostosowuje obserwowany poziom istotności do faktu przeprowadzania wielu porównań.
- *Sekwencyjna Bonferroniego.* Jest to sekwencyjne zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.
- *Sidak.* Metoda ta zapewnia węższe granice niż test Bonferroniego.
- *Sekwencyjna Sidaka.* Jest to sekwencyjnie zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.

## Ogólne równania estymujące: Zapisz

Zaznaczone elementy są zapisywane pod określoną nazwą; można nakazać nadpisywanie istniejących zmiennych o tej samej nazwie nowymi zmiennymi albo unikanie kolizji nazw poprzez dołączanie przyrostków, tak aby nazwy nowych zmiennych były unikalne. .

- **Przewidywana wartość średniej dla zależnej.** Zapisuje przewidywane przez model wartości poszczególnych obserwacji w pierwotnej metryce odpowiedzi. Gdy rozkład odpowiedzi jest dwumianowy, a zmienna zależna jest binarna, procedura zapisuje przewidywane prawdopodobieństwa. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **łącznie przewidywane prawdopodobieństwo**, a procedura zapisuje łączne przewidywane prawdopodobieństwo każdej kategorii odpowiedzi, z wyjątkiem ostatniej.
- **Dolna granica przedziału ufności średniej dla zależnej.** Zapisuje dolną granicę przedziału ufności średniej odpowiedzi. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **Dolna granica przedziału ufności dla skumulowanego przewidywanego prawdopodobieństwa**, a procedura zapisuje dolną granicę dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.
- **Górna granica przedziału ufności średniej dla zależnej.** Zapisuje górną granicę przedziału ufności średniej odpowiedzi. Gdy rozkład odpowiedzi jest wielomianowy, tekst etykiety zmienia się na **Górna granica przedziału ufności dla skumulowanego przewidywanego prawdopodobieństwa**, a procedura zapisuje górną granicę dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.
- **Przewidywana kategoria.** W przypadku modeli z rozkładem dwumianowym i binarnej zmiennej zależnej lub rozkładu wielomianowego zapisywane są przewidywane kategorie odpowiedzi dla każdej obserwacji. Opcja ta jest niedostępna w przypadku innych rozkładów odpowiedzi.
- **Przewidywana wartość predyktora liniowego.** Zapisuje przewidywane przez model wartości poszczególnych obserwacji w metryce predyktora liniowego (odpowiedzi przekształconej przez określoną funkcję łączenia). Gdy rozkład odpowiedzi jest wielomianowy, procedura zapisuje wartość przewidywaną dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.
- **Oszacowany błąd standardowy wartości przewidywanej predyktora liniowego.** Gdy rozkład odpowiedzi jest wielomianowy, procedura zapisuje błąd standardowy dla każdej kategorii odpowiedzi, z wyjątkiem ostatniej.

Następujące elementy są niedostępne, gdy odpowiedzi mają rozkład wielomianowy.

- *Reszta surowa.* Różnica pomiędzy wartością empiryczną, a wartością przewidywaną przez model.
- **Reszta Pearsona.** Pierwiastek kwadratowy udziału obserwacji w statystyce chi-kwadrat Pearsona, ze znakiem reszty surowej.

## Ogólne równania estymujące: Eksportuj

**Eksportuj model jako dane.** Zapisuje zbiór danych w formacie IBM SPSS Statistics zawierającym korelację parametru lub macierz kowariancji z oszacowaniami parametrów, błędami standardowymi, wartościami istotności oraz stopniami swobody. Kolejność zmiennych w pliku macierzowym jest następująca.

- **Zmienne dzielące.** Jeśli są używane, wszelkie zmienne definiujące zostają podzielone.
- **RowType\_.** Przyjmuje wartości (i etykiety wartości) *COV* (kowariancje), *CORR* (korelacje), *EST* (oszacowania parametrów), *SE* (błędy standardowe), *SIG* (poziomy istotności) i *DF* (stopnie swobody planu losowania). Dla każdego parametru modelu istnieje osobna obserwacja z typem wiersza *COV* (lub *CORR*), a także osobna obserwacja dla każdego z pozostałych typów wierszy.
- **VarName\_.** Przyjmuje wartości *P1*, *P2*, ..., odpowiadające uporządkowanej liście wszystkich oszacowanych parametrów modelu (z wyjątkiem parametrów skali lub ujemnych parametrów dwumianu) — dla wierszy typów *COV* lub *CORR*, z wartościami etykiet odpowiadającymi ciągłom znaków parametrów, które są pokazane w tabeli Oszacowania parametru. Te komórki są puste w przypadku wierszy innych typów.
- **P1, P2, ...** Te zmienne odpowiadają uporządkowanej liście wszystkich parametrów modelu (z wyjątkiem, odpowiednio, parametrów skali lub ujemnych parametrów dwumianu) z wartościami etykiet



odpowiadającymi ciągami znaków parametrów, które są pokazane w tabeli Oszacowania parametru, a ponadto przyjmują wartości zgodnie z typem wiersza.

W odniesieniu do parametrów nadmiarowych wszystkie kowariancje są ustawione na zero, korelacje są ustawione na systemowy brak danych; wszystkie oszacowania parametrów są ustawione na zero, a wszystkie błędy standardowe, poziomy istotności i stopnie swobody reszt są ustawione na systemowy brak danych.

W przypadku parametru skali kowariancje, korelacje, poziom istotności oraz poziomy swobody są ustawione na systemowy brak danych. Jeśli estymacja parametru skali zostanie przeprowadzona metodą największej wiarygodności, błąd standardowy będzie podany; w przeciwnym wypadku zostanie ustawiony na systemowy brak danych.

W przypadku ujemnego parametru dwumianu kowariancje, korelacje, poziom istotności oraz stopnie swobody są ustawione na systemowy brak danych. Jeśli estymacja ujemnego parametru dwumianu zostanie przeprowadzona metodą największej wiarygodności, błąd standardowy będzie podany; w przeciwnym wypadku zostanie ustawiony na systemowy brak danych.

Jeśli występują podziały, lista parametrów musi zostać skumulowana po wszystkich podziałach. W konkretnym podziale niektóre parametry mogą być nieistotne; nie oznacza to jednak, że są nadmiarowe. W przypadku parametrów nieistotnych wszystkie kowariancje i korelacje, oszacowania parametrów, błędy standardowe, poziomy istotności oraz poziomy swobody są ustawione na systemowy brak danych.

Można użyć tego pliku macierzowego jako wartości początkowych do dalszej estymacji modelu; tego pliku nie można użyć bezpośrednio do dalszych analiz w innych procedurach, które odczytują plik macierzowy, chyba że te procedury akceptują wszystkie typy wierszy eksportowane w tym przypadku. Nawet wówczas należy zadbać o to, aby wszystkie parametry w tym pliku macierzowym miały to samo znaczenie dla procedury odczytującej plik.

**Eksportuj model jako XML.** Jeśli ta opcja jest wybrana, oszacowania parametrów oraz macierz kowariancji parametrów są zapisywane w formacie XML (PMML). Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. .

## Dodatkowe właściwości komendy GENLIN

Język składni komend umożliwia również:

- Określanie wartości początkowych dla oszacowań parametru w postaci listy liczb (za pomocą opcji komendy CRITERIA).
- Określanie stałej działającej macierzy korelacji (za pomocą opcji komendy REPEATED).
- Ustalanie współzmiennych na wartościach innych niż ich średnie podczas obliczania szacowanych średnich brzegowych (za pomocą opcji komendy EMMEANS).
- Określanie niestandardowych kontrastów wielomianowych dla szacowanych średnich brzegowych (za pomocą opcji komendy EMMEANS).
- Określanie podzbioru czynników, dla których szacowane średnie brzegowe są wyświetlane w celu porównywania z użyciem podanego typu kontrastu (za pomocą słów kluczowych TABLES i COMPARE opcji komendy EMMEANS).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Uogólnione liniowe modele mieszane

---

Uogólnione liniowe modele mieszane rozszerzają model liniowy w następujący sposób:

- Zmienna przewidywana jest związana liniowo z czynnikami i współzmiennymi za pomocą określonej funkcji łączenia.
- Zmienna przewidywana może mieć rozkład inny niż normalny.
- Obserwacje mogą być skorelowane.

Uogólnione liniowe modele mieszane obejmują szeroki wachlarz modeli, począwszy od prostych modeli regresji liniowej, aż po złożone wielopoziomowe modele dla danych z obserwacji długofalowych nieposiadających rozkładu normalnego.

### Przykłady

Rada szkoły dzielnicowej może używać uogólnionego liniowego modelu mieszanego w celu ustalenia, czy eksperymentalna metoda nauczania skutecznie poprawia wyniki z matematyki. Uczniowie z tej samej klasy powinni zostać skorelowani, ponieważ są nauczani przez tego samego nauczyciela, a klasy w tej samej szkole również mogą być skorelowane, co umożliwi nam uwzględnienie efektów losowych na poziomie szkoły i klasy w celu uwzględnienia różnych źródeł zmienności.

Badacze pracujący w dziedzinach medycznych mogą używać uogólnionego liniowego modelu mieszanego w celu ustalenia, czy nowy lek przeciwdrgawkowy może zmniejszać częstotliwość ataków padaczkowych u pacjentów. Powtarzalne pomiary dla tego samego pacjenta są zwykle skorelowane pozytywnie, dlatego odpowiedni powinien być model mieszany z pewnymi efektami losowymi. Zmienna przewidywana — liczba ataków — przyjmuje dodatnie wartości całkowite, dlatego odpowiedni może być uogólniony liniowy model mieszany o rozkładzie Poissona i logarytmiczna funkcja łączenia.

Dostawcy telewizji kablowej, telefonii przewodowej i internetu przewodowego mogą używać uogólnionego liniowego modelu mieszanego w celu uzyskania dodatkowych informacji na temat potencjalnych klientów. Możliwe odpowiedzi charakteryzują się nominalnymi poziomami pomiaru, dlatego analityk firmy stosuje uogólniony liniowy mieszany model logit z losowym wyrazem wolnym w celu przechwytywania korelacji między odpowiedziami na pytania o korzystanie z usług różnych typów (telewizja, telefon, internet) w odpowiedziach konkretnego respondenta uczestniczącego w ankiecie.

Karta struktury danych umożliwia określenie zależności strukturalnych między rekordami w bazie danych po skorelowaniu obserwacji. Jeśli rekordy w zestawie danych reprezentują niezależne obserwacje, nie ma potrzeby określać niczego na tej karcie.

## Opcje efektów

### Obiekty

Połączenie wartości wybranych zmiennych jakościowych powinno w sposób jednoznaczny definiować obiekty w zbiorze danych. Na przykład jedna zmienna *ID pacjenta* powinna wystarczyć do zdefiniowania obiektów w jednym szpitalu. Jeśli jednak numery identyfikacyjne pacjenta nie identyfikują jednoznacznie pacjentów w różnych szpitalach, wówczas konieczna może być kombinacja *ID szpitala* i *ID pacjenta*. Przy wielokrotnych pomiarach dla każdego obiektu zapisywane są wielokrotne obserwacje. Z tego powodu jeden obiekt może być w zbiorze danych przedstawiany w wielu rekordach.

**Obiekt** jest jednostką obserwacyjną, która może być rozważana niezależnie od innych obiektów. Na przykład odczyty ciśnienia krwi pacjenta w badaniu medycznym mogą być traktowane jako niezależne od odczytów innych pacjentów. Definiowanie obiektów staje się szczególnie istotne, gdy istnieją wielokrotne pomiary na każdy obiekt i gdy wymagane jest modelowanie korelacji między tymi obserwacjami. Na przykład można oczekiwać, że pomiary ciśnienia krwi u jednego pacjenta podczas kolejnych wizyt u lekarza będą skorelowane.

Wszystkie zmienne określone jako **obiekty** w oknie dialogowym zmiennych są używane do definiowania obiektów dla struktury kowariancji rezydualnej i udostępniają listę możliwych zmiennych do definiowania obiektów dla struktur kowariancji efektów losowych w [bloku efektów losowych](#).

### OML Powtarzane pomiary

Zmienne określone w tym miejscu służą do identyfikowania obserwacji powtórzonych. Na przykład pojedyncza zmienna *Tydzień* może identyfikować 10 tygodni obserwacji w badaniu medycznym, a w celu identyfikacji obserwacji codziennych w ciągu roku można używać razem zmiennych *Miesiąc* i *Dzień*.

## Opcje kowariancji

### Definiuj grupy kowariancji według

Zmienne jakościowe określone w tym miejscu definiują niezależne zestawy parametrów kowariancji efektów powtarzanych; po jednej dla każdej kategorii zdefiniowanej przez klasyfikację krzyżową zmiennych grupujących. Wszystkie obiekty mają ten sam typ kowariancji; obiekty w tej samej grupie kowariancji będą miały te same wartości dla parametrów.

### Typ struktury kowariancji

Określa strukturę kowariancji dla reszt. Dostępne są różne opcje kowariancji w zależności od ustawienia wybranego w polu **Typ kowariancji powtórzonej**. Dostępne są następujące struktury:

- Autoregresja pierwszego rzędu (AR1)
- Iloczyn prosty AR1(UN\_AR1)
- Iloczyn prosty nieustrukturyzowany (UN\_UN)
- Iloczyn prosty - symetria złożona (UN\_CS)
- Heterogeniczna symetria złożona (CSH)
- Heterogeniczna autoregresyjna (ARH1)
- Autoregresyjna średnia ruchoma (1,1) (ARMA11)
- Symetria złożona
- Przekątna
- Tożsamość skalowana
- Toeplitz
- Nieustrukturalizowana
- Składowe wariancji
- Przestrzenna: potęgowa
- Przestrzenna: wykładnicza
- Przestrzenna: Gaussa
- Przestrzenna: liniowa
- Przestrzenna: liniowo-logarytmiczna
- Przestrzenna: sferyczna

### Miary Kroneckera

Wybierz zmienne określające strukturę obiektów dla pomiarów kowariancji Kroneckera i określ, jak skorelowane są błędy pomiaru. Pole jest dostępne tylko wtedy, gdy jako **Typ kowariancji powtórzonej** wybrany jest jeden z następujących typów:

- Iloczyn prosty AR1(UN\_AR1)
- Iloczyn prosty nieustrukturyzowany (UN\_UN)
- Iloczyn prosty - symetria złożona (UN\_CS)

### Współrzędne kowariancji przestrzennej

Gdy jeden z typów kowariancji przestrzennej wybrano jako typ struktury kowariancji, zmienne na tej liście określają współrzędne powtarzanych obserwacji.

Więcej informacji można znaleźć w temacie [“Struktury kowariancji” na stronie 103](#).

### Pseudo R<sup>2</sup> miary

Miary pseudo-R<sup>2</sup> i współczynnik korelacji wewnątrzklasowej są uwzględniane w danych wyjściowych GLMM (w razie potrzeby). Miary pseudo-R<sup>2</sup> oparte są w całości na oszacowaniach końcowych i są generowane po zakończeniu oszacowania. Współczynnik determinacji R<sup>2</sup> jest często zgłaszaną statystyką, ponieważ reprezentuje proporcję wariacji wyjaśnioną przez model liniowy. Wewnątrzklasowy współczynnik korelacji (ICC) jest statystyką pokrewną, która określa proporcję wariacji wyjaśnioną przez czynnik grupowania (losowy) w danych wielopoziomowych/hierarchicznych.

## Otrzymywanie uogólnionego liniowego modelu mieszanego

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Z menu wybierz:

### Analiza > Modele mieszane > Uogólnione liniowe...

1. Zdefiniuj strukturę obiektu zbioru danych na karcie **Struktura danych**. Karta udostępnia opcje do określania efektów modelu i typu kowariancji powtórzonej.
2. Na karcie **Zmienne i efekty** musi istnieć pojedyncza zmienna przewidywana, która może mieć dowolny poziom miary, albo specyfikacja zdarzeń/prób, ale w tym przypadku specyfikacja zdarzeń i prób musi być ciągła. Opcjonalnie określ jej rozkład i funkcję łączenia, efekty stałe i dowolne bloki oraz przesunięcie efektów losowych albo wagi analizy.
3. Kliknij **Opcje budowania**, aby podać opcjonalne ustawienia budowania.
4. Kliknij **Opcje modelu**, aby zapisać wyniki w aktywnym zbiorze danych i wyeksportować ten model do pliku zewnętrznego.
5. Kliknij opcję **Uruchom**, aby uruchomić procedurę i utworzyć Obiekty modelu.

## Zmienna przewidywana

Te ustawienia definiują zmienną przewidywaną, jej rozkład, a także jej relację z predyktorami przez funkcję łączenia.

**Zmienna przewidywana.** Zmienna przewidywana jest wymagana. Może mieć dowolny poziom pomiaru, a poziom pomiaru zmiennej przewidywanej ogranicza to, które rozkłady i funkcje łączenia są odpowiednie.

- **Użyj liczby prób jako mianownika.** Gdy zmienna przewidywana odpowiedzi jest liczbą zdarzeń występujących w zbiorze prób, zmienna przewidywana zawiera liczbę zdarzeń. Można wybrać dodatkową zmienną zawierającą liczbę prób. Na przykład podczas testowania nowego pestycydu można wystawiać mrówki na działanie pestycydu w różnych stężeniach, a następnie rejestrować liczbę mrówek zabitych i liczbę mrówek w każdej próbie. W tym przypadku zmienna rejestrująca liczbę mrówek zabitych powinna być określona jako zmienna przewidywana (zdarzenia), a zmienna rejestrująca liczbę mrówek w każdej próbie powinna być określona jako zmienna prób. Jeśli liczba mrówek jest taka sama dla każdej próby, wówczas liczba prób może być określona przy użyciu wartości stałej.

Liczba prób powinna być większa niż lub równa liczbie zdarzeń w każdym rekordzie. Zdarzenia powinny być nieujemnymi liczbami całkowitymi, a próby powinny być dodatnimi liczbami całkowitymi.

- **Dostosuj kategorię odniesienia.** W przypadku przewidywanej zmiennej jakościowej można wybrać kategorię odniesienia. To może wpłynąć na niektóre wyniki, takie jak oszacowania parametrów, ale nie powinno zmienić dopasowania modelu. Na przykład, jeśli zmienna przewidywana przyjmuje domyślnie wartości 0, 1 i 2, wówczas procedura ustawia ostatnią kategorię (o najwyższej wartości) — czyli 2 — jako kategorię odniesienia. W tej sytuacji oszacowania parametrów powinny być interpretowane jako odnoszące się do wiarygodności kategorii 0 lub 1 w odniesieniu do wiarygodności kategorii 2. Jeśli zostanie określona kategoria niestandardowa, a zmienna przewidywana zawiera zdefiniowane etykiety, można ustawić kategorię odniesienia, wybierając wartość z listy. To może być wygodne, jeśli podczas określania modelu użytkownik nie pamięta dokładnie sposobu zakodowania konkretnej zmiennej.

**Rozkład zmiennej przewidywanej i związek (łącze) z modelem liniowym.** Na podstawie wartości predyktorów model oczekuje, że rozkład wartości zmiennej przewidywanej będzie zgodny z określonym kształtem, a w przypadku wartości zmiennej przewidywanej, że będą powiązane liniowo z predyktorami przez określoną funkcję łączenia. Dostępne są skróty dla kilku typowych modeli. Jeśli istnieje konkretna kombinacja rozkładu i funkcji łączenia, dla której użytkownik planuje znaleźć dopasowanie, a która nie jest dostępna na liście skrótów, wówczas można również wybrać ustawienie **Użytkownika**.

- **Model liniowy.** Określa rozkład normalny z łączem tożsamości, które jest użyteczne, gdy zmienna przewidywana może zostać przewidziana z użyciem modelu regresji liniowej lub modelu ANOVA.
- **Regresja gamma.** Określa rozkład gamma z logarytmiczną funkcją łączenia, który powinien być używany, gdy zmienna przewidywana zawiera wszystkie wartości dodatnie i wykazuje skośność w stronę wyższych wartości.

- **Analiza logliniowa (LOGLINEAR).** Określa rozkład Poissona z logarytmiczną funkcją łączenia, który powinien być używany, gdy zmienna przewidywana reprezentuje liczbę wystąpień w stałym okresie czasu.
- **Regresja ujemna dwumianowa.** Określa rozkład ujemny dwumianowy z logarytmiczną funkcją łączenia, który powinien być używany, gdy zmienna przewidywana i mianownik reprezentują liczbę prób wymaganych do zaobserwowania  $k$  sukcesów.
- **Wielomianowa regresja logistyczna.** Określa rozkład wielomianowy, który powinien być używany, gdy zmienna przewidywana jest odpowiedzią wielokategoryjną. Używa skumulowanej funkcji łączenia logit (wyniki porządkowe) lub uogólnionej funkcji łączenia logit (wielokategoryjne odpowiedzi nominalne).
- **Binarna regresja logistyczna.** Określa rozkład dwumianowy z funkcją łączenia logit, który powinien być używany, gdy zmienna przewidywana jest odpowiedzią binarną przewidywaną przez model regresji logistycznej.
- **Binarny probit.** Określa rozkład dwumianowy z funkcją łączenia probit, który powinien być używany, gdy zmienna przewidywana jest odpowiedzią binarną z bazowym rozkładem binarnym.
- **Przeżycia obcięte przedziałowe.** Określa rozkład dwumianowy z funkcją łączenia komplementarny log-log, która jest użyteczna w analizie przeżycia, gdy niektóre obserwacje nie mają zdarzenia kończącego.

#### Dystrybucja

Ta opcja określa rozkład zmiennej przewidywanej. Możliwość określenia rozkładu innego niż normalny i funkcji łączenia innej niż tożsamość stanowi kluczowe udoskonalenie uogólnionego liniowego modelu mieszanego w porównaniu do liniowego modelu mieszanego. Istnieje wiele możliwych kombinacji rozkład-funkcja łączenia, a kilka z nich może być odpowiednich dla dowolnego zbioru danych, dlatego wybór może być zależny od rozważań teoretycznych a priori lub tego, która kombinacja wydaje się zapewniać najlepsze dopasowanie.

- **Dwumianowy.** Ten rozkład jest odpowiedni tylko dla zmiennej przewidywanej, która reprezentuje odpowiedź binarną lub liczbę zdarzeń.
- **Gamma.** Ten rozkład jest odpowiedni dla zmiennych przewidywanych z wartościami w skali dodatniej, które wykazują skośność w stronę większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Odwrócony Gaussa.** Ten rozkład jest odpowiedni dla zmiennych przewidywanych z wartościami w skali dodatniej, które wykazują skośność w stronę większych wartości dodatnich. Jeśli wartość danych jest mniejsza niż lub równa 0 lub występuje brak wartości, wówczas dana obserwacja nie jest wykorzystywana w analizie.
- **Wielomianowy.** Ten rozkład jest odpowiedni dla zmiennej przewidywanej, która reprezentuje odpowiedź wielokategoryjną. Forma modelu będzie zależna od poziomu pomiaru zmiennej przewidywanej.

Wynikiem dla **nominalnej** zmiennej przewidywanej będzie nominalny model wielomianowy, w którym szacowany jest osobny zestaw parametrów modelu dla każdej kategorii zmiennej przewidywanej (z wyjątkiem kategorii odniesienia). Oszacowania parametrów dla konkretnego predyktora przedstawiają związek między predyktorem a wiarygodnością dla każdej kategorii zmiennej przewidywanej, względem kategorii odniesienia.

Wynikiem dla **porządkowej** zmiennej przewidywanej będzie porządkowy model wielomianowy, w którym tradycyjny składnik stałej jest zastępowany przez zestaw parametrów **progowych**, które odnoszą się do prawdopodobieństwo skumulowanego kategorii zmiennej przewidywanej.

- **Ujemny dwumianowy.** Ujemna regresja dwumianowa wykorzystuje ujemny rozkład dwumianowy z logarytmiczną funkcją łączenia, które powinny być używane, gdy zmienna zależna reprezentuje liczebność wystąpień z dużą wariancją.
- **Normalny.** Jest odpowiedni w przypadku ilościowej zmiennej przewidywanej, której wartości przyjmują symetryczny rozkład w kształcie dzwona, z centralną wartością średnią.
- **Poissona.** Ten rozkład można traktować jako liczbę wystąpień zdarzenia badanego w ustalonym okresie i jest odpowiedni dla zmiennych o nieujemnych wartościach całkowitych. Jeśli wartość danych nie jest

liczbą całkowitą, jest mniejsza od 0 lub występuje brak danych, wówczas dana obserwacja nie jest wykorzystywana w analizie.

## Funkcje łączenia

Funkcja łączenia to transformacja zmiennej przewidywanej, która umożliwi estymację modelu. Dostępne są następujące funkcje:

- **Tożsamość.**  $f(x)=x$ . Zmienna przewidywana nie jest transformowana. Ta funkcja łączenia może być używana z dowolnym rozkładem, z wyjątkiem rozkładu wielomianowego.
- **Komplementarny logarymiczno-logarymiczny.**  $f(x)=\log(-\log(1-x))$ . Jest odpowiednia tylko w przypadku rozkładu dwumianowego i wielomianowego.
- **Cauchit.**  $f(x) = \tan(\pi(x - 0.5))$ . Jest odpowiednia tylko w przypadku rozkładu dwumianowego i wielomianowego.
- **Logarytm.**  $f(x)=\log(x)$ . Ta funkcja łączenia może być używana z dowolnym rozkładem, z wyjątkiem rozkładu wielomianowego.
- **Logarymiczny dopełnienia.**  $f(x)=\log(1-x)$ . Ma zastosowanie tylko w przypadku rozkładu dwumianowego.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Jest odpowiednia tylko w przypadku rozkładu dwumianowego i wielomianowego.
- **Ujemny logarymiczno-logarymiczny.**  $f(x)=-\log(-\log(x))$ . Jest odpowiednia tylko w przypadku rozkładu dwumianowego i wielomianowego.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , gdzie  $\Phi^{-1}$  jest odwrotnością funkcji skumulowanego rozkładu standardowego normalnego. Jest odpowiednia tylko w przypadku rozkładu dwumianowego i wielomianowego.
- **Potęgowy.**  $f(x)=x^\alpha$ , jeśli  $\alpha \neq 0$ .  $f(x)=\log(x)$ , jeśli  $\alpha=0$ .  $\alpha$  jest wymaganą specyfikacją liczbową i musi być liczbą rzeczywistą. Ta funkcja łączenia może być używana z dowolnym rozkładem, z wyjątkiem rozkładu wielomianowego.

## Efekty stałe

Współczynniki efektów stałych są zwykle traktowane jako zmienne, których wszystkie wartości badane są reprezentowane w zbiorze danych, i mogą być używane podczas oceniania. Domyślnie zmienne z predefiniowaną rolą w danych wejściowych, które nie są określone w innym miejscu w oknie dialogowym, są wprowadzane w części modelu dotyczącej efektów stałych. Zmienne jakościowe (nominalne i porządkowe) są używane jako współczynniki w modelu, a zmienne ilościowe są używane jako współzmiennie.

Wprowadź efekty do modelu, zaznaczając jedną zmienną lub większą liczbę zmiennych na liście źródłowej i przeciągając do listy efektów. Typ tworzonego efektu jest zależny od tego, do którego obszaru aktywnego zostanie przeciągnięte zaznaczenie.

- **Główne.** Zmienne odrzucane są wyświetlane jako osobne efekty główne u dołu listy efektów.
- **2. rzędu.** Wszystkie możliwe pary zmiennych odrzucanych pojawiają się jako interakcje 2. rzędu u dołu listy efektów.
- **3. rzędu.** Wszystkie możliwe trójki zmiennych odrzucanych pojawiają się jako interakcje 3. rzędu u dołu listy efektów.
- **\***. Połączenie wszystkich usuniętych pól jest wyświetlane jako pojedyncza interakcja na dole listy efektów.

Przyciski po prawej stronie kreatora efektów umożliwiają wykonywanie różnych działań.

Tabela 1. Opisy przycisków kreatora efektów




Ikona	Opis
	Umożliwia usuwanie składników z modelu efektów stałych poprzez wybranie składników przeznaczonych do usunięcia i kliknięcie przycisku usuwania.

Tabela 1. Opisy przycisków kreatora efektów (kontynuacja)

Ikona	Opis
	<p>Umożliwia reorganizację składników w modelu efektów statycznych poprzez wybranie składników przeznaczonych do reorganizacji i kliknięcie strzałki w górę lub w dół.</p>
	<p>Umożliwia dodawanie składników zagnieżdżonych przy użyciu okna dialogowego "Dodaj składnik zdefiniowany przez użytkownika" na stronie 57 poprzez kliknięcie przycisku Dodaj składnik zdefiniowany przez użytkownika.</p>

**Uwzględnij wyraz wolny.** Wyraz wolny jest zwykle uwzględniany w modelu. Jeśli można założyć, że dane przechodzą przez początek układu współrzędnych, wyraz wolny można wyłączyć z modelu.

## Dodaj składnik zdefiniowany przez użytkownika

W tej procedurze można zbudować składniki zagnieżdżone dla modelu. Składniki zagnieżdżone są przydatne do modelowania efektu czynników lub współzmiennych, których wartości nie wchodzi w interakcje z poziomami innego czynnika. Na przykład sieć sklepów spożywczych może analizować zwyczaje zakupowe swoich klientów w kilku sklepach. Ponieważ każdy klient bywa regularnie tylko w jednym z tych sklepów, efekt *Klient* jest **zagnieżdżony w** efekcie *Lokalizacja sklepu*.

Ponadto można uwzględnić efekty interakcji, takie jak składniki wielomianowe z tą samą współzmienną, lub dodać wiele poziomów zagnieżdżenia do składnika zagnieżdżonego.

**Ograniczenia.** W odniesieniu do składników zagnieżdżonych obowiązują następujące ograniczenia:

- Wszystkie czynniki w interakcji muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A^*A$ .
- Wszystkie czynniki w efekcie zagnieżdżonym muszą być unikalne. A zatem, jeśli  $A$  jest czynnikiem, to niedozwolone jest określenie  $A(A)$ .
- Efekt nie może być zagnieżdżony w obrębie współzmiennnej. A zatem, jeśli  $A$  jest czynnikiem, a  $X$  jest współzmienną, to określenie  $A(X)$  jest niedozwolone.

Tworzenie zagnieżdżonego składnika

1. Wybierz czynnik lub współzmienną zagnieżdżony/-ą w innym czynniku, a następnie kliknij przycisk ze strzałką.
2. Kliknij opcję **(W)**.
3. Wybierz czynnik, w którym zagnieżdżony jest poprzedni czynnik lub współzmienna, a następnie kliknij przycisk ze strzałką.
4. Kliknij opcję **Dodaj składnik**.

Opcjonalnie można uwzględnić efekty interakcji lub dodać wiele poziomów zagnieżdżenia do zagnieżdżonego składnika.

## Efekty losowe

Współczynniki efektów losowych to zmienne, których wartości w pliku danych mogą być traktowane jako próba losowa z większej populacji wartości. Są użyteczne, gdy konieczne jest wyjaśnienie nadmiernej zmienności w zmiennej przewidywanej. Domyślnie w przypadku wyboru więcej niż jednego obiektu na karcie Struktura danych dla każdego obiektu poza obiektem najbardziej wewnętrznym zostanie utworzony blok efektów losowych. Na przykład, jeśli wybrano Szkołę, Klasę i Ucznia jako obiekty a karcie Struktura danych, wówczas następujące bloki efektów losowych są tworzone automatycznie:

- Efekt losowy 1: obiektem jest szkoła (brak efektów, tylko wyraz wolny)
- Efekt losowy 2: obiektem jest szkoła \* klasa (brak efektów, tylko wyraz wolny)

Z bloków efektów losowych można korzystać na dwa sposoby:

1. Aby dodać nowy blok, kliknij przycisk **Dodaj blok...** Spowoduje to otwarcie okna dialogowego “Blok efektów losowych” na stronie 58.
2. Aby edytować istniejący blok, wybierz blok, który ma być edytowany, i kliknij opcję **Edytuj blok...** Spowoduje to otwarcie okna dialogowego “Blok efektów losowych” na stronie 58.
3. W celu usunięcia jednego lub większej liczby bloków, wybierz bloki, które chcesz usunąć, a następnie kliknij przycisk usuwania.

## Blok efektów losowych

Wprowadź efekty do modelu, zaznaczając jedną zmienną lub większą liczbę zmiennych na liście źródłowej i dodając na listę **Konstruktora efektów**.

Typ tworzonego efektu jest zależny od wyboru opcji **Typ**. Zmienne jakościowe (nominalne i porządkowe) są używane jako współczynniki w modelu, a zmienne ilościowe są używane jako współzmiennie.

### Interakcje

Kombinacja wszystkich zmiennych odrzucanych pojawia się jako pojedyncza interakcja u dołu listy efektów.

### Efekty główne

Zmienne odrzucane są wyświetlane jako osobne efekty główne u dołu listy efektów.

### Wszystkie 2. rzędu

Wszystkie możliwe pary zmiennych odrzucanych pojawiają się jako interakcje 2. rzędu u dołu listy efektów.

### Wszystkie 3. rzędu

Wszystkie możliwe trójki zmiennych odrzucanych pojawiają się jako interakcje 3. rzędu u dołu listy efektów.

### Wszystkie 4. rzędu

Wszystkie możliwe trójki zmiennych odrzucanych pojawiają się jako interakcje 4. rzędu u dołu listy efektów.

### Wszystkie 5. rzędu

Wszystkie możliwe trójki zmiennych odrzucanych pojawiają się jako interakcje 5. rzędu u dołu listy efektów.

### Uwzględnij wyraz wolny

Wyraz wolny nie jest domyślnie uwzględniony w modelu efektów losowych. Jeśli można założyć, że dane przechodzą przez początek układu współrzędnych, wyraz wolny można wyłączyć z modelu.

### Wyświetl parametry predykcji dla tego bloku

Pozwala wyświetlać szacowania parametrów według estymatora efektów losowych.

### Kombinacja obiektów

Dzięki tej opcji możliwe jest określanie obiektów efektów losowych z obecnej kombinacji obiektów z okna dialogowego zmiennych. Na przykład, jeśli *Szkoła*, *Klasa* i *Uczeń* są w tej kolejności zdefiniowane jako obiekty w oknie dialogowym zmiennych, wówczas lista rozwijana **Kombinacja obiektów** będzie zawierała opcje **Brak**, **Szkoła**, **Szkoła \* Klasa** oraz **Szkoła \* Klasa \* Uczeń**.

### Typ kowariancji efektów losowych

Określa strukturę kowariancji dla reszt. Dostępne są następujące struktury:

- Autoregresja pierwszego rzędu (AR1)
- Autoregresywna średnia ruchoma (1,1) (ARMA11)
- Symetria złożona
- Przekątna
- Skalowana tożsamość
- Toeplitz
- Nieustrukturalizowana



- Składowe wariancji

### Definiuj grupy kowariancji według

Zmienne jakościowe określone w tym miejscu definiują niezależne zestawy parametrów kowariancji efektów losowych; po jednej dla każdej kategorii zdefiniowanej przez klasyfikację krzyżową zmiennych grupujących. Dla każdego bloku efektów losowych można określić inny zestaw zmiennych grupujących. Wszystkie obiekty mają ten sam typ kowariancji; obiekty w tej samej grupie kowariancji będą miały te same wartości dla parametrów.

## Waga i przesunięcie

**Waga analizy.** Parametr skali to oszacowanie parametru modelu w odniesieniu do wariancji odpowiedzi. Wagi analiz są wartościami „znanymi”, które mogą się różnić między obserwacjami. Jeśli określona jest zmienna wagi analizy, wówczas parametr skali, który jest powiązany z wariancją odpowiedzi, jest dzielony przez wartości wagi analizy dla każdej obserwacji. W analizie nie są używane rekordy z wartościami wagi analizy mniejszymi od zera lub równymi zeru ani obserwacje brakujące.

**Przesunięcie.** Składnik przesunięcia jest predyktorem „strukturalnym”. Jego wskaźnik nie jest szacowany przez model, ale przyjmuje się, że ma wartość 1; dlatego wartości przesunięcia są po prostu dodawane do predyktora liniowego zmiennej przewidywanej. Jest to szczególnie przydatne w modelach regresji Poissona, w których każda obserwacja może mieć inny poziom ekspozycji na badane zdarzenie.

Na przykład przy modelowaniu częstości wypadków wśród poszczególnych kierowców należy pamiętać o istotnej różnicy między kierowcą, który miał jeden wypadek w ciągu trzech lat, a kierowcą, który miał jeden wypadek w ciągu 25 lat! Liczba wypadków może być modelowana jako odpowiedź o rozkładzie Poissona lub odpowiedź o rozkładzie ujemnym dwumianowym z logarytmiczną funkcją łączenia, jeśli logarytm naturalny doświadczenia kierowcy jest uwzględniony w składniku przesunięcia.

Inne kombinacje typów rozkładu i funkcji łączenia będą wymagały przekształcenia zmiennej przesunięcia.

## Ogólne opcje budowania

Te opcje określają niektóre bardziej zaawansowane kryteria używane do budowania modelu.

### Porządek sortowania

Te elementy sterujące określają porządek kategorii dla zmiennych przewidywanych i czynników (jakościowych zmiennych wejściowych) na potrzeby określenia „ostatniej” kategorii. Ustawienie porządku sortowania zmiennych przewidywanych jest ignorowane, jeśli zmienne przewidywane nie jest jakościowe lub jeśli w ustawieniach [“Zmienna przewidywana ” na stronie 54](#) określona jest niestandardowa kategoria odniesienia.

### Reguły zatrzymujące

Możliwe jest określenie maksymalnej liczby iteracji, jakie wykona algorytm. W algorytmie wykorzystywany jest proces z podwójną iteracją, który obejmuje pętlę wewnętrzną i zewnętrzną. Wartość określona dla maksymalnej liczby iteracji ma zastosowanie do obu pętli. Podaj nieujemną liczbę całkowitą. Wartość domyślna to: 100.

### Ustawienia po estymacji

Te ustawienia określają sposób obliczania niektórych wyników modelu na potrzeby wyświetlania.

### Poziom ufności (%)

Jest to poziomy ufności używany do wyliczania oszacowań przedziałów współczynników modelu. Należy podać wartość większą od 0 i mniejszą od 100. Domyślną wartością jest 95.

### Stopnie swobody

Określa sposób obliczania stopni swobody dla testów istotności. Wybierz opcję **Metoda resztowa**, jeśli próba jest wystarczająco duża lub dane są zrównoważone, albo model używa kowariancji prostszego typu — na przykład przekątna lub tożsamość skalowana. To jest ustawienie domyślne. Wybierz opcję **Przybliżenie Satterthwaite’a**, jeśli próba jest niewielka, dane są niezrównoważone lub w modelu używany jest skomplikowany typ kowariancji — na przykład nieustrukturyzowana. Wybierz opcję **Przybliżenie Kenwarda-Rogera**, jeśli próba jest mała, a model jest modelem o ograniczonej maksymalnej wiarygodności (REML).

## Testy efektów stałych i współczynników

Jest to metoda obliczania macierzy kowariancji oszacowań parametrów. Wybierz mocne oszacowanie, jeśli martwi Cię możliwość naruszenia założeń modelu.

## Oszacowanie

W algorytmie budowania modelu wykorzystywany jest proces z podwójną iteracją, który obejmuje pętlę wewnętrzną i zewnętrzną. Względem pętli wewnętrznej zastosowanie mają następujące ustawienia.

### Zbieżność parametru.

Zbieżność jest zakładana, jeśli maksymalna zmiana bezwzględna lub względna w oszacowaniach parametru jest mniejsza niż podana wartość, która musi być nieujemna. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### Zbieżność logarytmu wiarygodności.

Zbieżność jest zakładana, jeśli zmiana bezwzględna lub względna w funkcji logarytmu wiarygodności jest mniejsza niż podana wartość, która musi być nieujemna. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### Zbieżność Hessego.

W przypadku specyfikacji **Wartości bezwzględne** zakładana jest zbieżność, jeśli statystyka oparta na macierzy Hessego jest mniejsza niż określona wartość. W przypadku specyfikacji **Względne** zbieżność jest zakładana, jeśli statystyka jest mniejsza niż iloczyn wartości określonej i wartości bezwzględnej logarytmu wiarygodności. Kryterium nie jest stosowane, jeśli określona wartość jest równa 0.

### Maksymalna liczba kroków oceny Fishera.

Podaj nieujemną liczbę całkowitą. Wartość 0 określa metodę Newtona-Raphsona. Wartości większe od 0 określają użycie algorytmu oceny Fishera aż do iteracji o numerze  $n$ , gdzie  $n$  jest podaną liczbą całkowitą, a następnie metody Newtona-Raphsona.

### Tolerancja osobliwości.

Ta wartość jest stosowana jako tolerancja podczas kontroli osobliwości. Podaj wartość dodatnią.

**Uwaga:** Domyślnie używana jest zbieżność parametru, gdzie sprawdzana jest maksymalna **bezwzględna** zmiana przy tolerancji  $1E-6$ . To ustawienie może zwracać wyniki różniące się od wyników uzyskiwanych w wersjach wcześniejszych niż wersja 22. W celu odtworzenia wyników z wersji wcześniejszych niż wersja 22 należy użyć opcji **Względne** dla kryterium zbieżności parametru i zachować domyślną wartość tolerancji równą  $1E-6$ .

## Oszacowane średnie

Ta karta umożliwia wyświetlenie szacowanych średnich brzegowych dla poziomów współczynników i interakcji czynników. Szacowane średnie brzegowe są niedostępne w przypadku modeli wielomianowych.

**Składniki.** Ta lista zawiera składniki modelu w efektach stałych, które są w całości zbudowane ze zmiennych jakościowych. Należy sprawdzić każdy składnik, dla którego model ma zwracać szacowane średnie brzegowe.

- **Typ kontrastu.** Określa typ kontrastu używanego do poziomów zmiennej kontrastu. Jeśli wybrano opcję **Brak**, żaden kontrast nie jest generowany. **Parami** powoduje uzyskiwanie porównań parami dla kombinacji wszystkich poziomów podanych czynników. Jest to jedyny kontrast dostępny dla interakcji czynników. Kontrasty **Odchylenie** porównują każdy poziom czynnika do średniej głównej. Kontrasty **Proste** porównują każdy poziom czynnika, z wyjątkiem ostatniego, do ostatniego poziomu. Poziom „ostatni” jest określony przez porządek sortowania czynników określonych w opcjach budowania. Należy zwrócić uwagę na to, że żaden z tych typów kontrastów nie jest ortogonalny.
- **Zmienna kontrastu.** Określa czynnik, którego poziomy są porównywane z użyciem wybranego typu kontrastu. Jeśli jako typ kontrastu wybrano **Brak**, wówczas nie można (i nie trzeba) wybrać żadnej zmiennej kontrastu.

**Zmienne ilościowe.** Zmienne ilościowe z listy są wyodrębniane ze składników w efektach stałych, w których stosowane są zmienne ilościowe. Podczas obliczania szacowanych średnich brzegowych współzmiennie są ustalone na podanych wartościach. Wybierz średnią lub podaj wartość użytkownika.

**Wyświetlaj estymowane średnie w odniesieniu do.** Ta opcja określa, czy szacowane średnie brzegowe są obliczane na podstawie pierwotnej skali zmiennej przewidywanej, czy na podstawie transformacji funkcji łączenia. **Pierwotna skala zmiennej przewidywanej** oblicza szacowane średnie brzegowe dla zmiennej przewidywanej. Należy zwrócić uwagę na to, że jeśli zmienna przewidywana zostanie określona z użyciem opcji zdarzeń/prób, zwróci szacowane średnie brzegowe dla proporcji zdarzeń/prób, a nie dla liczby zdarzeń. **Transformacja funkcji łączenia** oblicza szacowane średnie brzegowe dla predyktora liniowego.

**Dla porównań wielokrotnych skoryguj metodą.** Podczas testowania hipotez z użyciem kontrastów wielokrotnych całościowy poziom istotności można skorygować na podstawie poziomów istotności dla uwzględnionych kontrastów. Dzięki temu możliwe jest wybranie metody korekty.

- **Najmniejsza istotna różnica.** Ta metoda nie kontroluje ogólnego prawdopodobieństwa odrzucenia hipotez, które stwierdzają, że niektóre kontrasty liniowe różnią się od wartości hipotezy zerowej.
- *Sekwencyjna Bonferroniego.* Jest to sekwencyjne zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.
- *Sekwencyjna Sidaka.* Jest to sekwencyjnie zstępująca odrzucająca procedura Bonferroniego, która jest mniej konserwatywna w zakresie odrzucania indywidualnych hipotez, ale zachowuje identyczny całościowy poziom istotności.

Metoda najmniejszej istotnej różnicy jest mniej konserwatywna niż metoda liczby sekwencyjnej Sidaka, która z kolei jest mniej konserwatywna niż metoda sekwencyjna Bonferroniego; oznacza to, że najmniejsza istotna różnica odrzuci co najmniej taką liczbę pojedynczych hipotez, co metoda sekwencyjna Sidaka, która z kolei odrzuci co najmniej taką liczbę pojedynczych hipotez, co metoda Sekwencyjna Bonferroniego.

## Zapisz

Zaznaczone elementy są zapisywane z podaną nazwą; konflikty z istniejącymi nazwami pól są niedozwolone.

**Wartości przewidywane-** Zapisuje przewidywaną wartość zmiennej przewidywanej. Domyślna nazwa pola to *PredictedValue*.

**Przewidywane prawdopodobieństwo dla jakościowych zmiennych przewidywanych-** Jeśli zmienna przewidywana jest jakościowa, to słowo kluczowe zapisuje przewidywane prawdopodobieństwa pierwszych  $n$  kategorii, aż do wartości określonej jako **Maksymalna liczba kategorii do zapisania**. Obliczone wartości są skumulowanymi prawdopodobieństwami dla porządkowej zmiennej przewidywanej. Domyślną nazwą główną jest *PredictedProbability*. Aby zapisać przewidywane prawdopodobieństwo przewidywanej kategorii, zapisz ufność (patrz poniżej).

**Przedziały ufności-** Zapisuje górną i dolną granicę przedziału ufności dla przewidywanej wartości lub przewidywanego prawdopodobieństwa. W przypadku wszystkich rozkładów oprócz dwumianowych taki sposób działania powoduje utworzenie dwóch zmiennych, a domyślna nazwa główna ma wartość *CI*, przy czym *\_Lower* i *\_Upper* są dołączane jako przyrostki.

W przypadku rozkładu wielomianowego oraz nominalnej zmiennej przewidywanej tworzone jest jedno pole dla każdej kategorii zmiennej zależnej. Dzięki temu zapisywane są dolne i górne granice przewidywanego prawdopodobieństwa dla pierwszych  $n$  kategorii aż do wartości określonej jako **Maksymalna liczba kategorii zapisywanych**. Domyślną nazwą główną jest *CI*, a domyślne nazwy pól to *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* itd. – odpowiednio do kolejności kategorii przewidywanych.

W przypadku rozkładu wielomianowego oraz nominalnej zmiennej przewidywanej tworzone jest jedno pole dla każdej kategorii zmiennej zależnej (więcej informacji można znaleźć w temacie [“Ogólne opcje budowania”](#) na stronie 59). Dzięki temu zapisywane są dolne i górne granice skumulowanego przewidywanego prawdopodobieństwa  $n$  kategorii, aż do ostatniej, ale bez ostatniej, i aż do wartości określonej jako **Maksymalna liczba kategorii zapisywanych**. Domyślną nazwą główną jest *CI*, a domyślne nazwy pól to *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* itd. – odpowiednio do kolejności kategorii przewidywanych.

**Reszty Pearsona-** zapisuje resztę Pearsona dla każdego rekordu, która może być użyta w diagnostyce po estymacji dopasowania modelu. Domyślna nazwa pola to *PearsonResidual*.

**Ufności-** Zapisuje ufność w wartości przewidywanej dla jakościowej zmiennej przewidywanej. Obliczona ufność może być oparta na prawdopodobieństwie przewidywanej wartości (najwyższe przewidywane prawdopodobieństwo) lub różnicy między najwyższym przewidywanym prawdopodobieństwem a drugim co do wysokości przewidywanym prawdopodobieństwem. Domyślna nazwa pola to *Confidence*.

## Eksportuj

**Eksportuj model.** Powoduje to zapisanie modelu w zewnętrznym pliku *.zip* . Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych. Więcej informacji zawiera temat [Kreator oceniania](#) . Podaj niepowtarzalną prawidłową nazwę pliku. Jeśli specyfikacja pliku odpowiada istniejącemu plikowi, zostanie on nadpisany.

### Eksportowanie raportu wynikowego

**Eksportuj EBLUPS.** Ta sekcja jest włączona, jeśli co najmniej jeden blok efektów losowych został utworzony za pomocą okna dialogowego Efekty losowe na karcie Zmienne i efekty, a pole wyboru służące do wyświetlania predykcji parametrów dla tego bloku jest zaznaczone. Jeśli określono wiele takich bloków, można zapisać wyniki do oddzielnych zbiorów danych lub plików albo połączyć je w jeden zbiór danych lub plik.

## Widok modelu

Procedura tworzy obiekt modelu w oknie raportów. Przez aktywację (dwukrotne kliknięcie) obiektu użytkownik uzyskuje interaktywny przegląd modelu.

Domyślnie wyświetlany jest widok Podsumowanie modelu. Aby wyświetlić inny widok, należy wybrać go z miniatur widoków.

Alternatywą dla obiektu Model jest generowanie tabel przestawnych i wykresów, wybierając opcję **Tabele krzyżowe i wykresy** w grupie Wyświetlanie wyników na karcie Dane wyjściowe w oknie dialogowym Opcje (Edycja > Opcje). W poniższych tematach obiekt Model.

### Podsumowanie modelu

Widok jest obrazem stanu, szybkim podsumowaniem modelu i jego dopasowania.

**Tabela.** Tabela identyfikuje zmienną przewidywaną, rozkład prawdopodobieństwa oraz funkcja łączenia, które są określone w obszarze Ustawienia zmiennej przewidywanej. Jeśli zmienna przewidywana jest zdefiniowana przez zdarzenia i próby, komórka zostanie podzielona w celu przedstawienia zmiennej zdarzeń i zmiennej prób albo stałą liczbę prób. Dodatkowo zostanie wyświetlone skorygowane kryterium informacyjne Akaike (AICC) i Bayesowskie kryterium informacyjne (BIC) dla próby skończonej.

- *Skorygowane Akaike.* Miara wybierania i porównywania modeli mieszanych tworzonych na podstawie  $-2$  logarytmu wiarygodności (ograniczonej). Mniejsze wartości oznaczają lepszy model. Wartość AICC „poprawia” wartość AIC w przypadku małych prób. Przy wzroście wielkości próby wartość AICC zbiega do wartości AIC.
- *Bayesowskie.* Miara wybierania i porównywania modeli tworzonych na podstawie  $-2$  logarytmu wiarygodności. Mniejsze wartości oznaczają lepszy model. Wartość BIC „karze” także modele przeparametryzowane (na przykład złożone modele z dużą liczbą danych wejściowych), jednak silniej niż miara AIC.

**Wykres.** Jeśli zmienna przewidywana jest jakościowa, na wykresie zostanie wyświetlona dokładność modelu końcowego, która jest procentem poprawnych klasyfikacji.

### Struktura danych

Ten widok udostępnia podsumowanie struktury danych określonej przez użytkownika i ułatwia sprawdzenie, czy obiekty i powtarzane pomiary zostały poprawnie określone. Informacje obserwowane dla pierwszego obiektu są wyświetlane dla każdej zmiennej obiektowej i każdej zmiennej pomiarów

powtarzanych, a także dla zmiennej przewidywanej. Dodatkowo wyświetlana jest liczba poziomów dla każdej zmiennej obiektowej i zmiennej pomiarów powtarzanych.

## Przewidywane według obserwowanych

W przypadku ilościowych zmiennych przewidywanych określonych jako zdarzenia/próby ten wykres przedstawia wykres rozrzutu z kategoryzacją wartości przewidywanych (na osi pionowej) na wartości obserwowane (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.

## Klasyfikacja

W przypadku jakościowych zmiennych przewidywanych wyświetlana jest klasyfikacja krzyżowa wartości obserwowanych względem przewidywanych w mapie natężeń, a dodatkowo przedstawiany jest ogólny procent poprawnych.

**Style tabel.** Istnieje kilka różnych stylów wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Procent w wierszu.** Wyświetla procent w wierszu (liczby komórek wyrażone jako procent sum z wierszy) w komórkach. Jest to wartość domyślna.
- **Liczby komórek.** Wyświetla liczby komórek w komórkach. Cieniowanie mapy komórek jest w dalszym ciągu wyrażone jako procent w wierszu.
- **Mapa natężeń.** W komórkach nie są wyświetlane żadne wartości, tylko cieniowanie.
- **Skompresowane.** Nagłówki kolumn ani wierszy nie są wyświetlane. Nie są wyświetlane również wartości w komórkach. Taki styl może być użyteczny, gdy zmienna przewidywana zawiera wiele kategorii.

**Braki danych.** Jeśli w jakichkolwiek rekordach brakuje wartości w zmiennej przewidywanej, wówczas takie rekordy są wyświetlane w wierszu (**Braki danych**), który jest wyświetlany pod poprawnymi wierszami. Rekordy z brakami danych nie zwiększają ogólnego procentu poprawnych.

**Wiele zmiennych przewidywanych.** Jeśli istnieje wiele przewidywanych zmiennych jakościowych, wówczas każda zmienna przewidywana jest wyświetlana w osobnej tabeli i istnieje lista rozwijana **Zmienna przewidywana**, która kontroluje wyświetlane zmienne przewidywane.

**Duże tabele.** Jeśli wyświetlana zmienna przewidywana zawiera ponad 100 kategorii, żadna tabela nie jest wyświetlana.

## Efekty stałe

Ten widok przedstawia rozmiar każdego efektu stałego w modelu.

**Style.** Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Diagram.** Jest to wykres, w którym efekty są posortowane od góry do dołu w kolejności, w jakiej zostały określone w ustawieniach Efekty stałe. Linie łączące w diagramie są ważone na podstawie istotności efektu, gdzie większa szerokość linii odpowiada bardziej istotnym efektom (niższe wartości  $p$ ). Jest to wartość domyślna.
- **Tabela.** Jest to tabela ANOVA dla ogólnego całkowitych i pojedynczych efektów modelu. Poszczególne efekty są posortowane od góry do dołu w kolejności, w jakiej zostały określone w ustawieniach Efekty stałe.

**Istotność.** Dostępny jest suwak Istotność, który steruje widocznością efektów w widoku. 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 stałe

Widok ten przedstawia wartość każdego współczynnika stałego 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 współczynnikowi nadmiarowemu.

**Style.** Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Diagram.** Jest to wykres, w którym najpierw wyświetlany jest wyraz wolny, a następnie sortowane są efekty od góry do dołu w kolejności, w jakiej zostały określone w ustawieniach Efekty stałe. W efektach zawierających czynniki, współczynniki są posortowane rosnąco według wartości danych. Linie łączące w diagramie są kolorowane i wazone na podstawie istotności współczynnika, gdzie większa szerokość linii odpowiada bardziej istotnym współczynnikom (niższe wartości  $p$ ). Jest to domyślny styl.
- **Tabela.** Pokazuje ona wartości, testy istotności i przedziały ufności dla poszczególnych współczynników modelu. Po wyrazie wolnym poszczególne efekty są posortowane od góry do dołu w kolejności, w jakiej zostały określone w ustawieniach Efekty stałe. W efektach zawierających czynniki, współczynniki są posortowane rosnąco według wartości danych.

**Wielomianowy.** Jeśli obowiązuje rozkład wielomianowy, wówczas lista rozwijana Wielomianowy kontroluje to, która kategoria zmiennych przewidywanych będzie wyświetlana. Kolejność wyświetlania wartości na liście jest określona przez specyfikację w ustawieniach Opcje budowania.

**Wykładnicza.** Wyświetlane są oszacowania współczynników wykładniczych i przedziały ufności dla niektórych typów modeli, w tym dla binarnej regresji logistycznej (rozkład dwumianowy i funkcja łączenia logit), nominalnej regresji logistycznej (rozkład wielomianowy i funkcja łączenia logit), negatywnej regresji dwumianowej (negatywna regresja dwumianowa i logarytmiczna funkcja łączenia) oraz modelu logarytmiczno-liniowego (rozkład Poissona i logarytmiczna funkcja łączenia).

**Istotność.** Dostępny jest suwak Istotność, który steruje widocznością współczynników w widoku. 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.

## Kowariancje efektów losowych

W tym widoku przedstawiana jest macierz kowariancji efektów losowych (**G**).

**Style.** Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Wartości kowariancji.** Jest to mapa natężeń macierzy kowariancji, w której efekty są posortowane od góry do dołu w kolejności, w jakiej zostały określone w ustawieniach Efekty stałe. Kolory na diagramie corrgram odpowiadają wartościom komórek przedstawionych w kluczu. Jest to wartość domyślna.
- **Corrgram.** Jest to mapa natężeń macierzy kowariancji.
- **Skompresowane.** Jest to mapa natężeń macierzy kowariancji bez nagłówek wierszy i kolumn.

**Bloki.** Jeśli istnieje wiele bloków efektów losowych, wówczas istnieje lista rozwijana Blok przeznaczona do wyboru bloku do wyświetlenia.

**Grupy.** Jeśli blok efektów losowych obejmuje specyfikację grupy, wówczas istnieje lista rozwijana Grupa przeznaczona do wyboru poziomu grupy do wyświetlenia.

**Wielomianowy.** Jeśli obowiązuje rozkład wielomianowy, wówczas lista rozwijana Wielomianowy kontroluje to, która kategoria zmiennych przewidywanych będzie wyświetlana. Kolejność wyświetlania wartości na liście jest określona przez specyfikację w ustawieniach Opcje budowania.

## Parametry kowariancji

Ten widok kontroluje oszacowania parametrów kowariancji i powiązane statystyki dla efektów reszty i efektów losowych. Są to zaawansowane, ale fundamentalne wyniki, które udostępniają informację na temat tego, czy struktura kowariancji jest odpowiednia.

**Tabela podsumowań.** Jest to krótki przegląd liczby parametrów w macierzach kowariancji reszt (**R**) i efektów losowych (**G**), ranga (liczba kolumn) w efekcie losowym (**X**) oraz macierzach projektów efektu losowego (**Z**), a także liczba obiektów zdefiniowanych przez zmienne obiektowe, które definiują strukturę danych.

**Tabela parametrów kowariancji.** Dla wybranego efektu wyświetlany jest wybrany efekt, oszacowanie, błąd standardowy oraz przedział ufności dla każdego parametru kowariancji. Liczba widocznych parametrów jest zależna od struktury kowariancji dla efektu, bloków efektów losowych, a także od liczby efektów w bloku. Jeśli widoczne jest, że parametry, które nie leżą na przekątnej, są nieistotne, być może oznacza to, że możliwe będzie użycie prostszej struktury kowariancji.

**Efekty.** Jeśli istnieją bloki efektów losowych, wówczas dostępna jest lista rozwijana Efekt, z której można wybrać efekt reszty lub efekt losowy do wyświetlenia. Efekt reszty jest zawsze dostępny.

**Grupy.** Jeśli blok efektów reszty lub losowych obejmuje specyfikację grupy, wówczas istnieje lista rozwijana Grupa przeznaczona do wyboru poziomu grupy do wyświetlenia.

**Wielomianowy.** Jeśli obowiązuje rozkład wielomianowy, wówczas lista rozwijana Wielomianowy kontroluje to, która kategoria zmiennych przewidywanych będzie wyświetlana. Kolejność wyświetlania wartości na liście jest określona przez specyfikację w ustawieniach Opcje budowania.

## Oszacowane średnie: efekty istotne

Są to wykresy wyświetlane dla 10 „najbardziej istotnych” efektów o wszystkich czynnikach ustalonych, począwszy od interakcji trzeciego rzędu, następnie interakcje drugiego rzędu, a na koniec efekty główne. Na osi pionowej wykresu przedstawiona jest wartość zmiennej przewidywanej oszacowana przez model dla każdej wartości efektu głównego (lub efektu, który znajduje się pierwszy na liście w interakcji) na osi poziomej; osobna linia jest generowana dla każdej wartości drugiego efektu z listy w interakcji; osobny wykres jest generowany dla każdej wartości trzeciego efektu z listy w interakcji trzeciego rzędu; wszystkie pozostałe predyktory są utrzymywane jako stałe. Zapewnia on przydatną wizualizację efektów współczynników docelowej każdego predyktora. Należy zwrócić uwagę, że jeśli żaden predyktor nie jest istotny, nie jest generowana żadna oszacowana średnia.

**Ufność.** Przedstawiane są górne i dolne granice ufności dla średnich brzegowych — na podstawie poziomu ufności określonego w ramach opcji budowania.

## Oszacowane średnie: efekty użytkownika

Są to tabele i wykresy dla żądanych przez użytkownika efektów o wszystkich czynnikach ustalonych.

**Style.** Dostępne są różne style wyświetlania, które są dostępne z poziomu listy rozwijanej **Styl**.

- **Diagram.** W przypadku tego stylu na osi pionowej wykresu liniowego przedstawiona jest wartość zmiennej przewidywanej oszacowana przez model dla każdej wartości efektu głównego (lub efektu, który znajduje się pierwszy na liście w interakcji) na osi poziomej; osobna linia jest generowana dla każdej wartości drugiego efektu z listy w interakcji; osobny wykres jest generowany dla każdej wartości trzeciego efektu z listy w interakcji trzeciego rzędu; wszystkie pozostałe predyktory są utrzymywane jako stałe.

Jeśli zażądano kontrastów, wówczas wyświetlany jest inny wykres w celu porównania poziomów zmiennej kontrastu; na potrzeby interakcji wykres jest wyświetlany dla każdego poziomu kombinacji efektów innych niż zmienna kontrastu. W przypadku kontrastów **parami** wykres sieci odległości to graficzne odzwierciedlenie tabeli porównań, w przypadku której odległości między węzłami w sieci odpowiadają różnicom między próbami. Żółte linie odpowiadają różnicom istotnym pod względem statystycznym. Linie czarne odpowiadają różnicom nieistotnym. Ustawienie kursora na sieci powoduje wyświetlenie podpowiedzi zawierającej skorygowaną istotność różnicy między węzłami połączonymi za pomocą linii.

W przypadku kontrastów **odchylenie** wyświetlany jest wykres słupkowy z oszacowaną przez model wartością zmiennej przewidywanej na osi pionowej oraz wartościami zmiennej kontrastu na osi poziomej; na potrzeby interakcji wyświetlany jest wykres dla każdego poziomu kombinacji efektów

innych niż zmienna kontrastu. Słupki przedstawiają różnice między poszczególnymi poziomami zmiennej kontrastu oraz średnią ogólną, którą reprezentuje czarna linia pozioma.

W przypadku kontrastów **prostych** wyświetlany jest wykres słupkowy z oszacowaną przez model wartością zmiennej przewidywanej na osi pionowej oraz wartościami zmiennej kontrastu na osi poziomej; na potrzeby interakcji wyświetlany jest wykres dla każdego poziomu kombinacji efektów innych niż zmienna kontrastu. Słupki przedstawiają różnice między poszczególnymi poziomami zmiennej kontrastu (za wyjątkiem ostatniego) oraz poziomem ostatnim, który reprezentuje czarna linia pozioma.

- **Tabela.** W tym stylu wyświetlana jest tabela zawierająca oszacowaną przez model wartość zmiennej przewidywanej, jej błąd standardowy oraz przedział ufności dla każdego poziomu kombinacji obowiązujących zmiennych; wszystkie pozostałe predyktory są utrzymywane jako stałe.

Jeśli żądane były kontrasty, wówczas wyświetlana jest inna tabela z oszacowaniem, błędem standardowym, testem istotności i przedziałem ufności dla każdego kontrastu; na potrzeby interakcji istnieje osobny zestaw wierszy dla każdego poziomu kombinacji efektów innych niż zmienna kontrastu. Dodatkowo wyświetlana jest tabela z wynikami testu ogólnego; na potrzeby interakcji istnieje osobny test ogólny dla każdego poziomu kombinacji efektów innych niż zmienna kontrastu.

**Ufność.** Ta opcja umożliwia przełączanie górnych i dolnych przedziałów ufności dla średnich brzegowych – na podstawie poziomu ufności określonego w ramach opcji budowania.

**Układ.** Ta opcja umożliwia przełączanie układu diagramu kontrastów par. Układ kołowy przedstawia mniej informacji na temat kontrastów niż układ sieciowy, ale eliminuje problem przecinających się linii.

## Analiza logliniowa: Wybór modelu

---

Procedura Analiza logliniowa: Wybór modelu analizuje tabele krzyżowe wielu rzędów (tabele kontyngencji). Dopasowuje hierarchiczne modele nieliniowe do wielowymiarowych tabel krzyżowych, używając iteracyjnego algorytmu dopasowywania proporcjonalnego. Ta procedura ułatwia ustalenie, które zmienne kategoriaalne są skojarzone. Do budowania modeli dostępne są metody wprowadzania wymuszonego i eliminacji wstecznej. W przypadku modeli nasyconych można zażądać oszacowań parametrów i testów pod kątem cząstkowych powiązań. Model nasycony dodaje 0,5 do wszystkich komórek.

**Przykład.** W badaniu preferencji jednego z dwóch detergentów do prania badacze policzyli ludzi w każdej grupie, łącząc różne kategorie twardości wody (miękka, średnia, twarda), poprzednie użycie produktów jednej z marek oraz temperaturę prania (pranie w niskiej lub wysokiej temperaturze). Odkryli zależność między temperaturą a twardością wody oraz preferencją w odniesieniu do marek.

**Statystyki.** Częstotliwości, reszty, oszacowania parametrów, błędy standardowe, przedziały ufności oraz testy powiązania cząstkowego. W przypadku modeli niestandardowych wykresy reszt i wykresy prawdopodobieństwa normalnego.

Wymagania dotyczące danych do analizy logliniowej: wybór modelu

**Dane.** Zmienne czynnika są jakościowe. Wszystkie zmienne przeznaczone do analizowania muszą być numeryczne. Jakościowe zmienne łańcuchowe muszą zostać zarejestrowane w zmiennych liczbowych przed rozpoczęciem analizy wyboru modelu.

Należy unikać określania wielu zmiennych o wielu poziomach. Takie specyfikacje mogą prowadzić do sytuacji, w której wiele komórek zawiera niewielką liczbę obserwacji, a wartości chi-kwadrat mogą być nieużyteczne.

**Procedury pokrewne.** Procedura wyboru modelu może ułatwić zidentyfikowanie składników wymaganych w modelu. Następnie można kontynuować ocenę modelu przy użyciu ogólnej analizy logliniowej albo analizy logliniowej Logit. W celu rekodowania zmiennych łańcuchowych można korzystać z opcji automatycznego rekodowania. Jeśli zmienna liczbową zawiera puste kategorie, należy użyć opcji Rekoduj w celu utworzenia ciągłych zmiennych całkowitych.

Uzyskiwanie analizy logliniowej: wybór modelu

Z menu wybierz:



## Analiza > Analizy logliniowe > Wybór modelu...

1. Wybierz co najmniej dwa czynniki jakościowe numeryczne.
2. Wybierz co najmniej jedną zmienną czynnika z listy Czynniki, a następnie kliknij przycisk **Definiuj zakres**.
3. Zdefiniuj zakres wartości każdej zmiennej czynnika.
4. Zaznacz opcję w grupie Budowa modelu.

Opcjonalnie można wybrać zmienną wagi komórek, aby określić zera strukturalne.

## Analiza logliniowa: Definiuj zakres

Musisz wskazać zakres kategorii dla każdej zmiennej czynnikowej. Wartości Minimum i Maksimum odpowiadają minimalnym i maksymalnym kategoriom zmiennej czynnika. Obie wartości muszą być całkowite, a wartość minimalna musi być mniejsza niż wartość maksymalna. 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 1, 2 oraz 3. Procedurę należy powtórzyć dla każdej zmiennej czynnika.

## Analiza logliniowa: Model

**Określ model.** Model nasycony obejmuje efekty główne wszystkich czynników oraz interakcje wszystkich czynników. Wybierz opcję **Niestandardowa**, aby określić klasę tworzącą dla modelu nienasyconego.

**Klasa tworząca.** Klasa tworząca jest listą składników najwyższego rzędu, w którym występują czynniki. Model hierarchiczny zawiera składniki, które definiują klasę tworzącą oraz wszystkie wartości względne niższego rzędu. Załóżmy, że wybrano zmienne  $A$ ,  $B$  oraz  $C$  na liście Czynniki, a następnie wybrano opcję **Interakcje** na liście rozwijanej Budowanie składników. Wynikowy model będzie zawierał określoną 3-drożną interakcję  $A*B*C$ , interakcje 2-drożne  $A*B$ ,  $A*C$  i  $B*C$  oraz główne wyniki dla  $A$ ,  $B$  i  $C$ . Nie należy określać krewnych niższego rzędu w klasie generującej.

Dla wybranych czynników:

### 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:

## Opcje analizy logliniowej: wybór modelu

**Pokaż.** Można wybrać opcję **Częstości, Reszty** lub obie te opcje. W modelu nasyconym częstości obserwowane i oczekiwane są równe, a reszty są równe 0.

**Wykres.** W przypadku modeli niestandardowych można wybrać jeden lub oba typy wykresów — **Reszty** i **Prawdopodobieństwo normalne**. Dzięki nim można ustalić stopień dopasowania modelu do danych.

**Pokaż dla modelu nasyconego.** W przypadku modelu nasyconego można wybrać opcję **Oceny parametrów**. Oceny parametrów mogą pomóc w ustaleniu składników, które można usunąć z modelu. Dostępna jest także tabela skojarzeń, która zawiera listę testów o skojarzeniu częściowym. Ta opcja bardziej obciąża zasoby obliczeniowe w przypadku tabel o wielu czynnikach.

**Kryteria modelu.** W celu uzyskania oszacowań parametrów używany jest interaktywny algorytm dopasowania proporcjonalnego. Co najmniej jedno kryterium oszacowania można zastąpić poprzez określenie wartości **Maksymalna liczba iteracji**, **Zbieżność** lub **Delta** (wartość dodawana do częstości komórkowych w przypadku modeli nasyconych).

## Dodatkowe właściwości komendy HILOGLINEAR

Język składni komend umożliwia również:

- Określenie wag komórek w postaci macierzy (za pomocą opcji komendy CWEIGHT).
- Generowanie analiz kilku modeli przy użyciu jednej komendy (za pomocą opcji komendy DESIGN).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Ogólna analiza log-liniowa

Procedura Ogólna analiza logliniowa analizuje częstości występowania (liczebności) obserwacji należących do poszczególnych kategorii klasyfikacji krzyżowej w tabeli krzyżowej lub tabeli kontyngencji. Każda klasyfikacja krzyżowa w tabeli stanowi jedną komórkę, a każda zmienna jakościowa nazywana jest czynnikiem. Zmienną zależną jest liczba obserwacji (częstość) w komórce tabeli krzyżowej, a zmiennymi objaśniającymi są czynniki i współzmienne. Ta procedura szacuje parametry największej wiarygodności hierarchicznych i niehierarchicznych modeli logliniowych przy użyciu metody Newtona-Raphsona. Można analizować albo rozkład Poissona, albo wielomianowy.

Można wybrać maksymalnie 10 czynników definiujących komórki tabeli. Zmienna struktury komórki umożliwia zdefiniowanie zer strukturalnych dla tabel niekompletnych, uwzględnienie składnika przesunięcia w modelu, dopasowanie modelu typu log-rate lub zaimplementowanie metody korygowania tabel brzegowych. Kontrasty umożliwiają obliczanie uogólnionych ilorazów logarytmów szans (GLOR).

Automatycznie wyświetlane są informacje o modelu i statystyka dobroci dopasowania. Można także wyświetlać różne statystyki i wykresy oraz zapisywać reszty i wartości przewidywane w aktywnym zbiorze danych.

**Przykład.** Dane z raportu o wypadkach samochodowych na Florydzie są podstawą do ustalania zależności między zapięciem pasów bezpieczeństwa a tym, czy wypadek był śmiertelny, czy nie. Iloraz szans jest istotnym wskaźnikiem takiej zależności.

**Statystyki.** Częstości obserwowane i oczekiwane; reszty pierwotne, skorygowane i niedopasowania; macierz planu; oszacowania parametrów; iloraz szans; iloraz logarytmu szans; GLOR; statystyka Walda i przedziały ufności. Wykresy: reszt skorygowanych, reszt niedopasowania i prawdopodobieństwa normalnego.

Wymagania dotyczące danych dla ogólnej analizy logliniowej

**Dane.** Czynniki są zmiennymi jakościowymi, a współzmienne komórek są zmiennymi ciągłymi. Gdy model zawiera współzmienną, względem komórki stosowana jest średnia wartość współzmiennnej dla obserwacji w komórce. Kontrasty są zmiennymi ciągłymi. Zazwyczaj są one używane do wyliczania zmiennych typu flaga. Wartości kontrastów są współczynnikami liniowej kombinacji logarytmów oczekiwanych liczebności komórek.

Zmienna struktury komórki przypisuje wagi. Na przykład, jeśli niektóre z komórek są zerami strukturalnymi, zmienna struktury komórki ma wartość 0 lub 1. Nie należy używać zmiennej struktury komórek do ważenia danych zagregowanych. Zamiast tego należy wybrać opcję **Ważenie obserwacji** w menu Dane.

**Założenia.** W ogólnej analizie logliniowej dostępne są dwa rozkłady: Poissona i wielomianowy.

Przy założeniu o rozkładzie Poissona:

- Całkowity rozmiar próby nie jest ustalony przed badaniem lub analiza nie jest zależna od całkowitego rozmiaru próby.
- Zdarzenie obserwacji w komórce jest statystycznie niezależne od liczebności w innych komórkach.

Przy założeniu o rozkładzie wielomianowym:

- Całkowity rozmiar próby jest ustalony lub analiza jest zależna od całkowitego rozmiaru próby.
- Liczebności w komórkach nie są statystycznie niezależne.

**Procedury pokrewne.** Do badania tabel krzyżowych należy używać procedury Tabele krzyżowe. Z procedury Analiza logliniowa Logit należy korzystać, gdy co najmniej jedną zmienną jakościową można w naturalny sposób traktować jako zmienne odpowiedzi, a pozostałe jako zmienne objaśniające.

Uzyskiwanie ogólnej analizy logliniowej

1. Wybierz z menu następującą opcję:

**Analiza > Analizy logliniowe > Ogólna...**

2. W oknie dialogowym Ogólna analiza logliniowa wybierz maksymalnie 10 zmiennych czynników.

Opcjonalnie możesz wykonać następujące czynności:

- Wybierz współzmiennie komórek.
- Wybierz zmienną struktury komórki, aby zdefiniować zera strukturalne lub uwzględnić składnik przesunięcia.
- Wybierz kontrast.

## Ogólna analiza logliniowa: Model

**Określ model.** Model nasycony zawiera wszystkie efekty główne i interakcje z udziałem zmiennych czynnika. Nie zawiera współzmiennych. Aby samodzielnie określić tylko podzbiór interakcji lub interakcje czynnik-współzmienna, wybierz opcję **Użytkownika**.

**Czynniki i współzmiennie.** Lista zawiera czynniki i współzmiennie.

**Składniki w modelu.** Model zależy od charakteru danych. Po wybraniu opcji **Użytkownika** można wybrać efekty główne oraz interakcje będące przedmiotem zainteresowania w czasie analizy. Należy określić wszystkie składniki modelu.

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:

## Ogólna analiza logliniowa: Opcje

Procedura Ogólna analiza logliniowa wyświetla informacje o modelu i statystyki dobroci dopasowania. Ponadto można wybrać jedną lub więcej spośród następujących opcji:

**Pokaż.** Do wyświetlania dostępnych jest kilka statystyk: obserwowane i oczekiwane liczebności w komórkach; reszty pierwotne, skorygowane i niedopasowania; macierz planu modelu; oraz oszacowania parametrów modelu.

**Wykres.** Wykresy (dostępne tylko w przypadku modeli użytkownika) obejmują dwie macierze rozrzutu (reszty skorygowane lub reszty niedopasowania względem obserwowanych i oczekiwanych liczebności w komórkach). Można także wyświetlić normalny wykres prawdopodobieństwa i normalne wykresy bez trendu reszt skorygowanych lub reszt niedopasowania.

**Oszacowanie przedziału ufności.** Przedział ufności oszacowań parametrów można skorygować.

**Kryteria.** Do uzyskania oszacowań parametrów największej wiarygodności używana jest metoda Newtona-Raphsona. Nowe wartości można wprowadzać dla maksymalnej liczby iteracji, dla kryterium zbieżności oraz dla delty (stała dodawana do wszystkich komórek na potrzeby początkowych przybliżeń). W przypadku modeli nasyconych delta pozostaje w komórkach.

## Ogólna analiza logliniowa: Zapisz

Wybierz wartości, które chcesz zapisać jako nowe zmienne w aktywnym zbiorze danych. Przyrostek  $n$  w nazwach nowych zmiennych jest zwiększany o jeden, aby każda zapisana zmienna miała unikalną nazwę.

Zapisane wartości odnoszą się do danych zagregowanych (komórek w tabeli kontyngencji), nawet jeśli dane są zarejestrowane w poszczególnych obserwacjach w Edytorze danych. Jeśli użytkownik zapisze reszty lub wartości przewidywane dla danych niezagregowanych, to zapisana wartość komórki w tabeli kontyngencji zostanie w Edytorze danych wprowadzona dla każdej obserwacji w tej komórce. Aby prawidłowo zinterpretować zapisane wartości, należy zagregować dane w celu uzyskania liczebności w komórkach.

Można zapisywać cztery typy reszt: pierwotne, standaryzowane, skorygowane i niedopasowania. Możliwe jest także zapisywanie wartości przewidywanych.

- *Reszty.* Nazywana też resztą zwykłą lub surową, definiowana jako różnica pomiędzy oczekiwaną a obserwowaną liczbą komórek.
- *Reszty standaryzowane.* Reszta podzielona przez oszacowanie jej błędu standardowego. Standaryzowane reszty są także znane jako reszty Pearsona.
- *Reszty skorygowane.* Standaryzowana reszta podzielona przez jej oszacowany błąd standardowy. Ponieważ przy prawidłowym wyborze modelu reszty skorygowane mają rozkład standaryzowany asymptotycznie normalny, lepiej niż reszty standaryzowane nadają się do sprawdzania normalności.
- *Reszty odchyień.* Pierwiastek kwadratowy ze znakiem, z indywidualnego udziału w statystyce ilorazu wiarygodności chi-kwadrat (G kwadrat), gdzie znak jest znakiem reszty (liczebność obserwowana minus liczebność oczekiwana). Reszty odchylenia mają standardowy rozkład asymptotycznie normalny.

## Dodatkowe właściwości komendy GENLOG

Język składni komend umożliwia również:

- Obliczanie liniowych kombinacji obserwowanych i oczekiwanych częstości komórkowych, a także drukowanie reszt, reszt standaryzowanych i reszt skorygowanych danej kombinacji (za pomocą opcji komendy GERESID).
- Zmianę domyślnej wartości progowej dla kontroli nadmiarowości (za pomocą opcji komendy CRITERIA).
- Wyświetlanie reszt standaryzowanych (za pomocą opcji komendy PRINT).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Analiza logliniowa Logit

Procedura Analiza logliniowa Logit analizuje relację między zmiennymi zależnymi (odpowiedzi) a zmiennymi niezależnymi (objaśniającymi). Zmienne zależne są zawsze jakościowe, natomiast zmienne niezależne mogą być zmiennymi jakościowymi (czynniki). Pozostałe zmienne niezależne (współzmiennie w komórce) mogą być ciągłe, ale nie są stosowane obserwacja po obserwacji. Ważona średnia współzmiennej dla komórki jest stosowana względem tej komórki. Logarytm szans zmiennych zależnych jest wyrażony jako liniowa kombinacja parametrów. Automatycznie przyjmuje się rozkład wielomianowy; modele te nazywane są niekiedy wielomianowymi modelami logit. Ta procedura szacuje parametry modeli logliniowych logit przy użyciu algorytmu Newtona-Raphsona.

Można wybrać od 1 do 10 zmiennych zależnych i zmiennych czynników. Zmienna struktury komórki umożliwia zdefiniowanie zer strukturalnych dla tabel niekompletnych, uwzględnienie składnika przesunięcia w modelu, dopasowanie modelu typu log-rate lub zaimplementowanie metody korygowania tabel brzegowych. Kontrasty umożliwiają obliczanie uogólnionych ilorazów logarytmów szans (GLOR). Wartości kontrastów są współczynnikami liniowej kombinacji logarytmów oczekiwanych liczebności komórek.

Automatycznie wyświetlane są informacje o modelu i statystyka dobroci dopasowania. Można także wyświetlać różne statystyki i wykresy oraz zapisywać reszty i wartości przewidywane w aktywnym zbiorze danych.

**Przykład.** Badaniem na Florydzie objęto 219 aligatorów. Jaka jest zależność między wielkością ciała aligatora i jeziorem (jednym z czterech), które zamieszkuje, a rodzajem jego pożywienia. Badanie wykazało, że szansa, iż mniejszy aligator będzie chętniej jadł gady niż ryby, jest 0,70 raza mniejsza niż w przypadku większych aligatorów; ponadto szanse, że pierwszym wyborem aligatora będą gady, a nie ryby, były najwyższe w jeziorze nr 3.

**Statystyki.** Częstości obserwowane i oczekiwane; reszty pierwotne, skorygowane i niedopasowania; macierz planu; oszacowania parametrów; uogólniony iloraz logarytmu szans; statystyka Walda i przedziały ufności. Wykresy: reszt skorygowanych, reszt niedopasowania i prawdopodobieństwa normalnego.

Wymagania dotyczące danych dla analizy logliniowej Logit

**Dane.** Zmienne zależne są jakościowe. Czynniki są typu jakościowego. Współzmiennie w komórkach mogą być ciągłe, jednak gdy model zawiera współzmienną, względem komórki stosowana jest średnia wartość współzmiennej dla obserwacji w komórce. Kontrasty są zmiennymi ciągłymi. Służą do obliczania uogólnionych ilorazów logarytmów szans (GLOR). Wartości kontrastów są współczynnikami liniowej kombinacji logarytmów oczekiwanych liczebności komórek.

Zmienna struktury komórki przypisuje wagi. Na przykład, jeśli niektóre z komórek są zerami strukturalnymi, zmienna struktury komórki ma wartość 0 lub 1. Nie należy używać zmiennej struktury komórek do ważenia danych zagregowanych. Zamiast tego należy użyć opcji Ważenie obserwacji w menu Dane.

**Założenia.** Zakłada się, że liczebności w każdej kombinacji kategorii zmiennych objaśniających, mają rozkład wielomianowy. Przy założeniu o rozkładzie wielomianowym:

- Całkowity rozmiar próby jest ustalony lub analiza jest zależna od całkowitego rozmiaru próby.

- Liczebności w komórkach nie są statystycznie niezależne.

**Procedury pokrewne.** Za pomocą procedury Tabele krzyżowe można wyświetlić tabele kontyngencji. Procedurę Ogólna analiza logliniowa stosuje się, chcąc przeanalizować relację między liczebnością obserwowaną a zbiorem zmiennych objaśniających.

Uzyskiwanie analizy logliniowej Logit

1. Wybierz z menu następującą opcję:

**Analiza > Analizy logliniowe > Logit...**

2. W oknie dialogowym Analiza logliniowa Logit wybierz co najmniej jedną zmienną zależną.
3. Wybierz co najmniej jedną zmienną czynnikową.

Suma zmiennych zależnych i czynników musi być mniejsza lub równa 10.

Opcjonalnie możesz wykonać następujące czynności:

- Wybierz współzmiennie komórek.
- Wybierz zmienną struktury komórki, aby zdefiniować zera strukturalne lub uwzględnić składnik przesunięcia.
- Wybierz co najmniej jedną zmienną będącą kontrastem.

## Analiza logliniowa Logit: Model

**Określ model.** Model nasycony zawiera wszystkie efekty główne i interakcje z udziałem zmiennych czynnikowych. Nie zawiera współzmiennych. Aby samodzielnie określić tylko podzbiór interakcji lub interakcje czynnik-współzmienna, wybierz opcję **Użytkownika**.

**Czynniki i współzmiennie.** Lista zawiera czynniki i współzmiennie.

**Składniki w modelu.** Model zależy od charakteru danych. Po wybraniu opcji **Użytkownika** można wybrać efekty główne oraz interakcje będące przedmiotem zainteresowania w czasie analizy. Należy określić wszystkie składniki modelu.

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.

Składniki dodaje się do planu, uwzględniając wszystkie możliwe kombinacje składników zależnych i dopasowując każdą kombinację do każdego składnika na liście modelu. Jeśli wybrana jest opcja **Uwzględnij stałą dla zmiennej zależnej**, to do listy modelu dodawany jest także składnik jednostkowy (1).

Załóżmy na przykład, że  $D1$  i  $D2$  są zmiennymi zależnymi. Procedura Analiza logliniowa Logit tworzy listę składników zależnych ( $D1$ ,  $D2$ ,  $D1*D2$ ). Jeśli lista Składniki w modelu zawiera  $M1$  oraz  $M2$  i uwzględniona jest stała, to lista modelu zawiera  $1$ ,  $M$  i  $M2$ . Wynikowy plan zawiera kombinacje każdego składnika modelu z każdym składnikiem zależnym.

$D1$ ,  $D2$ ,  $D1*D2$

$M1*D1, M1*D2, M1*D1*D2$

$M2*D1, M2*D2, M2*D1*D2$

**Uwzględnij stałą dla zmiennej zależnej.** Powoduje uwzględnienie stałej dla zmiennej zależnej w modelu użytkownika.

## 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:

## Analiza logliniowa Logit: Opcje

Procedura Analiza logliniowa Logit wyświetla informacje o modelu i statystyki dobroci dopasowania. Ponadto można wybrać jedną lub więcej spośród następujących opcji:

**Pokaż.** Do wyświetlania dostępnych jest kilka statystyk: obserwowane i oczekiwane liczebności w komórkach; reszty pierwotne, skorygowane i niedopasowania; macierz planu modelu; oraz oszacowania parametrów modelu.

**Wykres.** Wykresy (dostępne w przypadku modeli użytkownika) obejmują dwie macierze rozrzutu (reszty skorygowane lub reszty niedopasowania względem obserwowanych i oczekiwanych liczebności w komórkach). Można także wyświetlić normalny wykres prawdopodobieństwa i normalne wykresy bez trendu reszt skorygowanych lub reszt niedopasowania.

**Oszacowanie przedziału ufności.** Przedział ufności oszacowań parametrów można skorygować.

**Kryteria.** Do uzyskania oszacowań parametrów największej wiarygodności używana jest metoda Newtona-Raphsona. Nowe wartości można wprowadzać dla maksymalnej liczby iteracji, dla kryterium zbieżności oraz dla delty (stała dodawana do wszystkich komórek na potrzeby początkowych przybliżeń). W przypadku modeli nasyconych delta pozostaje w komórkach.

## Analiza logliniowa Logit: Zapisz

Wybierz wartości, które chcesz zapisać jako nowe zmienne w aktywnym zbiorze danych. Przyrostek  $n$  w nazwach nowych zmiennych jest zwiększany o jeden, aby każda zapisana zmienna miała unikalną nazwę.

Zapisane wartości odnoszą się do danych zagregowanych (komórek w tabeli kontyngencji), nawet jeśli dane są zarejestrowane w poszczególnych obserwacjach w Edytorze danych. Jeśli użytkownik zapisze reszty lub wartości przewidywane dla danych niezagregowanych, to zapisana wartość komórki w tabeli kontyngencji zostanie w Edytorze danych wprowadzona dla każdej obserwacji w tej komórce. Aby prawidłowo zinterpretować zapisane wartości, należy zagregować dane w celu uzyskania liczebności w komórkach.

Można zapisywać cztery typy reszt: pierwotne, standaryzowane, skorygowane i niedopasowania. Możliwe jest także zapisywanie wartości przewidywanych.

- *Reszty.* Nazywana też resztą zwykłą lub surową, definiowana jako różnica pomiędzy oczekiwaną a obserwowaną liczbą komórek.
- *Reszty standaryzowane.* Reszta podzielona przez oszacowanie jej błędu standardowego. Standaryzowane reszty są także znane jako reszty Pearsona.
- *Reszty skorygowane.* Standaryzowana reszta podzielona przez jej oszacowany błąd standardowy. Ponieważ przy prawidłowym wyborze modelu reszty skorygowane mają rozkład standaryzowany asymptotycznie normalny, lepiej niż reszty standaryzowane nadają się do sprawdzania normalności.

- *Reszty odchyłeń*. Pierwiastek kwadratowy ze znakiem, z indywidualnego udziału w statystyce ilorazu wiarygodności chi-kwadrat (G kwadrat), gdzie znak jest znakiem reszty (liczebność obserwowana minus liczebność oczekiwana). Reszty odchylenia mają standardowy rozkład asymptotycznie normalny.

## Dodatkowe właściwości komendy GENLOG

Język składni komend umożliwia również:

- Obliczanie liniowych kombinacji obserwowanych i oczekiwanych częstości komórkowych, a także drukowanie reszt, reszt standaryzowanych i reszt skorygowanych danej kombinacji (za pomocą opcji komendy GERESID).
- Zmianę domyślnej wartości progowej dla kontroli nadmiarowości (za pomocą opcji komendy CRITERIA).
- Wyświetlanie reszt standaryzowanych (za pomocą opcji komendy PRINT).

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Tabele trwania życia

W wielu sytuacjach celowe jest zbadanie rozkładu czasów, jakie upływają między dwoma zdarzeniami, np. długości zatrudnienia (czasu między rozpoczęciem a zakończeniem pracy w firmie). Jednak w tego rodzaju danych uwzględnione są zwykle także takie obserwacje, dla których drugie zdarzenie nie zostało zarejestrowane (np. osoby nadal pracujące dla firmy w chwili zakończenia badania). Istnieje kilka potencjalnych przyczyn występowania takich obserwacji: niekiedy zdarzenie po prostu nie wystąpiło przed końcem badania; w innych przypadkach utracono możliwość dalszej obserwacji przed końcem badania; w jeszcze innych sytuacjach kontynuowanie obserwacji może być niemożliwe z przyczyn niezależnych od badania (np. przebywania pracownika na zwolnieniu lekarskim). Łącznie takie obserwacje nazywane są **obserwacjami ocenzurowanymi**, a ich obecność powoduje, że tego rodzaju badania nie można przeprowadzić przy użyciu tradycyjnych testów, takich jak testy  $t$  lub regresja liniowa.

Techniką statystyczną przydatną do pracy z tego rodzaju danymi jest **tabela trwania życia**. Istotą tej metody jest podział całego czasu trwania obserwacji na krótsze przedziały czasowe. Dla każdego przedziału, na podstawie obserwacji wszystkich osób obserwowanych przez co najmniej ten przedział, obliczane jest prawdopodobieństwo, że w przedziale tym wystąpi zdarzenie terminalne. Następnie na podstawie prawdopodobieństw oszacowanych dla każdego z przedziałów szacowane jest ogólne prawdopodobieństwo wystąpienia zdarzenia w różnych punktach czasowych.

**Przykład.** Czy nowy plaster nikotynowy pomaga w rzuceniu palenia skuteczniej niż tradycyjne plastry? Można przeprowadzić badanie z udziałem dwóch grup palaczy — jednej korzystającej z tradycyjnej terapii i drugiej korzystającej z metody eksperymentalnej. Zbudowanie tabel trwania życia na podstawie zebranych danych umożliwiłoby porównanie ogólnych wskaźników abstynencji między dwiema grupami i określenie, czy terapia eksperymentalna jest pod tym względem lepsza od tradycyjnej. Można także wykreślić funkcje przeżycia lub hazardu i porównać je wzrokowo, by uzyskać bardziej szczegółowe informacje.

**Statystyki.** Liczba wprowadzanych, liczba opuszczających, liczba narażonych na ryzyko, liczba zdarzeń terminalnych, proporcja kończących, proporcja przeżywających, skumulowana proporcja przeżywających (i błąd standardowy), gęstość prawdopodobieństwa (i błąd standardowy) oraz wskaźnik hazardu (i błąd standardowy) dla każdego przedziału dla każdej grupy; mediana czasu przeżycia dla każdej grupy; i test Wilcozona (Gehana) porównujący rozkłady przeżycia między grupami. Wykresy: wykresy funkcji przeżycia, logarytmu przeżycia, gęstości, wskaźnika hazardu i jeden minus przeżycie.

Wymagania dotyczące danych dla tabel trwania życia

**Dane.** Zmienna czasu powinna być ilościowa. Zmienna statusu powinna być dychotomiczna lub jakościowa, zakodowana jako wartości całkowite, przy czym zdarzenia powinny być kodowane jako pojedyncze wartości lub zakresy kolejnych wartości. Zmienne czynników powinny być jakościowe i zakodowane jako wartości całkowite.

**Założenia.** Prawdopodobieństwo interesującego zdarzenia powinno zależeć tylko od czasu, po zajściu zdarzenia początkowego — zakłada się, że ewentualny wpływ czasu bezwzględnego jest stabilny.



Oznacza to, że obserwacje wprowadzane do badania w różnych momentach (np. pacjenci rozpoczynający leczenie w różnych terminach) powinny zachowywać się podobnie. Ponadto nie powinny występować systematyczne różnice między obserwacjami ocenianymi i nieocenianymi. Jeśli, na przykład, wiele obserwacji ocenianych odzwierciedla pacjentów, u których choroba ma cięższy przebieg niż u pozostałych, wyniki mogą być obciążone.

**Procedury pokrewne.** W procedurze Tabele trwania życia stosowane jest podejście aktuariale do tego rodzaju analizy (ogólnie nazywanej analizą przeżycia). W procedurze Analiza przeżycia Kaplana-Meiera stosowana jest nieco inna metoda obliczania tabel trwania życia, w której łączny czas obserwacji nie jest dzielony na mniejsze przedziały. Ta metoda jest zalecana, jeśli istnieje mniejsza liczba obserwacji, w związku z czym każdy przedział przeżycia zawierałby tylko niewielką ich liczbę. Jeśli podejrzewa się, że niektóre zmienne są powiązane z czasem przeżycia lub chce się przeprowadzić kontrolę (współzmienną) względem niektórych zmiennych, należy użyć procedury Regresja Coxa. Jeśli dla tej samej obserwacji współzmienną mogą mieć różne wartości w różnych punktach czasowych, należy użyć regresji Coxa ze współzmiennymi zależnymi od czasu.

Tworzenie tabel trwania życia

1. Wybierz z menu następującą opcję:

**Analiza > Analiza przeżycia > Tabele trwania życia...**

2. Wybierz jedną *liczbową* zmienną przeżycia.

3. Określ przedziały czasowe do badania.

4. Wybierz zmienną statusu, aby zdefiniować obserwacje, dla których zaszło zdarzenie terminalne.

5. Kliknij przycisk **Zdefiniuj zdarzenie**, aby określić wartość zmiennej statusu oznaczającą, że zdarzenie wystąpiło.

Opcjonalnie można wybrać zmienną czynnika pierwszego rzędu. Dla każdej kategorii zmiennej czynnika generowane są aktuariale tabele zmiennej przeżycia.

Można także wybrać zmienną *przez czynnik* drugiego rzędu. Dla każdej kombinacji zmiennej czynnika pierwszego i drugiego rzędu generowane są aktuariale tabele zmiennej przeżycia.

## Tabele trwania życia: Definiuj zakres

Obserwacje z wartościami zmiennej czynnikowej w podanym zakresie będą uwzględnione w analizie, a dla każdej unikalnej wartości w zakresie zostaną wygenerowane osobne tabele (i wykresy, jeśli ich zażądano).

## Opcja tabel trwania życia

Można wpływać na różne aspekty analizy tabel trwania życia.

**Tabele trwania życia.** Aby w wynikach nie były widoczne tabele trwania życia, anuluj zaznaczenie opcji **Tabele trwania życia**.

**Wykres.** Umożliwia utworzenie wykresów funkcji przeżycia. Jeśli zdefiniowano zmienne czynniki, to dla każdej podgrupy zdefiniowanej przez te zmienne wygenerowane zostaną wykresy. Dostępne są wykresy przeżycia, logarytmu przeżycia, hazardu, gęstości i jeden minus przeżycie.

- *Przeżycie.* Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.
- *Czas przeżycia dziennika.* Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali logarytmicznej.
- *Zagrożenie.* Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.
- *Gęstość.* Umożliwia wyświetlanie funkcji gęstości.
- *Jeden minus przeżycie.* Umożliwia wykreślenie funkcji Jeden minus przeżycie na skali liniowej.

**Porównanie poziomów pierwszego czynnika.** Jeśli istnieje zmienna kontrolna pierwszego rzędu, można wybrać jedną z opcji w tej grupie, aby wykonać test Wilcozona (Gehana), który porównuje przeżycia podgrup. Testy wykonywane są na czynniku pierwszego rzędu. Jeśli zdefiniowano czynnik drugiego rzędu, testy wykonywane są dla każdego poziomu zmiennej drugiego rzędu.

## Dodatkowe właściwości komendy SURVIVAL

Język składni komend umożliwia również:

- Określenie więcej niż jednej zmiennej zależnej.
- Określenie nierówno rozmieszczonych przedziałów.
- Określenie więcej niż jednej zmiennej statusu.
- Określanie porównań, które nie uwzględniają wszystkich czynników ani wszystkich zmiennych sterujących.
- Obliczanie porównań przybliżonych zamiast dokładnych.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Modele parametrycznego czasu przyspieszonego niepowodzenia

Analiza modelu parametrycznego czasu przyspieszonego niepowodzenia (AFT) wywołuje procedurę parametrycznych modeli przeżycia z nierekurencyjnymi danymi czasu życia. Parametryczne modele przeżycia zakładają, że czas przeżycia jest zgodny ze znanym rozkładem, a ta analiza dopasowuje modele przyspieszonego czasu awarii do ich efektów modelu proporcjonalnych do czasu przeżycia.

### Uzyskiwanie analizy modeli parametrycznych przyspieszonego czasu awarii

1. Z menu wybierz:

**Analizuj > Przeżycie > Modele AFT (parametric Accelerated Failure Time)**

2. Wybierz zmienną źródłową.

#### Czas

##### Przeżycie

Pojedyncza zmienna numeryczna oznaczająca czas przeżycia.

##### Początek/koniec

Zmienne numeryczne oznaczające **Czas rozpoczęcia** i **Czas zakończenia**.

#### Status

Pojedyncza opcjonalna zmienna łańcuchowa lub numeryczna, która określa jedno z następujących ustawień statusu:

##### Niepowodzenie/zdarzenie

Odwzorowuje rekord na kategorię niepowodzeń/zdarzeń. Wartością domyślną zmiennej statusu łańcucha jest F.

##### Obciążenie prawostronne

Odwzorowuje rekord na odpowiednią kategorię cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest R.

##### Obciążenie lewostronne

Odwzorowuje rekord na lewą kategorię cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest L.

##### Obciążenie odstępu

Odwzorowuje rekord na kategorię cenzurowania interwałów. Tylko dla opcji **Początek/Koniec**. Wartością domyślną zmiennej statusu łańcucha jest I.

##### Przetwarzanie nieodwzorowanych wartości

Steruje kategorią, na którą mają zostać odwzorowane nieodwzorowane rekordy. Aby usunąć rekordy, których odwzorowanie nie powiodło się, wybierz opcję **Wyklucz je z analizy**.

W przypadku opcji **Survival** domyślny status dla wszystkich obserwacji to **Niepowodzenie/Zdarzenie**. Dla opcji **Początek/Koniec** domyślnym statusem jest **Cenzurowanie interwałowe**. Kliknij przycisk **Zdefiniuj zdarzenie**, aby zdefiniować zdarzenie dla zmiennej statusu.

### **Współzmiennie**

Jedna lub więcej opcjonalnych zmiennych numerycznych, które mają być traktowane jako współzmiennie. Należy zauważyć, że zmienna nie może być określona jednocześnie przez **Współzmiennie** i **Czynniki stałe**.

### **Czynniki stałe**

Co najmniej jedna zmienna opcjonalna, która ma być traktowana jako czynnik. Zmienna nie może być określona jednocześnie przez **czynniki stałe** i **współzmiennie**.

### **Obcinanie lewostronne**

Pojedyncza opcjonalna zmienna numeryczna dla lewego obciążenia tylko dla **Survival**.

## **Modele parametrycznego czasu przyspieszonego niepowodzenia: kryteria**

### **Kryteria**

Opcjonalny panel służący do określania kryteriów ogólnych.

#### **Przedział ufności**

Opcjonalna wartość procentowa określająca poziom przedziałów ufności parametrów regresji. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 100. Domyślną wartością jest 95.

#### **Braki danych**

Opcja sterująca sposobem traktowania braków danych zdefiniowanych przez użytkownika:

##### **Wyklucz systemowe i zdefiniowane przez użytkownika braki danych**

Traktuje braki danych zdefiniowane przez użytkownika jako poprawne wartości. Jest to ustawienie domyślne.

##### **Zdefiniowane braki danych są traktowane jako ważne wartości**

Ignoruje oznaczenia braków danych zdefiniowanych przez użytkownika i traktuje je jako poprawne wartości.

#### **Traktowanie statusu**

Tylko dla opcji **Początek/Koniec**. Opcja sterująca sposobem postępowania z rekordami z niepoprawnymi polami statusu:

##### **Odrzuć rekord powodujący konflikt**

Usuwa rekordy powodujące konflikt. Jest to ustawienie domyślne.

##### **Uzyskaj informacje o czasie zgodnie ze statusem**

Pobiera informacje o czasie zgodnie ze statusem.

##### **Wylicz status zgodnie z informacjami o czasie**

Zmienia status zgodnie z informacjami o czasie.

## **Modele parametrycznego czasu przyspieszonego niepowodzenia: Model**

### **operacyjny**

Opcjonalny panel służący do określania opcji i ustawień modelu.

#### **Rozkład czasu przeżycia**

Opcja służąca do określania rozkładu czasu przeżycia.

##### **Weibulla**

Określa rozkład Weibulla. Jest to ustawienie domyślne.

##### **Wykładniczy**

Określa rozkład wykładniczy.

##### **Log-normalny**

Określa rozkład logarytmiczno-normalny.

##### **Log-logistyczny**

Określa rozkład logarytmicznie logistyczny.

### **Ustawienia współzmiennych**

Określ współzmiennie.

### **Ustawienia czynników**

Określ zmienne czynnika.

### **Początkowa wartość przechwyty**

Opcja określająca początkową wartość wyrazu wolnego. Jeśli jest określona, musi być pojedynczą wartością numeryczną i nie może być równa 0.

### **Początkowa wartość parametru skali**

Opcja sterująca ustawieniem parametru skali.

### **Błąd standardowy odpowiedniej regresji OLS**

Używa błędu standardowego odpowiedniej regresji metodą najmniejszych kwadratów jako wartości początkowej.

### **Odwraca błąd standardowy odpowiedniej regresji OLS**

Używa odwrotności błędu standardowego.

### **Wartość podana przez użytkownika**

Jeśli zostanie podana pojedyncza wartość liczbowa, zostanie ona użyta jako wartość początkowa. Jeśli jest określona, musi być większa od 0.

## **Modele parametrycznego czasu przyspieszonego niepowodzenia: Oszacowanie**

### **Oszacowanie**

Opcjonalny panel służący do określania ustawień sterujących szacowaniem modeli przyspieszonego czasu awarii i opcjonalnego procesu wyboru predyktorów.

### **Metoda kierunku naprzemiennego lub Mnożniki (ADMM)**

#### **Uproszczone**

Stosuje metodę szybkiego naprzemiennego kierunku mnożników (ADMM). Jest to ustawienie domyślne.

#### **Tradycyjny**

Stosuje tradycyjny algorytm ADMM.

### **Zastosuj regularyzację L-1**

Przeprowadza proces w celu sterowania wyborem funkcji. Pole **Parametr kary** określa parametr kary, który steruje procesem regularyzacji. Musi to być pojedyncza wartość większa od 0. Ustawieniem domyślnym jest 0.001.

### **Modelowanie kryteriów zbieżności**

#### **Zbieżność parametru:**

Określa kryteria zbieżności dla parametru. Musi to być pojedyncza wartość liczbowa należąca do  $[0, 1)$ . Ustawieniem domyślnym jest 0,000001. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

#### **Zbieżność funkcji celu**

Określa kryteria zbieżności dla funkcji celu. Musi to być pojedyncza wartość liczbowa należąca do  $[0, 1)$ . Ustawieniem domyślnym jest 0, co nie powoduje zastosowania kryteriów zbieżności. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

#### **Zbieżność Hessego**

Określa kryteria zbieżności dla macierzy Hessego. Musi to być pojedyncza wartość liczbowa należąca do  $[0, 1)$ . Ustawieniem domyślnym jest 0, co nie powoduje zastosowania kryteriów zbieżności. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną

zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalne pole **Wartość** określa słowo kluczowe.

### **Kryteria zbieżności resztowej**

Opcja sterująca procesem optymalizacji.

#### **Reszty pierwotne i dualne**

Stosuje zarówno pierwotne, jak i podwójne kryterium zbieżności reszt. Jest to ustawienie domyślne.

#### **Tylko reszty pierwotne**

Stosuje pierwotne kryterium zbieżności reszt.

#### **Tylko reszty dualne**

Stosuje podwójne kryterium zbieżności reszt.

### **Metoda**

Opcjonalny parametr służący do określania metody estymacji.

#### **Automatycznie**

Automatycznie wybiera metodę na podstawie przykładowego zestawu danych. Jest to ustawienie domyślne. Zmienna **Liczba progowa predyktorów** określa próg liczby predyktorów i musi być jedną liczbą całkowitą większą od 1. Wartość domyślna: 1000.

#### **Newton-Raphson**

Stosuje metodę Newtona-Raphsona.

#### **L-BFGS**

Stosuje algorytm BFGS z ograniczoną pamięcią. Pole **Aktualizacja** określa liczbę wcześniejszych aktualizacji obsługiwanych przez algorytm BFGS z ograniczoną pamięcią i musi być jedną liczbą całkowitą większą lub równą 1. Wartość domyślna to 5.

### **Iteracja**

#### **Maksymalna liczba iteracji**

Określa maksymalną liczbę iteracji. Musi to być pojedyncza liczba całkowita należąca do [ 1, 100]. Ustawieniem domyślnym jest 20.

#### **Maksymalna liczba kroków połowienia**

Określa maksymalną liczbę kroków połowienia. Musi to być pojedyncza liczba całkowita należąca do [ 1, 20]. Ustawienie domyślne to 5.

#### **Maksymalna liczba wyszukiwań liniowych**

Określa maksymalną liczbę wyszukiwań liniowych. Musi to być pojedyncza liczba całkowita należąca do [ 1, 100]. Ustawieniem domyślnym jest 20.

#### **Bezwzględna zbieżność dla procesu iteracji**

Określa bezwzględną zbieżność procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowa należąca do (0, 1). Ustawieniem domyślnym jest 0.0001.

#### **Względna zbieżność dla procesu iteracji**

Określa względną zbieżność procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowa należąca do (0, 1). Ustawieniem domyślnym jest 0.01.

## **Modele parametrycznego czasu niepowodzenia akceleracji: Drukuj**

### **Drukuj**

Opcjonalny panel do sterowania danymi wyjściowymi tabeli.

#### **Szczegóły kodowania czynników**

Wybranie tej opcji powoduje wyświetlenie i wydrukowanie szczegółów kodowania czynników. Proces jest ignorowany, jeśli nie ma żadnych czynników.

#### **Wartości początkowe przypisane do parametrów regresji**

Jeśli ta opcja jest wybrana, wyświetlane są wartości początkowe używane w procesie estymacji.

### Historia iteracji modelu

Jeśli ta opcja jest wybrana, wyświetlana jest historia iteracji analizy przeżycia. W polu **Liczba kroków** podaj liczbę kroków z zakresu od 1 do 99999999. Ustawieniem domyślnym jest 1.

### Wyniki wyboru zawierające

Steruje wyświetlaniem szczegółów wyboru predyktora.

#### Zarówno wybrane, jak i niewybrane zmienne

Wyświetl w tabeli zarówno wybrane, jak i niewybrane zmienne.

#### Tylko wybrane zmienne

Wyświetla tylko wybrane zmienne.

#### Tylko niewybrane zmienne

Wyświetl tylko niewybrane zmienne. Pole **Maksymalna liczba zmiennych do wyświetlenia** określa maksymalną liczbę zmiennych wyświetlanych w tabeli. Ustawienie domyślne to 30.

## Modele parametrycznego przyspieszonego czasu awarii: przewidywanie

### Predykcja

Opcjonalny panel służący do oceniania i zapisywania przewidywanych statystyk w aktywnym zestawie danych.

#### Wartości czasu do oceny

##### Wartości czasu zdefiniowane przez zmienne zależne

Ocenia **Predykcje** na podstawie zmiennej czasu określonej dla parametrycznego modelu przeżycia.

##### Przedziały regularne

Ocenia **Predykcje** na podstawie przyszłych wartości czasu. Pole **Przedział czasu** określa przedział czasu i musi być pojedynczą wartością liczbową większą od 0. Pole **Liczba okresów** określa liczbę okresów i musi być jedną liczbą całkowitą z zakresu od 2 do 100.

##### Przedział czasu

Ocenia **Predykcje** na podstawie przedziału czasu w celu zdefiniowania przyszłych wartości czasu. Musi to być pojedyncza zmienna numeryczna.

### Predykcje

#### Przeżycie

Ocenia i zapisuje przewidywane statystyki przeżycia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredSurvival.

#### Hazard

Ocenia i zapisuje przewidywane zagrożenia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredHazard.

#### Skumulowane zagrożenie

Ocenia i zapisuje przewidywane skumulowane zagrożenia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredCumHazard.

#### Przeżycie warunkowe

Ocenia i zapisuje przewidywane statystyki przeżycia warunkowego w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredConditionalSurvival. Jeśli parametr PASTTIME nie zostanie określony, proces zostanie zignorowany. Wymagana jest wartość **przeszłego czasu przeżycia** i określa przeszłe wartości czasu dla oceniania. Musi to być pojedyncza zmienna numeryczna.

## Parametryczne modele przyspieszonego czasu awarii: wykres

### Wykres

#### Wykres reszt Coxa-Snella

Wybierz opcję **Wyświetl wykres**, aby utworzyć wykres reszt Coxa-Snella. W polu **Liczba punktów podziału** określ liczbę od 1 do 10000. Ustawieniem domyślnym jest 100.

#### Wykresy funkcji

Opcja sterująca wykresami funkcji.

##### Typ

###### Przeżycie

Tworzy wykres dla funkcji przeżycia.

###### Hazard

Tworzy wykres dla funkcji hazardu.

###### Gęstość

Tworzy wykres dla funkcji gęstości.

#### Liczba punktów do wyświetlenia

Określa liczbę punktów funkcji z zakresu od 1 do 200. Ustawieniem domyślnym jest 100.

#### Wartości współzmiennych do wykresu

Opcjonalne, aby określić wartości podane przez użytkownika i przypisać je do predyktorów. Domyślnie wyznaczone wykresy zostaną utworzone przy wartości **Średnia** dla każdej uwzględnionej współzmiennnej i częstości kategorii każdego uwzględnionego czynnika. Jeśli określono, wyznaczone wykresy zostaną utworzone na podstawie ustawień wzorca. W przypadku występowania zduplikowanych zmiennych zostanie rozpoznana pierwsza określona zmienna, a pozostałe zmienne zostaną zignorowane. Poprawna zmienna musi być zawarta w efekcie modelu. W przypadku współzmiennnej wartość podana przez użytkownika musi być wartością liczbową. Pominięcie zmiennej w efekcie wskazuje, że częstotliwość kategorii i **Średnia** będą używane domyślnie odpowiednio dla czynnika i współzmiennnej. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

#### Wartości czynników do wykresu

Opcjonalne, aby określić wartości podane przez użytkownika i przypisać je do predyktorów. W przypadku występowania zduplikowanych zmiennych zostanie rozpoznana pierwsza określona zmienna, a pozostałe zmienne zostaną zignorowane. Poprawna zmienna musi być zawarta w efekcie modelu. Pominięcie zmiennej w efekcie wskazuje, że częstotliwość kategorii i średnia będą domyślnie używane odpowiednio dla czynnika i współzmiennnej. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

#### Oddzielne wiersze dla

Opcja określania zmiennej kategoryjnej, według której będą rysowane wykresy liniowe.

#### Maksymalna liczba linii na wykresie

Określa maksymalną liczbę linii na wykresie, jeśli określono opcję **Rozdziel linie dla**. Wartością domyślną jest 10.

## Parametryczne modele czasu przyspieszonego niepowodzenia: Eksportuj

### Eksportuj

Wybierz opcję **Eksportuj informacje o modelu do pliku XML**, aby zapisać informacje o modelu i parametrze w pliku PMML w celu przeprowadzenia oceny. Należy określić katalog i nazwę pliku PMML, który ma zostać zapisany.

## Przeżycie AFT Definiuj zdarzenia dla zmiennych statusu

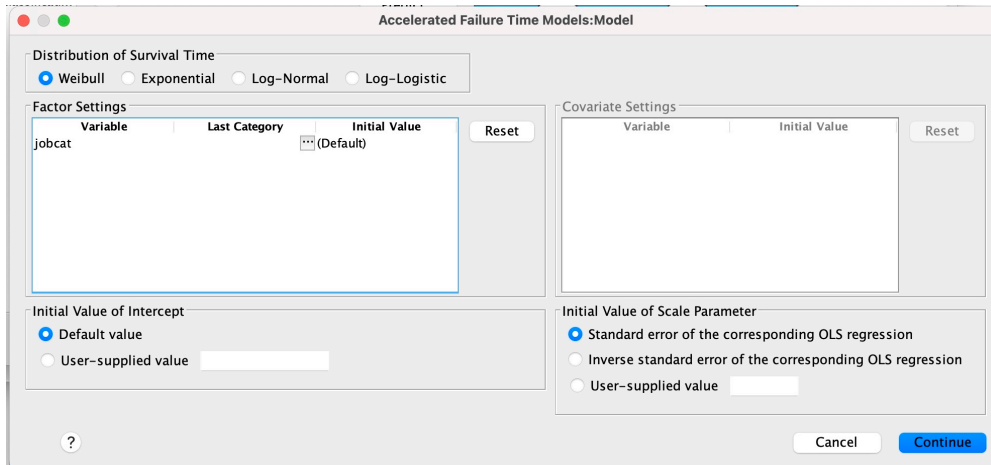
Wystąpienia wybranej wartości lub wartości dla zmiennej statusu wskazują, że w tych przypadkach wystąpiło zdarzenie końcowe. Wszystkie pozostałe obserwacje uznaje się za ocenzurowane. Wprowadź jedną wartość lub zakres wartości identyfikujący interesujące zdarzenie.

## Parametryczne modele przyspieszonego czasu awarii: wybierz kategorię

Ustawienie Wybierz kategorię udostępnia opcje umożliwiające wybór wartości oznaczającej kategorię, która ma być modelowana jako linia bazowa do porównania.

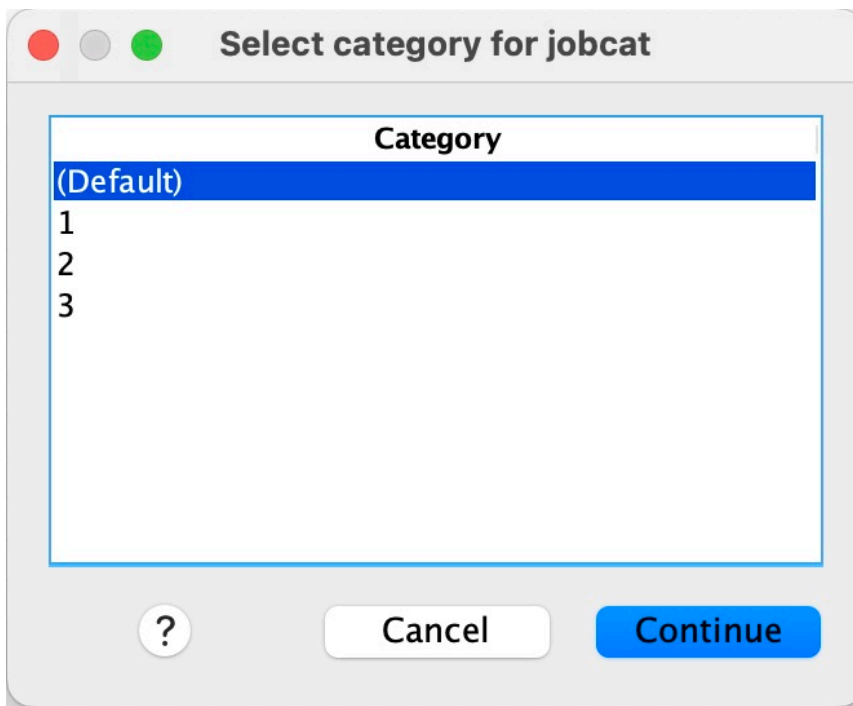
Wybieranie kategorii

Kliknij opcję 'Ostatnia kategoria', aby otworzyć okno dialogowe 'Wybierz kategorię'.



Rysunek 3. Modele przyspieszonego czasu życia-okno dialogowe-kategoria

Aby wyznaczyć kategorię jako linię bazową, wybierz wartość w oknie dialogowym 'Wybierz kategorię'.



Rysunek 4. Modele przyspieszonego czasu życia-okno dialogowe-wybierz kategorię

Kliknij przycisk Kontynuuj.



## Parametryczne współużytkowane modele Frailty

Analiza parametrycznych współużytkowanych modeli frailty przeżycia rozpoczyna procedurę parametrycznych modeli przeżycia z danymi wejściowymi czasu życia. Parametryczne modele przeżycia zakładają, że czas przeżycia jest zgodny ze znanym rozkładem, a analiza ta uwzględnia składnik frailty w parametrycznym modelu przeżycia. Jest on traktowany jako komponent losowy w celu uwzględnienia efektu nieobserwowanego z powodu zmienności poziomu indywidualnego lub grupy.

### Uzyskiwanie analizy parametrycznych współużytkowanych modeli Frailty

1. Z menu wybierz opcję:

**Analizuj > Survival > Parametryczne współużytkowane modele Frailty**

2. Wybierz zmienną źródłową.

#### Czas

##### Przeżycie

Czas przeżycia jest reprezentowany przez jedną zmienną oznaczającą czas zakończenia. Czas rozpoczęcia zostanie ustawiony na 0.

##### Początek/koniec

Zmienne numeryczne oznaczające **Czas rozpoczęcia** i **Czas zakończenia**.

#### Obiekt

Wymagane do uruchomienia procedury. Określa pojedynczą zmienną dla identyfikatora podmiotu.

#### Przedział

Określa pojedynczą i numeryczną zmienną dla numeru przedziału, który jest używany do identyfikowania różnych rekordów powtarzających się, które mają ten sam identyfikator obiektu.

#### Status

Pojedyncza opcjonalna zmienna łańcuchowa lub numeryczna, która określa jedno z następujących ustawień statusu:

##### Niepowodzenie/zdarzenie

Odwzorowuje rekord na kategorię niepowodzeń/zdarzeń. Wartością domyślną zmiennej statusu łańcucha jest F.

##### Obciążenie prawostronne

Odwzorowuje rekord na odpowiednią kategorię cenzurowania. Wartością domyślną zmiennej statusu łańcucha jest R.

##### Przetwarzanie nieodwzorowanych wartości

Steruje kategorią, na którą mają zostać odwzorowane nieodwzorowane rekordy. Aby usunąć rekordy, których odwzorowanie nie powiodło się, wybierz opcję **Wyklucz je z analizy**.

Kliknij przycisk **Zdefiniuj zdarzenie**, aby zdefiniować zdarzenie dla zmiennej statusu.

#### Współzmiennie

Jedna lub więcej opcjonalnych zmiennych numerycznych, które mają być traktowane jako współzmiennie. Należy zauważyć, że zmienna nie może być określona jednocześnie przez **Współzmiennie** i **Czynniki stałe**.

#### Czynniki stałe

Co najmniej jedna zmienna opcjonalna, która ma być traktowana jako czynnik. Zmienna nie może być określona jednocześnie przez **czynniki stałe** i **współzmiennie**.

## Parametryczne współużytkowane modele Frailty: Kryteria

### Kryteria

Opcjonalny panel służący do określania kryteriów ogólnych.

### **Przedział ufności**

Opcjonalna wartość procentowa określająca poziom przedziałów ufności parametrów regresji. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 100. Domyślną wartością jest 95.

### **Poziom istotności**

Opcja służąca do określania poziomu istotności testu ilorazu wiarygodności dla składnika frailty. Musi to być pojedyncza wartość liczbowa z zakresu od 0 do 1. Ustawieniem domyślnym jest 0.05.

### **Braki danych**

Opcja sterująca sposobem traktowania braków danych zdefiniowanych przez użytkownika:

#### **Wyklucz systemowe i zdefiniowane przez użytkownika braki danych**

Traktuje braki danych zdefiniowane przez użytkownika jako poprawne wartości. Jest to ustawienie domyślne.

#### **Zdefiniowane braki danych są traktowane jako ważne wartości**

Ignoruje oznaczenia braków danych zdefiniowanych przez użytkownika i traktuje je jako poprawne wartości.

### **Traktowanie odstępu czasu**

Opcja sterująca sposobem postępowania z rekordami, których interwał powoduje konflikt z czasem rozpoczęcia i zakończenia. Ma ona zastosowanie, jeśli istnieją dwie zmienne czasu ze zmienną przedziału czasu określoną w głównym oknie dialogowym.

#### **Odrzuć rekordy z konfliktem**

Jeśli wartość przedziału czasu jest w konflikcie z czasem rozpoczęcia i zakończenia, odrzuca całe rekordy numeru seryjnego podmiotu. Jest to ustawienie domyślne.

#### **Wykryj wartości przedziałów w oparciu o czas rozpoczęcia i zakończenia**

Wykrywa wartość odstępu czasu od czasu rozpoczęcia i zakończenia.

## **Parametryczne współużytkowane modele frailty: Model**

### **operacyjny**

Opcjonalny panel służący do określania opcji i ustawień modelu.

#### **Rozkład czasu przeżycia**

Opcja służąca do określania rozkładu czasu przeżycia.

##### **Weibulla**

Określa rozkład Weibulla. Jest to ustawienie domyślne.

##### **Wykładniczy**

Określa rozkład wykładniczy.

##### **Log-normalny**

Określa rozkład log-normalny.

##### **Log-logistyczny**

Określa rozkład logarytmicznie logiczny.

#### **Ustawienia współzmiennych**

Określ współzmiennie.

#### **Ustawienia czynników**

Określ zmienne czynnika.

#### **Początkowa wartość przechwyty**

Opcja określająca początkową wartość wyrazu wolnego. Jeśli jest określona, musi być pojedynczą wartością numeryczną i nie może być równa 0.

#### **Początkowa wartość parametru skali**

Opcja sterująca ustawieniem parametru skali.

**Błąd standardowy odpowiedniej regresji OLS**

Używa błędu standardowego odpowiedniej regresji metodą najmniejszych kwadratów jako wartości początkowej.

**Odwraca błąd standardowy odpowiedniej regresji OLS**

Używa odwrotności błędu standardowego.

**Wartość podana przez użytkownika**

Jeśli zostanie podana pojedyncza wartość liczbowa, zostanie ona użyta jako wartość początkowa. Jeśli jest określona, musi być większa od 0.

**Komponent słabości**

Opcjonalny parametr służący do określania **rozkładu** komponentu frailty.

**Gamma**

Określa rozkład gamma. Jest to ustawienie domyślne.

**Odwrotność rozkładu gaussowskiego**

Określa odwrotny rozkład Gaussa.

**Początkowa wartość wariancji**

Określa początkową wartość wariancji składnika frailty. Musi to być pojedyncza wartość liczbowa większa od 0. Wartością domyślną jest 1.0 dla rozkładu gamma i 0.1 dla rozkładu odwrotnego Gaussa.

## Parametryczne współużytkowane modele Frailty: Oszacowanie

**Oszacowanie**

Opcjonalny panel służący do określania ustawień, które sterują oszacowaniem współużytkowanych modeli frailty i opcjonalnym procesem wyboru predyktorów.

**Metoda kierunku naprzemiennego lub Mnożniki (ADMM)****Uproszczone**

Stosuje metodę szybkiego naprzemiennego kierunku mnożników (ADMM). Jest to ustawienie domyślne.

**Tradycyjny**

Stosuje tradycyjny algorytm ADMM.

**Zastosuj regularyzację L-1**

Przeprowadza proces w celu sterowania wyborem funkcji. Pole **Parametr kary** określa parametr kary, który steruje procesem regularyzacji. Musi to być pojedyncza wartość większa od 0. Ustawieniem domyślnym jest 0.001.

**Modelowanie kryteriów zbieżności****Zbieżność parametru:**

Określa kryteria zbieżności dla parametru. Musi to być pojedyncza wartość liczbowa należąca do  $[0, 1)$ . Ustawieniem domyślnym jest 0,000001. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna **Wartość** określa próg liczbowy dla typu zbieżności.

**Zbieżność funkcji celu**

Określa kryteria zbieżności dla funkcji celu. Musi to być pojedyncza wartość liczbowa, która należy do  $[0, 1)$ . Ustawieniem domyślnym jest 0, co nie powoduje zastosowania kryteriów zbieżności. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna **Wartość** określa próg liczbowy dla typu zbieżności.

**Zbieżność Hessego**

Określa kryteria zbieżności dla macierzy Hessego. Musi to być pojedyncza wartość liczbowa, która należy do  $[0, 1)$ . Ustawieniem domyślnym jest 0, co nie powoduje zastosowania kryteriów

zbieżności. W polu **Typ** można wybrać opcję **BEZWZGLĘDNE**, aby zastosować bezwzględną zbieżność do optymalizacji wewnętrznej, lub opcję **WZGLĘDNE**, aby zastosować względną zbieżność do optymalizacji wewnętrznej. Opcjonalna **Wartość** określa próg liczbowy dla typu zbieżności.

#### **Kryteria zbieżności resztowej**

Opcja sterująca procesem optymalizacji.

#### **Reszty pierwotne i dualne**

Stosuje zarówno pierwotne, jak i podwójne kryterium zbieżności reszt. To ustawienie jest domyślnie.

#### **Tylko reszty pierwotne**

Stosuje pierwotne kryterium zbieżności reszt.

#### **Tylko reszty dualne**

Stosuje podwójne kryterium zbieżności reszt.

#### **Metoda**

Opcjonalny parametr służący do określania metody estymacji.

#### **Automatycznie**

Automatycznie wybiera metodę na podstawie przykładowego zestawu danych. Ta metoda jest domyślnie wybrana. Zmienna **Liczba progowa predyktorów** określa próg liczby predyktorów i musi być jedną liczbą całkowitą większą od 1. Wartość domyślna: 1000.

#### **Newton-Raphson**

Stosuje metodę Newtona-Raphsona.

#### **L-BFGS**

Stosuje algorytm BFGS z ograniczoną pamięcią. Pole **Aktualizacja** określa liczbę poprzednich aktualizacji, które są obsługiwane przez algorytm BFGS z ograniczoną pamięcią i musi być jedną liczbą całkowitą większą lub równą 1. Wartość domyślna to 5.

#### **Iteracja**

##### **Maksymalna liczba iteracji**

Określa maksymalną liczbę iteracji. Musi to być pojedyncza liczba całkowita należąca do [ 1, 300]. Ustawieniem domyślnym jest 20.

##### **Maksymalna liczba kroków połowienia**

Określa maksymalną liczbę kroków połowienia. Musi to być pojedyncza liczba całkowita należąca do [ 1, 200]. Ustawienie domyślne to 5.

##### **Maksymalna liczba wyszukiwań liniowych**

Określa maksymalną liczbę wyszukiwań linii. Musi to być pojedyncza liczba całkowita należąca do [ 1, 300]. Ustawieniem domyślnym jest 20.

##### **Bezwzględna zbieżność dla procesu iteracji**

Określa bezwzględną zbieżność procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowo, która należy do (0, 1). Ustawieniem domyślnym jest 0.0001.

##### **Względna zbieżność dla procesu iteracji**

Określa względną zbieżność procesu iteracji zewnętrznej. Musi to być pojedyncza wartość liczbowo, która należy do (0, 1). Ustawieniem domyślnym jest 0.01.

## **Parametryczne współużytkowane modele frailty: Drukuj**

### **Drukuj**

Opcjonalny panel, który steruje danymi wyjściowymi tabeli.

#### **Szczegóły kodowania czynników**

Wybranie tej opcji powoduje wyświetlenie i wydrukowanie szczegółów kodowania czynników. Proces jest ignorowany, jeśli nie ma żadnych czynników.

### Wartości początkowe przypisane do parametrów regresji

Po wybraniu tej opcji wyświetlane są wartości początkowe, które są używane w procesie estymacji.

### Historia iteracji modelu

Jeśli ta opcja jest wybrana, wyświetlana jest historia iteracji analizy przeżycia. W polu **Liczba kroków** podaj liczbę kroków z zakresu od 1 do 99999999. Ustawieniem domyślnym jest 1.

## Parametryczne współzależne modele frailty: przewidywanie

### Przewidywanie

Opcjonalny panel służący do oceniania i zapisywania przewidywanych statystyk w aktywnym zestawie danych.

#### Wartości czasu do oceny

##### Wartości czasu zdefiniowane przez zmienne zależne

Ocenia **Przewidywanie** na podstawie zmiennej czasu określonej dla parametrycznego modelu przeżycia.

##### Przedziały regularne

Ocenia **Przewidywanie** na podstawie przyszłych wartości czasu. Pole **Przedział czasu** określa przedział czasu i musi być pojedynczą wartością liczbową większą od 0. Pole **Liczba okresów** określa liczbę okresów i musi być jedną liczbą całkowitą z zakresu od 2 do 100.

##### Przedział czasu

Ocenia **Przewidywanie** na podstawie przedziału czasu w celu zdefiniowania przyszłych wartości czasu. Musi to być pojedyncza zmienna numeryczna.

#### Przewidywanie

##### Przeżycie

Ocenia i zapisuje przewidywane statystyki przeżycia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredSurvival.

##### Hazard

Ocenia i zapisuje przewidywane zagrożenia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredHazard.

##### Skumulowane zagrożenie

Ocenia i zapisuje przewidywane skumulowane zagrożenia w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredCumHazard.

##### Przeżycie warunkowe

Ocenia i zapisuje przewidywane statystyki przeżycia warunkowego w aktywnym zbiorze danych. Domyślna nazwa zmiennej użytkownika (lub trzon nazwy) to PredConditionalSurvival. Proces jest ignorowany, jeśli nie określono parametru PASTTIME. Wymagana jest wartość **przeszłego czasu przeżycia** i określa przeszłe wartości czasu dla oceniania. Musi to być pojedyncza zmienna numeryczna.

##### Przeżycie bezwarunkowe

Ocenia i zapisuje przewidywane bezwarunkowe statystyki przeżycia w aktywnym zbiorze danych. Słowo kluczowe jest domyślnie pomijane. Jeśli jest określony, może po nim następować opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa główna) podana w nawiasach. Nazwa domyślna to PredUnCondSurvival.

##### Hazard bezwarunkowy

Ocenia i zapisuje przewidywane bezwarunkowe statystyki zagrożeń w aktywnym zbiorze danych. Słowo kluczowe jest domyślnie pomijane. Jeśli jest określony, może po nim następować opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa główna) podana w nawiasach. Nazwa domyślna to PredUncondHazard.

##### Bezwarunkowy skumulowany hazard

Ocenia i zapisuje przewidywane bezwarunkowe skumulowane statystyki hazardu w aktywnym zbiorze danych. Słowo kluczowe jest domyślnie pomijane. Jeśli jest określony, może po nim

następować opcjonalna nazwa zmiennej podana przez użytkownika (lub nazwa główna podana w nawiasach). Nazwa domyślna to PredUncondCumHazard.

## Parametryczne współużytkowane modele frailty: Wykres

### Wykres

#### Wykresy funkcji

Opcja sterująca wykresami funkcji.

#### Typ

##### Przeżycie

Tworzy wykres dla bezwarunkowych funkcji przeżycia.

##### Hazard

Tworzy wykres dla funkcji bezwarunkowego zagrożenia.

##### Gęstość

Tworzy wykres dla funkcji gęstości.

#### Liczba punktów do wyświetlenia

Określa liczbę punktów funkcji z zakresu od 1 do 200. Ustawieniem domyślnym jest 100.

#### Wartości współzmiennych do wykresu

Opcja umożliwiająca określenie wartości podanych przez użytkownika i przypisanie ich do predyktorów. Domyślnie wyznaczone wykresy będą tworzone dla średniej każdej współzmiennnej, która obowiązuje. Jeśli określono, wyznaczone wykresy zostaną utworzone na podstawie ustawień wzorca. W przypadku występowania zduplikowanych zmiennych zostanie rozpoznana pierwsza określona zmienna, a pozostałe zmienne zostaną zignorowane. Poprawna zmienna musi być zawarta w efekcie modelu. W przypadku współzmiennnej wartość podana przez użytkownika musi być wartością liczbową. Pominięcie zmiennej w efekcie wskazuje, że średnia będzie domyślnie używana dla współzmiennnej. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

#### Wartości czynników do wykresu

Opcja umożliwiająca określenie wartości podanych przez użytkownika i przypisanie ich do predyktorów. Domyślnie wyznaczone wykresy będą tworzone z częstotliwością kategorii dla każdego z uwzględnianych czynników. Jeśli określono, wyznaczone wykresy zostaną utworzone na podstawie ustawień wzorca. W przypadku występowania zduplikowanych zmiennych zostanie rozpoznana pierwsza określona zmienna, a pozostałe zmienne zostaną zignorowane. Poprawna zmienna musi być zawarta w efekcie modelu. Pominięcie zmiennej w efekcie wskazuje, że częstotliwość kategorii będzie domyślnie używana dla czynnika. Jeśli do zmiennej zostanie przypisana niepoprawna wartość, żądany wzorzec nie zostanie wykreślony.

#### Oddzielne wiersze dla

Opcja określania zmiennej kategoryjnej, według której będą rysowane wykresy liniowe.

#### Maksymalna liczba linii na wykresie

Określa maksymalną liczbę linii na wykresie, jeśli określono opcję **Rozdziel linie dla** .  
Wartością domyślną jest 10.

## Parametryczne współużytkowane modele Frailty: Eksportuj

### Eksportuj

Wybierz opcję **Eksportuj informacje o modelu do pliku XML** , aby zapisać informacje o modelu i parametrze w pliku PMML w celu przeprowadzenia oceny. Należy określić katalog i nazwę pliku PMML, który ma zostać zapisany.

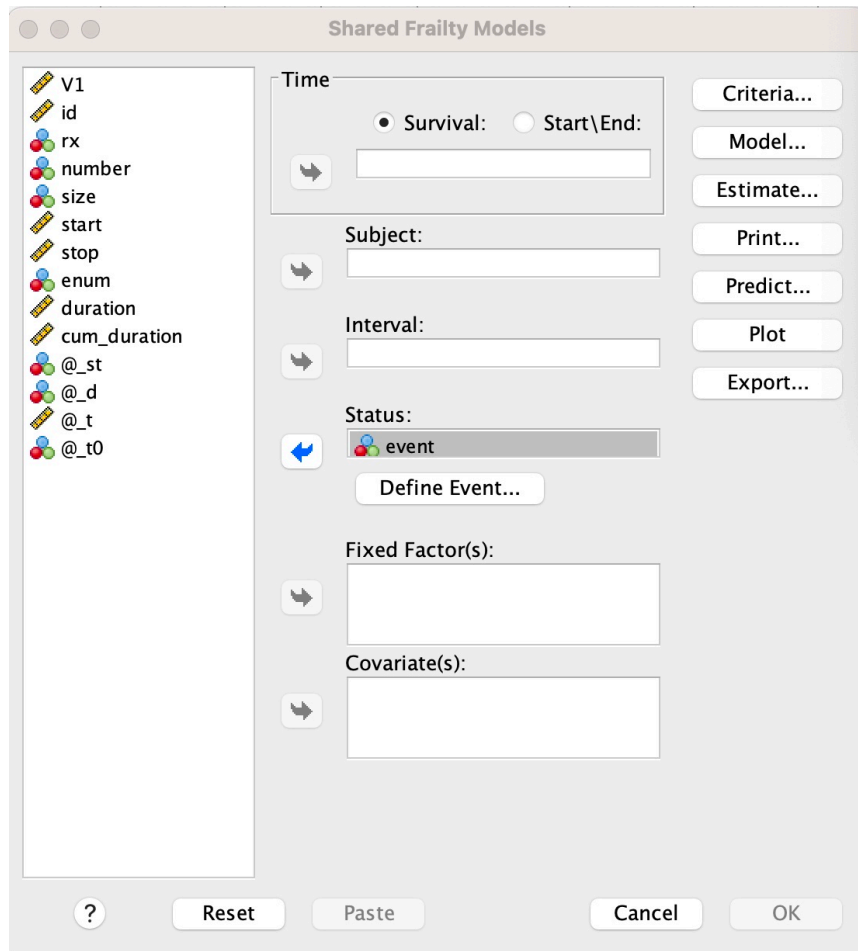
## Parametryczne współużytkowane modele frailty: Definiuj zdarzenia

Opcja definiowania statusu. Jeśli zmienna statusu zostanie pominięta, niepowodzenie lub zdarzenie stanie się domyślnym statusem dla wszystkich obserwacji.

1. Z menu wybierz:

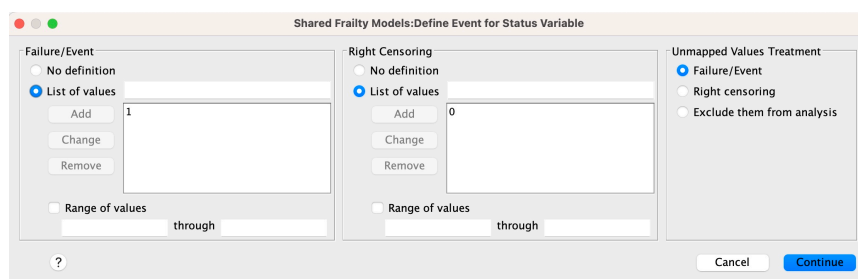
**Analizuj > Survival > Parametryczne współużytkowane modele Frailty ...**

2.



Rysunek 5. Współużytkowane modele Frailty-okno dialogowe-Status

3.



Rysunek 6. Współużytkowane modele Frailty-okno dialogowe-Status-Definiuj zdarzenie

## Parametryczne współużytkowane modele frailty-przykłady

### Przykład 1

```
SURVREG CURRENT Y WITH x1 BY x2
```

```
/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL.
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ .  
Czas przeżycia jest reprezentowany przez pojedynczą zmienną  $y$ .  
Obiekty są identyfikowane przez identyfikator zmiennej.  
Zakłada się, że czas przeżycia jest zgodny z rozkładem Weibulla.  
Zakłada się, że wariancja frailty jest zgodna z rozkładem Gamma.  
Wszystkie poprawne rekordy są używane w analizie przeżycia.

### **Przykład 2**

```
SURVREG CURRENT Y WITH x1 BY x2  
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ .  
Czas przeżycia jest reprezentowany przez dwie zmienne  $y_1$  i  $y_2$  oznaczające czas rozpoczęcia i zakończenia.  
Obiekty są identyfikowane przez identyfikator zmiennej.  
Zakłada się, że czas przeżycia jest zgodny z rozkładem logarytmiczno-normalnym.  
Przyjmuje się, że wariancja frailty jest zgodna z odwrotnym rozkładem Gaussa.  
Przedziały czasu są definiowane przez zmienną  $z$ . Dla każdego obiektu procedura używa tylko rekordów, które nie powodowały konfliktów, i wyklucza z analizy wszystkie rekordy po statusie pierwszego niepowodzenia.

### **Przykład 3**

```
SURVREG RECURRENT y1 y2 Z x1 BY x2(1)  
/MODEL SUBJECT = id FRAILTY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC  
/STATUS VARIABLE=zdarzenie FAILURE=1 RIGHT=0.
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ . Przeżycie czas jest reprezentowany przez dwie zmienne  $y_1$  i  $y_2$  oznaczające odpowiednio czas rozpoczęcia i czas zakończenia. Dla czynnika  $x_2$  kategoria "1" jest wyznaczana jako plan bazowy do modelowania.  
Obiekty są identyfikowane przez identyfikator zmiennej.  
Przyjmuje się, że czas przeżycia jest zgodny z rozkładem log-logistycznym.  
Przyjmuje się, że wariancja frailty jest zgodna z odwrotnym rozkładem Gaussa.  
Zdarzenie zmiennej jest określane w celu zdefiniowania statusu z odpowiednio 1 i 0 oznaczającymi niepowodzenie i cenzurowanie po prawej stronie.

### **Przykład 4**

```
SURVREG CURRENT Y WITH x1 BY x2  
/MODEL SUBJECT = id  
/STATUS VARIABLE=event FAILURE=1 RIGHT=0  
/PREDICT UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD  
/FUNCTIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ . Czas przeżycia jest reprezentowany przez pojedynczą zmienną  $y$ .



Obiekty są identyfikowane przez identyfikator zmiennej.

Bezwarunkowe lub oparte na populacji przeżycia, hazardu i skumulowanego hazardu są oceniane i zapisywane w aktywnym zbiorze danych.

Krzywe przeżycia i hazardu, bezwarunkowe lub oparte na populacji, są wykreślane oddzielnie dla kategorii w  $x_2$ .

### Przykład 5

```
SURVREG CURRENT Y WITH x1 BY x2
```

```
/MODEL SUBJECT = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABLE=event FAILURE=1 RIGHT=0
```

```
/ESTYMACJA HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE  
PENALTY=0.01.
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ . Czas przeżycia jest reprezentowany przez pojedynczą zmienną  $y$ .

Obiekty są identyfikowane przez identyfikator zmiennej.

Zakłada się, że czas przeżycia jest zgodny z rozkładem Weibulla.

Zakłada się, że wariancja frailty jest zgodna z rozkładem Gamma.

Kryteria zbieżności są oparte na macierzy Hessego. Jako zbieżność względna używana jest wartość  $1e-12$ .

Model zawiera składnik kary, który steruje procesem regularyzacji. Parametr kary ma wartość 0.01.

### Przykład 6

```
SURVREG CURRENT Y WITH x1 BY x2
```

```
/MODEL SUBJECT = id
```

```
/STATUS VARIABLE=infekcja FAILURE=1 RIGHT=0
```

```
/OSZACOWANIE MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

Parametryczny model przeżycia ze współzmienną  $x_1$  i czynnikiem  $x_2$  jest dopasowany do  $y$ . Czas przeżycia jest reprezentowany przez pojedynczą zmienną  $y$ .

Obiekty są identyfikowane przez identyfikator zmiennej.

Procedura określa maksymalną liczbę przeszukiwanych wierszy na 100, maksymalną liczbę iteracji na 50 i maksymalną liczbę kroków połowienia na 20.

## Parametryczne współużytkowane modele frailty-studium przypadku dla danych okresowych

Parametryczne współużytkowane modele frailty-studium przypadku dla danych okresowych

Nazwa przypadku użycia-efekt uboczny leczenia.

Aktorzy-badacz i praktykant w dziedzinie zdrowia publicznego.

Warunki wstępne-wyczyszczony zestaw danych dostępny na podstawie czasu przeżycia, statusu efektu ubocznego i predyktorów, które mają zostać skorygowane.

Opis-Patrick, badacz zdrowia publicznego, bada próbkę danych, która obejmuje 20 uczestników.

Uczestnicy ci są rekrutowani do badania nad łagodnym efektem ubocznym, który może być spowodowany nowym leczeniem. Projektant leczenia twierdzi, że nie byłoby różnic między mężczyznami i kobietami,

jeśli chodzi o efekt uboczny. Patrick chciałby ocenić taką hipotezę. Zmienne, które są uwzględnione w próbie danych, są wymienione w następujący sposób:

- patID: numer identyfikatora identyfikujący unikalnego uczestnika.
- endTime: Czas przeżycia (w dniach) działania niepożądanego po leczeniu, który jest mierzony od początku leczenia do zgłaszanego lub cenzurowania działania niepożądanego w ciągu 60 dni.
- sideEffect: Status efektu ubocznego, status = 0, jeśli ocenzurowany i status = 1, jeśli zgłaszany jest łagodny efekt boczny.
- wiek: wiek uczestnika w okresie badań.
- samica: samica = 0 jeśli samiec i samica = samica = 1 jeśli samica.

Możliwe jest zastosowanie wielu metod leczenia, co powoduje powstanie wielu rekordów czasów powtarzania, które są mierzone dla określonego uczestnika. Czas rozpoczęcia jest zawsze równy 0 dla każdego rekordu, który jest pomijany w próbie danych. Patrick jest zainteresowany wizualizacją funkcji przeżycia i ryzyka w celu porównania płci męskiej i żeńskiej poprzez kontrolowanie ich wieku i kruchości. Jest świadomy, że te zabiegi, które są podawane temu samemu uczestnikowi, są bardziej skorelowane. Zakładając, że czas przeżycia jest zgodny z rozkładem Weibulla, Patrick postanawia zbudować parametryczny model przeżycia ze współdzieloną frailty w programie SPSS Statistics, aby uwzględnić zależność leczenia od tego samego uczestnika.

Składnia-

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILITY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

### Podsumowanie:

Składnia określona przez Patricka określa endTime jako pojedynczą zmienną czasu. Procedura automatycznie zakłada, że czas rozpoczęcia dla każdego rekordu wynosi 0. Zmienne age i female są

modelowane odpowiednio jako współzmienna i czynnik. Zakłada się, że czas przeżycia dla nawrotu jest zgodny z rozkładem Weibulla. Zakłada się, że nieobstrużony składnik frailty jest zgodny z rozkładem Gamma, a jego składnik wariancji jest modelowany. Jeśli chodzi o wyniki, tabela Podsumowanie modelu zawiera informacje o procedurze i modelu. Tabela Podsumowanie przetwarzania przypadków zawiera obszerny wykaz statusu niepowodzenia/cenzurowania, a także przypadków wykluczonych z analizy.

W próbie danych Patricka wszystkie rekordy są poprawne i uwzględnione w analizie. Porównując logarytm wiarygodności z modelem odpowiadającym bez komponentu frailty, model shared-frailty nie osiąga istotnego poziomu ( $p$ -value = 0.168). Patrick zastanawia się, czy konieczne jest uwzględnienie w modelu terminu "shared-frailty". Szacowany współczynnik przyspieszenia uczestnika płci męskiej wynosi 1.017, co uzyskuje się przez obliczenie wykładnika szacowanego współczynnika regresji 0.017 dla [kobieta = 0.0]. Powiązany 95% przedział ufności to (0,688, 1.504). Wyniki te sugerują, że samiec ma prawie taki sam współczynnik przyspieszenia jak samica z tym samym wiekiem i kruchością. Na poziomie populacji Patryk wykreśla bezwarunkowe krzywe przeżycia i hazardu oddzielnie dla mężczyzn i kobiet, które są oceniane na podstawie średniej wieku z próby (37.45 roku życia).

Patrick potwierdza, że dla każdej stałej wartości czasu przeżycia, mężczyzna i kobieta średnio mają takie samo prawdopodobieństwo przeżycia. Co ciekawe, pomimo jednomodalnego kształtu, który jest pokazany na bezwarunkowym wykresie zagrożeń, Patrick odkrywa, że w ciągu 60 dni zagrożenie populacją wzrasta. Takie zachowanie może oznaczać istnienie słabego efektu. Aby dokładniej zbadać efekt uboczny wywołany przez leczenie, Patrick może kontynuować pracę z modelem bez składnika kruchości i porównać zachowanie mężczyzn i kobiet. Ponadto może on rozważyć możliwość współpracy z uczestnikami przez okres dłuższy niż 60 dni w celu zebrania większej ilości danych.

## Analiza przeżycia Kaplana-Meiera

W wielu sytuacjach celowe jest zbadanie rozkładu czasów, jakie upływają między dwoma zdarzeniami, np. długości zatrudnienia (czasu między rozpoczęciem a zakończeniem pracy w firmie). Jednak dane tego rodzaju zawierają zwykle obserwacje ocenzurowane. Obserwacje ocenzurowane to takie, dla których drugie zdarzenie nie zostało zarejestrowane (np. osoby nadal pracujące dla firmy w chwili zakończenia badania). Procedura Kaplana-Meiera jest metodą szacowania modeli czasu do wystąpienia zdarzenia w obecności obserwacji ocenzurowanych. Model Kaplana-Meiera oparty jest na estymacji prawdopodobieństwa warunkowego w każdym punkcie czasowym wystąpienia zdarzenia i estymacji wskaźnika przeżycia w każdym punkcie czasowym na podstawie granicznego estymatora iloczynowego tych prawdopodobieństw.

**Przykład.** Czy nowa metoda leczenia AIDS przedłuża życie pacjentów? Można przeprowadzić badanie z udziałem dwóch grup pacjentów chorych na AIDS: jednej leczonej metodami tradycyjnymi i drugiej leczonej metodą eksperymentalną. Zbudowanie na podstawie danych modelu Kaplana-Meiera umożliwiłoby porównanie ogólnych wskaźników przeżycia między dwiema grupami i określenie, czy terapia eksperymentalna jest pod tym względem lepsza od tradycyjnej. Można także wykreślić funkcje przeżycia lub hazardu i porównać je wzrokowo, by uzyskać bardziej szczegółowe informacje.

**Statystyki.** Tabela przeżycia, zawierająca czas, status, skumulowane przeżycia i błąd standardowy, skumulowane zdarzenia i pozostałą liczbę; a także średnią i medianę czasu przeżycia z błędem standardowym i 95% przedziałem ufności. Wykresy: przeżycie, hazard, logarytm przeżycia i jeden minus przeżycie

Wymagania dotyczące danych w procedurze Kaplana-Meiera

**Dane.** Zmienna czasu powinna być ciągła, zmienna statusu może być jakościowa lub ciągła, a zmienne czynników i warstwujące powinny być jakościowe.

**Założenia.** Prawdopodobieństwo interesującego zdarzenia powinno zależeć tylko od czasu, po zajściu zdarzenia początkowego — zakłada się, że ewentualny wpływ czasu bezwzględnego jest stabilny. Oznacza to, że obserwacje wprowadzane do badania w różnych momentach (np. pacjenci rozpoczynający leczenie w różnych terminach) powinny zachowywać się podobnie. Ponadto nie powinny występować systematyczne różnice między obserwacjami ocenzurowanymi i nieocenzurowanymi. Jeśli, na przykład, wiele obserwacji ocenzurowanych odzwierciedla pacjentów, u których choroba ma cięższy przebieg niż u pozostałych, wyniki mogą być obciążone.

**Procedury pokrewne.** W procedurze Kaplana-Meiera stosowana jest metoda obliczania tabel trwania życia, które estymują funkcję przeżycia lub hazardu w momencie zajścia każdego zdarzenia. W procedurze Tabele trwania życia stosowana jest aktuariałna metoda analizy przeżycia, która bazuje na podziale okresu obserwacji na mniejsze przedziały czasowe i może być użyteczna przy operowaniu na dużych próbach. Jeśli podejrzewa się, że niektóre zmienne są powiązane z czasem przeżycia lub chce się przeprowadzić kontrolę (współzmiennie) względem niektórych zmiennych, należy użyć procedury Regresja Coxa. Jeśli dla tej samej obserwacji współzmiennie mogą mieć różne wartości w różnych punktach czasowych, należy użyć regresji Coxa ze współzmiennymi zależnymi od czasu.

Uzyskiwanie analizy przeżycia Kaplana-Meiera

1. Z menu wybierz:

**Analiza > Analiza przeżycia > Kaplan-Meier...**

2. Wybierz zmienną czasu.

3. Wybierz zmienną statusu, aby zidentyfikować obserwacje, dla których zaszło zdarzenie terminalne. Może to być zmienna liczbowa lub *krótki tańcuch*. Następnie kliknij opcję **Zdefiniuj zdarzenie**.

Opcjonalnie można wybrać zmienną czynnika, aby przeanalizować różnice między grupami. Można także wybrać zmienną warstwową, by wygenerować odrębną analizę dla każdego poziomu (każdej warstwy) zmiennej.

## Kaplan-Meier: Definiuj zdarzenie dla zmiennej statusu

Wprowadź wartość lub wartości wskazujące na wystąpienie zdarzenia terminalnego. Można wprowadzić jedną wartość, zakres wartości lub listę wartości. Opcja Zakres wartości jest dostępna tylko wtedy, gdy zmienna statusu jest liczbowa.

## Kaplan-Meier: Porównaj poziomy czynnik

Można zlecić obliczenie statystyki testującej równość rozkładów przeżycia dla różnych poziomów czynnika. Dostępne statystyki to: test rang logarytmicznych, test Breslowa i test Tarone'a-Ware'a. Wybierz jedną z alternatyw, aby określić sposób wykonywania porównań: sumarycznie poprzez warstwy, dla każdej warstwy, parami poprzez warstwy albo parami dla każdej warstwy.

**Liniowy trend poziomów czynnika.** Umożliwia testowanie w poszukiwaniu trendu liniowego między poziomami czynnika. Ta opcja jest dostępna tylko dla porównań sumarycznych (a nie parami) poziomów czynników.

- *Ranga logarytmiczna.* Test służący do porównywania rozkładów przeżycia. Wszystkie punkty osi czasu posiadają w trakcie testu tę samą wagę.
- *Breslow.* Test służący do porównywania rozkładów przeżycia. Momenty czasowe są ważone liczbą obserwacji zagrożonych w każdym z nich.
- *Tarone-Ware.* Test służący do porównywania rozkładów przeżycia. Każdemu momentowi w czasie przypisana jest waga równa pierwiastkowi kwadratowemu z liczby obserwacji zagrożonych w danym momencie.
- *Sumarycznie poprzez warstwy.* Umożliwia porównanie wszystkich poziomów czynników w jednym teście w celu przetestowania równości krzywych przeżycia.
- *Parami przez warstwy.* Umożliwia porównanie każdej odrębnej pary poziomów czynników. Testy trendów parami są niedostępne.
- *Dla każdej warstwy.* Dla każdej warstwy wykonuje oddzielny test równości wszystkich poziomów czynników. W razie braku zmiennej stratyfikacyjnej testy nie są wykonywane.
- *Parami dla każdej warstwy.* Umożliwia porównanie każdej odrębnej pary poziomów czynników dla każdej warstwy. Testy trendów parami są niedostępne. W razie braku zmiennej stratyfikacyjnej testy nie są wykonywane.

## Kaplan-Meier: Zapisz zmienne wynikowe

Informacje z tabeli Kaplana-Meiera można zapisać jako nowe zmienne, które następnie można będzie wykorzystać w następnych analizach do testowania hipotez lub sprawdzania założeń. Jako nowe zmienne można zapisywać przeżycie, błąd standardowy przeżycia, hazard i skumulowane zdarzenia.

- *Przeżycie*. Szacowanie prawdopodobieństwa skumulowanego przeżycia. Domyślna nazwa zmiennej składa się z przedrostka `sur_` i dołączonego do niego numeru kolejnego. Jeśli na przykład zmienna `sur_1` już istnieje procedura Kaplana-Meiera przypisuje zmiennej nazwę `sur_2`.
- *Standardowy błąd przeżycia*. Błąd standardowy oszacowania skumulowanego przeżycia. Domyślna nazwa zmiennej składa się z przedrostka `se_` i dołączonego do niego numeru kolejnego. Jeśli na przykład zmienna `se_1` już istnieje, procedura Kaplana-Meiera przypisuje zmiennej nazwę `se_2`.
- *Zagrożenie*. Szacowanie funkcji skumulowanego hazardu. Domyślna nazwa zmiennej składa się z przedrostka `haz_` i dołączonego do niego numeru kolejnego. Jeśli na przykład zmienna `haz_1` już istnieje procedura Kaplana-Meiera przypisuje zmiennej nazwę `haz_2`.
- *Zdarzenia skumulowane*. Częstość skumulowana zdarzeń dla obserwacji posortowanych według ich czasów przeżycia i kodów stanu. Domyślna nazwa zmiennej składa się z przedrostka `cum_` i dołączonego do niego numeru kolejnego. Jeśli na przykład zmienna `cum_1` już istnieje procedura Kaplana-Meiera przypisuje zmiennej nazwę `cum_2`.

## Kaplan-Meier: Opcje

Dostępne są różne rodzaje wyników analizy Kaplana-Meiera.

**Statystyki.** Można wybrać statystyki, które mają być prezentowane dla obliczonych funkcji przeżycia, w tym tabele przeżycia, średnią i medianę przeżycia oraz kwartyle. Jeśli uwzględniono zmienne czynników, to dla każdej grupy generowane są odrębne statystyki.

**Wykresy.** Korzystając z wykresów, można wizualnie badać funkcję przeżycia, jeden minus przeżycie, funkcję hazardu i logarytmiczną funkcję przeżycia. Jeśli uwzględniono zmienne czynników, to funkcje są wykreślane dla każdej grupy z osobna.

- *Przeżycie*. Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.
- *Jeden minus przeżycie*. Umożliwia wykreślenie funkcji Jeden minus przeżycie na skali liniowej.
- *Zagrożenie*. Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.
- *Czas przeżycia dziennika*. Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali logarytmicznej.

## Dodatkowe właściwości komendy KM

Język składni komend umożliwia również:

- Uzyskiwanie tabel częstości, które uwzględniają utracone obserwacje w dalszej analizie jako kategorię osobną od obserwacji ocenianych.
- Określanie nierównego rozmieszczenia dla testu pod kątem trendu liniowego.
- Uzyskiwanie percentyli innych niż kwartyle dla zmiennej czasu przeżycia.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Analiza metodą regresji Coxa

Regresja Coxa tworzy model predykcyjny dla danych w czasie do zdarzenia. Model generuje funkcję przeżycia przewidującą prawdopodobieństwo, że zdarzenie będące przedmiotem zainteresowania wystąpiło w określonym czasie ( $t$ ) dla danych wartości zmiennych predyktora. Kształt funkcji przeżycia i współczynniki regresji dla predyktorów są szacowane na podstawie obserwowanych obiektów; model może być następnie zastosowany do nowych obserwacji, które zawierają pomiary dla zmiennych predyktora. Należy zauważyć, że informacje z ocenianych obiektów, czyli takich, które nie są przedmiotem zainteresowania podczas obserwacji, są przydatne podczas szacowania modelu.

**Przykład.** Czy u kobiet i mężczyzn ryzyko rozwoju raka płuca w wyniku palenia tytoniu jest takie samo? Budując model regresji Coxa, w którym współzmiennymi będzie liczba papierosów wypalanych dziennie i płeć, można testować hipotezy dotyczące wpływu płci i palenia papierosów na czas, po jakim nastąpi zachorowanie na raka płuca.

**Statystyki.** Dla każdego modelu:  $-2LL$ , statystyka ilorazu wiarygodności i ogólna wartość chi-kwadrat. Dla zmiennych w modelu: oszacowania parametrów, błędy standardowe i statystyki Walda. W przypadku zmiennych nienależących do modelu: statystyki oceny i reszta chi-kwadrat.

Wymagania dotyczące danych w regresji Coxa

**Dane.** Zmienna czasu powinna być ilościowa, ale zmienna statusu może być jakościowa lub ciągła. Zmienne niezależne (współzmiennie) mogą być ciągłe lub jakościowe; w przypadku zmiennych jakościowych wymagane jest kodowanie przy użyciu wartości sztucznych lub wskaźników (w procedurze dostępna jest opcja automatycznego przekodowywania zmiennych). Zmienne warstwujące powinny być jakościowe i zakodowane jako wartości całkowite lub krótkie łańcuchy.

**Założenia.** Obserwacje powinny być niezależne, a iloraz hazardu powinien być stały w czasie; tj. proporcja hazardu między obserwacjami nie powinna zmieniać się w czasie. To ostatnie założenie znane jest jako **założenie o proporcjonalności hazardu**.

**Procedury pokrewne.** Jeśli założenie o proporcjonalności hazardu nie jest spełnione (patrz wyżej), to konieczne może być użycie procedury Coxa ze współzmiennymi zależnymi od czasu. Jeśli nie ma współzmiennych lub istnieje tylko jedna współzmienna jakościowa, można użyć Tabel trwania życia lub procedury Kaplana-Meiera do przeanalizowania funkcji przeżycia lub hazardu w próbie. Jeśli próba nie zawiera danych ocenzonego (tj. w każdej obserwacji zaszło zdarzenie terminalne), można użyć procedury Regresja liniowa do zamodelowania relacji między predyktorami a czasem do wystąpienia zdarzenia.

Uzyskiwanie analizy metodą regresji Coxa

1. Wybierz z menu następującą opcję:

**Analizuj > Przeżycie > Regresja Coxa ...**

2. Wybierz zmienną czasu. Obserwacje z ujemnymi wartościami czasu nie są analizowane.

3. Wybierz zmienną statusu, a następnie kliknij przycisk **Zdefiniuj zdarzenie**.

4. Wybierz co najmniej jedną współzmienną. Aby uwzględnić terminy interakcji, wybierz wszystkie zmienne związane z interakcją, a następnie kliknij przycisk **>a\*b>**.

Opcjonalnie można obliczyć odrębne modele dla poszczególnych grup, definiując zmienną warstwującą.

## Regresja Coxa: Definiuj współzmiennie jakościowe

Można szczegółowo określić sposób traktowania zmiennych jakościowych przez procedurę regresji Coxa.

**Współzmiennie.** Lista wszystkich współzmiennych określonych w głównym oknie dialogowym albo bezpośrednio, albo jako części interakcji, na dowolnej warstwie. Jeśli niektóre z tych zmiennych są łańcuchowe lub jakościowe, można użyć ich tylko jako współzmiennych jakościowych.

**Współzmiennie jakościowe.** Lista zmiennych określonych jako jakościowe. Przy każdej zmiennej w nawiasach wskazano sposób kodowania kontrastów. Zmienne łańcuchowe (oznaczone symbolem < po nazwie) są już wymienione na liście współzmiennych jakościowych. Wybierz dowolne inne współzmiennie jakościowe z listy Współzmiennie i przenieś je na listę Współzmiennie jakościowe.

**Zmień kontrast.** Umożliwia zmianę metody kontrastu. Dostępne są następujące metody kontrastu:

- **Wskaźnik.** Kontrasty wskazują na obecność lub nieobecność kategorii. Kategoria odniesienia jest przedstawiona w macierzy kontrastów jako rząd zer.
- **Prosty.** Każda kategoria predyktora z wyjątkiem kategorii odniesienia jest porównywana z kategorią odniesienia.
- **Różnica.** Każda kategoria predyktora z wyjątkiem pierwszej kategorii jest porównywana ze średnim efektem poprzednich kategorii. Znane również jako odwrotne kontrasty Helmerta.

- **Helmerta.** Każda kategoria predyktora z wyjątkiem ostatniej kategorii jest porównywana ze średnim efektem następných kategorii.
- **Powtórzoney.** Każda kategoria predyktora z wyjątkiem pierwszej kategorii jest porównywana z kategorią poprzedzającą.
- **Wielomianowy.** Kontrasty wielomianowe ortogonalne. Zakłada się, że kategorie są równo rozłożone. Kontrasty wielomianowe są dostępne wyłącznie dla zmienných liczbowych.
- **Odchylenie.** Każda kategoria predyktora z wyjątkiem kategorii odniesienia jest porównywana z efektem łącznym.

W przypadku wybrania opcji **Odchylenie, Proste** lub **Wskaźnik**, jako kategorię odniesienia należy wybrać **Pierwsza** albo **Ostatnia**. Należy zauważyć, że metoda zmienia się dopiero po kliknięciu przycisku **Zmień**.

Współzmiennie łańcuchowe muszą być współzmiennymi jakościowymi. Aby usunąć zmienną łańcuchową z listy Współzmiennie jakościowe, należy usunąć wszystkie składniki zawierające tę zmienną z listy Współzmiennie w głównym oknie dialogowym.

## Wykresy regresji Coxa

Wykresy mogą okazać się pomocne w ewaluacji oszacowanego modelu i interpretacji wyników. Można wykreślić funkcję przeżycia, hazardu, log-minus-log i jeden-minus-przeżycie.

### Typ wykresu

#### Przeżycie

Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.

#### Zagrożenie

Umożliwia wyświetlenie funkcji skumulowanego przeżycia na skali liniowej.

#### Ln minus Ln

Względem oszacowania stosowane jest skumulowane oszacowanie przeżycia po transformacji  $\ln(-\ln)$ .

#### Jeden minus przeżycie

Umożliwia wykreślenie funkcji Jeden minus przeżycie na skali liniowej.

### Wartości współzmienných wykreślane dla

Można wykreślić osobną linię dla każdej wartości współzmiennęj jakościowej, przenosząc tę współzmienną dla pola tekstowego **Oddzielne linie dla**. Ta opcja jest dostępna tylko w przypadku współzmienných jakościowych, które oznaczone są przyrostkiem **(Cat)** po nazwie na liście **Wartości współzmienných wykreślane dla**.

### Zmień wartość

Ponieważ funkcje te zależą od wartości współzmienných, to aby wykreślić funkcje względem czasu, należy użyć stałych wartości w miejsce współzmienných. Domyślnie jako wartość stała używana jest średnia każdej współzmiennęj, ale można wprowadzić własne wartości dla wykresu, korzystając z grupy elementów sterujących **Zmień wartość**.

## Regresja Coxa: Zapisz zmienne wynikowe

Różne wyniki analizy można zapisywać jako nowe zmienne. Zmienných tych można następnie używać w późniejszych analizach do testowania hipotez lub sprawdzania założeń.

### Zapisz zmienne modelu

Umożliwia zapisanie funkcji przeżycia i jej błędu standardowego, oszacowań log-minus-log, funkcji niezawodności, reszt cząstkowych, wartości DfBeta dla regresji oraz  $X \cdot \text{Beta}$  predyktora liniowego jako nowych zmienných.

#### Funkcja niezawodności

Wartość funkcji skumulowanego przeżycia dla danego czasu. Jest to równe prawdopodobieństwu przeżycia do danego czasu.

#### Błąd standardowy funkcji niezawodności

Błąd standardowy oszacowania skumulowanego przeżycia.

### **Log minus log funkcji niezawodności**

Szacowanie skumulowanego przeżycia po zastosowaniu transformacji  $\ln(-\ln)$  do oszacowania.

### **Funkcja hazardu**

Umożliwia zapisywanie oszacowania funkcji skumulowanego hazardu (zwaną także resztą Coxa-Snella).

### **Reszty cząstkowe**

Można wykreślić reszty cząstkowe względem czasu przeżycia, aby przetestować założenie o proporcjonalności hazardu. Dla każdej współzmiennnej w modelu finalnym zapisywana jest jedna zmienna. Reszty cząstkowe są dostępne tylko dla modeli zawierających co najmniej jedną współzmienną.

### **DfBeta**

Szacowana zmiana współczynnika w przypadku usunięcia obserwacji. Dla każdej współzmiennnej w modelu finalnym zapisywana jest jedna zmienna. Wartości DfBeta są dostępne tylko dla modeli zawierających co najmniej jedną współzmienną.

### **X\*beta**

Wynik predyktora liniowego. Jest sumą iloczynu znormalizowanych wartości współzmiennnej oraz odpowiadających im ocen parametrów dla każdej z obserwacji.

**Uwaga:** Jeśli wykonywana jest analiza Coxa ze współzmiennymi zależnymi od czasu, zapisywane są tylko DfBeta.

### **Eksportuj informacje o modelu do pliku XML**

Oszacowania parametrów są eksportowane do określonego pliku w formacie XML. Możesz użyć tego pliku modelu do stosowania informacji o modelu do innych plików danych w celach statystycznych.

## **Regresja Coxa: Opcje**

Można wpływać na różne aspekty analizy i wyników.

**Statystyki modelu.** Możliwe jest uzyskanie statystyk parametrów modelu, w tym przedziałów ufności dla  $\exp(B)$  i korelacji oszacowań. Można zażądać tych statystyk w każdym lub tylko w ostatnim kroku.

**Prawdopodobieństwo w metodzie krokowej.** Jeśli wybrano metodę krokową, można określić prawdopodobieństwo wprowadzenia do modelu lub usunięcia z modelu. Zmienna zostaje wprowadzona, jeśli oszacowany dla niej poziom istotności dla wartości  $F$  wprowadzenia jest mniejszy niż określona wartość kryterium wprowadzenia, a zostaje wyłączona, jeśli poziom istotności jest większy niż wartość przyjęta jako kryterium usunięcia. Wartość kryterium wprowadzenia musi być mniejsza od wartości kryterium usunięcia.

**Maksymalna liczba powtórzeń.** Umożliwia określenie maksymalnej liczby iteracji dla modelu. Wartość ta wpływa na czas poszukiwania rozwiązania.

**Wyświetl funkcję bazową.** Umożliwia wyświetlenie funkcji linii bazowej hazardu i skumulowanego przeżycia przy średniej współzmiennych. Ten sposób prezentacji jest niedostępny, jeśli określono współzmiennie zależne od czasu.

## **Regresja Coxa: Definiuj zdarzenie dla zmiennej statusu**

Wprowadź wartość lub wartości wskazujące na wystąpienie zdarzenia terminalnego. Można wprowadzić jedną wartość, zakres wartości lub listę wartości. Opcja Zakres wartości jest dostępna tylko wtedy, gdy zmienna statusu jest liczbowa.

## **Dodatkowe właściwości komendy COXREG**

Język składni komend umożliwia również:

- Uzyskiwanie tabel częstości, które uwzględniają utracone obserwacje w dalszej analizie jako kategorię osobną od obserwacji ocenianych.
- Wybranie kategorii odniesienia innej niż pierwsza lub ostatnia dla metody kontrastu Odchylenie, Prosty albo Wskaźnik.



- Określenie nierównego rozmieszczenia kategorii dla metody kontrastu Wielomianowy.
- Określenie dodatkowych kryteriów iteracji.
- Kontrolowanie sposobu postępowania z brakami danych.
- Określanie nazw zapisanych zmiennych.
- Zapisywanie danych wynikowych do zewnętrznego pliku danych IBM SPSS Statistics.
- Wstrzymywanie danych dla każdej grupy pliku dzielonego w zewnętrznym pliku roboczym podczas przetwarzania. Dzięki temu można oszczędzać miejsce w pamięci w przypadku wykonywania analiz względem dużych zbiorów danych. Takie rozwiązanie nie jest dostępne w przypadku współzmiennych zależnych od czasu.

Pełne informacje na temat składni znajdują się w podręczniku *Command Syntax Reference*.

## Obliczanie współzmiennych zależnych od czasu

W niektórych sytuacjach użytkownik chce obliczyć model regresji Coxa, ale założenie o proporcjonalności hazardu nie jest spełnione. tj. ilorazy hazardu zmieniają się w czasie; wartości co najmniej jednej współzmiennych są różne w różnych punktach czasowych. W takich przypadkach należy użyć rozszerzonego modelu regresji Coxa, w którym można określić **współzmiennie zależne od czasu**.

Aby przeanalizować taki model, należy najpierw zdefiniować współzmienną zależną od czasu. Aby to ułatwić, dostępna jest "zmienna systemowa", która reprezentuje czas. Ta zmienna jest nazywana  $T_..$ . Można użyć tej zmiennej do zdefiniowania współzmiennych zależnych od czasu na dwa ogólne sposoby:

- Aby przetestować założenie o proporcjonalności hazardu w odniesieniu do konkretnej współzmiennych lub oszacować rozszerzony model regresji Coxa, który dopuszcza nieproporcjonalne zagrożenia, należy zdefiniować współzmienną zależną od czasu jako funkcję zmiennej czasu  $T_.$  i danej współzmiennych. Typowym przykładem byłby prosty iloczyn zmiennej czasu i współzmiennych, można jednak określać również bardziej złożone zmienne. Testowanie istotności współczynnika współzmiennych zależnej od czasu wskazuje, czy założenie o proporcjonalności hazardu jest rozsądne.
- Niektóre zmienne mogą mieć różne wartości w różnych okresach, ale nie są systematycznie powiązane z czasem. W takich przypadkach należy zdefiniować **posegmentowaną współzmienną zależną od czasu**, którą można wykonać za pomocą **wyrażeń logicznych**. Wyrażenie logiczne przyjmuje wartość 1, jeśli jest prawdziwe, albo 0, jeśli jest fałszywe. Używając szeregu wyrażeń logicznych, można utworzyć współzmienną zależną od czasu na podstawie zbioru pomiarów. Na przykład, jeśli ciśnienie krwi jest mierzone raz w tygodniu przez cztery tygodnie badania (określane jako  $BP1$  do  $BP4$ ), można zdefiniować współzmienną zależną od czasu jako  $(T_ < 1) * BP1 + (T_ >= 1 \& T_ < 2) * BP2 + (T_ >= 2 \& T_ < 3) * BP3 + (T_ >= 3 \& T_ < 4) * BP4$ . Można zauważyć, że dokładnie jeden z terminów w nawiasach jest równy 1 dla każdego konkretnego przypadku, a reszta będzie równa 0. Podsumowując, ta funkcja oznacza, że jeśli czas jest krótszy niż jeden tydzień, należy użyć wartości  $BP1$ . Jeśli czas jest dłuższy niż jeden tydzień, ale krótszy niż dwa tygodnie, należy użyć wartości  $BP2$  itd.

W oknie dialogowym Oblicz współzmienną zależną od czasu można użyć elementów sterujących budowania funkcji do zbudowania wyrażenia dla współzmiennych zależnej od czasu lub wprowadzić je bezpośrednio w obszarze tekstowym Nazwa. Należy zauważyć, że stałe łańcuchowe muszą być ujęte w cudzysłowy lub apostrofy, a stałe liczbowe muszą być wpisane w formacie amerykańskim, z kropką dziesiętną. Wszystkie wynikowe współzmiennie zależne od czasu muszą być uwzględnione jako współzmiennie w modelu regresji Coxa.

## Obliczanie współzmiennych zależnej od czasu

1. Z menu wybierz opcję  
**Analizuj > Przeżycie > Cox w/Time-Dep Cov...**
2. Wprowadź wyrażenie współzmiennych zależnej od czasu.
3. Kliknij przycisk **Model**, aby wykonać procedurę Regresja Coxa.

Uwaga-Upewnij się, że w modelu regresji Coxa zostały uwzględnione nowe zmienne, które zostały dodane jako współzmiennie.

## Schematy kodowania zmiennych jakościowych

W wielu procedurach można zażądać automatycznego zastąpienia niezależnej zmiennej jakościowej zbiorem kontrastów, które będą następnie wprowadzane do równania lub usuwane z równania jako jeden blok. Można określić sposób kodowania kontrastów, zwykle w opcji komendy CONTRAST. W niniejszym dodatku omówiono działanie różnych typów kontrastów żądanych w opcji komendy CONTRAST oraz przedstawiono odpowiednie przykłady.

### Odchylenie

**Odchylenie od średniej ogólnej.** Macierzowa postać tych kontrastów jest następująca:

```
mean ( 1/k 1/k ... 1/k 1/k)
df(1) ( 1-1/k -1/k ... -1/k -1/k)
df(2) ( -1/k 1-1/k ... -1/k -1/k)
.
.
df(k-1) ( -1/k -1/k ... 1-1/k -1/k)
```

gdzie  $k$  jest liczbą kategorii zmiennej niezależnej, a ostatnia kategoria jest domyślnie pominięta. Na przykład kontrasty odchylenia dla zmiennej niezależnej z trzema kategoriami są następujące:

```
( 1/3 1/3 1/3)
( 2/3 -1/3 -1/3)
(-1/3 2/3 -1/3)
```

Aby pominąć kategorię inną niż ostatnia, podaj numer kategorii do pominięcia w nawiasach za słowem kluczowym DEVIATION. Na przykład poniższa opcja komendy wyznacza odchylenia z pierwszej i trzeciej kategorii, a pomija drugą:

```
/CONTRAST(FACTOR)=DEVIATION(2)
```

Założmy, że zmienna *factor* (czynnik) ma trzy kategorie: Oto wynikowa macierz kontrastów:

```
( 1/3 1/3 1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3 2/3)
```

### Proste

**Kontrasty proste.** Porównuje każdy poziom czynnika z ostatnim. Ogólna postać macierzy jest następująca:

```
mean (1/k 1/k ... 1/k 1/k)
df(1) ( 1 0 ... 0 -1)
df(2) ( 0 1 ... 0 -1)
.
.
df(k-1) ( 0 0 ... 1 -1)
```

gdzie  $k$  jest liczbą kategorii zmiennej niezależnej. Na przykład proste kontrasty dla zmiennej niezależnej z czterema kategoriami są następujące:

```
(1/4 1/4 1/4 1/4)
( 1 0 0 -1)
( 0 1 0 -1)
( 0 0 1 -1)
```

Aby jako kategorii odniesienia użyć kategorii innej niż ostatnia, podaj numer kolejnej kategorii odniesienia w nawiasach za słowem kluczowym SIMPLE. Numer kolejny nie musi być równy z wartością powiązaną z kategorią. Na przykład poniższa opcja komendy CONTRAST wyznacza macierz kontrastów z pominięciem drugiej kategorii:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Założmy, że zmienna *factor* (czynnik) ma cztery kategorie: Oto wynikowa macierz kontrastów:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
```

```
( 0 -1 1 0)
( 0 -1 0 1)
```

## Helmerta

**Kontrasty Helmerta.** Porównuje kategorie zmiennej niezależnej ze średnią z następnych kategorii. Ogólna postać macierzy jest następująca:

```
mean (1/k 1/k ... 1/k 1/k 1/k)
df(1) ( 1 -1/(k-1) ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) ( 0 1 ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
df(k-2) ( 0 0 ... 1 -1/2 -1/2)
df(k-1) ( 0 0 ... 0 1 -1)
```

gdzie  $k$  jest liczbą kategorii zmiennej niezależnej. Na przykład zmienna niezależna z czterema kategoriami ma macierz kontrastów Helmerta o następującej postaci:

```
(1/4 1/4 1/4 1/4)
( 1 -1/3 -1/3 -1/3)
( 0 1 -1/2 -1/2)
( 0 0 1 -1)
```

## Różnica

**Kontrasty różnicy lub odwrotne kontrasty Helmerta.** Porównuje kategorie zmiennej niezależnej ze średnią z poprzednich kategorii zmiennej. Ogólna postać macierzy jest następująca:

```
mean ( 1/k 1/k 1/k ... 1/k)
df(1) ( -1 1 0 ... 0)
df(2) ( -1/2 -1/2 1 ... 0)
.
.
df(k-1) (-1/(k-1) -1/(k-1) -1/(k-1) ... 1)
```

gdzie  $k$  jest liczbą kategorii zmiennej niezależnej. Na przykład kontrasty różnicy dla zmiennej niezależnej z czterema kategoriami są następujące:

```
( 1/4 1/4 1/4 1/4)
( -1 1 0 0)
(-1/2 -1/2 1 0)
(-1/3 -1/3 -1/3 1)
```

## Wielomianowe

**Ortogonalne kontrasty wielomianowe.** Pierwszy stopień swobody zawiera efekt liniowy obejmujący wszystkie kategorie; drugi stopień swobody zawiera efekt kwadratowy; trzeci stopień swobody — sześcienny; i tak dalej w przypadku efektów wyższego rzędu.

Można określić odstępy między poziomami terapii mierzone przez daną zmienną jakościową. Równe odstępy, które obowiązują domyślnie w przypadku pominięcia metryki, można określić jako kolejne liczby całkowite od 1 do  $k$ , gdzie  $k$  jest liczbą kategorii. Jeśli zmienna *drug* (lek), ma trzy kategorie, to opcja komendy

```
/CONTRAST(DRUG)=POLYNOMIAL
```

jest równoważna

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Jednak nie zawsze potrzebne są równe odstępy. Załóżmy na przykład, że zmienna *drug* reprezentuje różne dawki leku podawane trzem grupom. Jeśli dawka podawana drugiej grupie jest dwa razy większa od dawki podawanej pierwszej grupie, a dawka podawana trzeciej grupie jest trzy razy większa od dawki podawanej pierwszej grupie, to odstępy między kategoriami terapii są równe, a odpowiednia metryka w tej sytuacji składa się z kolejnych liczb całkowitych:

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

Jeśli jednak dawka podawana drugiej grupie jest cztery razy większa od dawki podawanej pierwszej grupie, a dawka podawana trzeciej grupie jest siedem razy większa od dawki podawanej pierwszej grupie, to odpowiednia będzie metryka:

```
/CONTRAST(DRUG)=POLYNOMIAL(1,4,7)
```

W obu przypadkach pierwszy stopień swobody zmiennej *drug* zawiera efekt liniowy poziomów dawki, a drugi stopień swobody zawiera efekt kwadratowy.

Kontrasty wielomianowe są szczególnie użyteczne przy testowaniu trendów i badaniu charakteru powierzchni odpowiedzi. Kontrastów wielomianowych można też używać do nieliniowego dopasowywania krzywych, np. regresji krzywoliniowej.

## Powtórzone

**Porównuje sąsiednie poziomy zmiennej niezależnej.** Ogólna postać macierzy jest następująca:

```
mean (1/k 1/k 1/k ... 1/k 1/k)
df(1) ( 1 -1 0 ... 0 0)
df(2) ( 0 1 -1 ... 0 0)
.
.
df(k-1) ( 0 0 0 ... 1 -1)
```

gdzie *k* jest liczbą kategorii zmiennej niezależnej. Na przykład kontrasty powtórzone dla zmiennej niezależnej z czterema kategoriami są następujące:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Kontrasty te są przydatne w analizie profili i wszędzie tam, gdzie potrzebne są oceny różnic.

## Specjalne

**Kontrast zdefiniowany przez użytkownika.** Umożliwia wprowadzanie kontrastów specjalnych w postaci macierzy kwadratowej z tyloma wierszami i kolumnami, ile kategorii ma dana zmienna niezależna.

W przypadku analizy MANOVA i LOGLINEAR pierwszy wprowadzony wiersz jest zawsze efektem średnim lub stałym i odzwierciedla zbiór wag wpływających na sposób uśredniania pozostałych zmiennych niezależnych (jeśli są wybrane) według danej zmiennej. Zasadniczo ten kontrast jest wektorem jedynek.

Pozostałe wiersze macierzy zawierają kontrasty specjalne określające porównania między kategoriami zmiennej. Zwykle najbardziej użyteczne są kontrasty ortogonalne. Kontrasty ortogonalne są statystycznie niezależne i nienadmiarowe. Kontrasty są ortogonalne, jeśli:

- Dla każdego wiersza suma współczynników kontrastu wynosi 0.
- Suma iloczynów odpowiednich współczynników ze wszystkich par rozłącznych wierszy również wynosi 0.

Założmy na przykład, że terapia ma cztery poziomy i chcemy porównać ze sobą nawzajem te różne poziomy terapii. Oto odpowiedni kontrast specjalny:

```
(1 1 1 1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

który definiuje się za pomocą następującej opcji komendy CONTRAST w przypadku analizy MANOVA, LOGISTIC REGRESSION i COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1 1 1 1
                             3 -1 -1 -1
                             0 2 -1 -1
                             0 0 1 -1 )
```

W przypadku analizy LOGLINEAR należy zdefiniować:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL ( 1 1 1 1
3 -1 -1 -1
0 2 -1 -1
0 0 1 -1 )
```

Każdy wiersz z wyjątkiem wiersza średnich sumuje się do 0. Ilorazy każdej pary rozłącznych wierszy sumują się do 0, a także:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

Kontrasty specjalne nie muszą być ortogonalne. Jednak nie mogą być liniowymi kombinacjami siebie nawzajem. W takim przypadku procedura poinformuje o zależności liniowej i zakończy działanie. Kontrasty Helmerta, różnic i wielomianowe są kontrastami ortogonalnymi.

## Wskaźnik

**Kodowanie zmiennych wskaźnikowych.** Nazywane także kodowaniem sztucznym, jest niedostępne w przypadku analiz LOGLINEAR i MANOVA. Liczba kodowanych nowych zmiennych wynosi  $k-1$ . Obserwacje w kategorii odniesienia mają kod 0 dla wszystkich  $k-1$  zmiennych. Obserwacja w  $i$ -tej kategorii ma kod 0 dla wszystkich zmiennych wskaźnikowych z wyjątkiem  $i$ -tej, która ma kod 1.

## Struktury kowariancji

Ta sekcja zawiera dodatkowe informacje na temat struktur kowariancji.

**Przed-zależność: Pierwszego rzędu.** Struktura kowariancji o heterogenicznej i heterogenicznej korelacji między elementami. Korelacja między dwoma niesąsiadującymi elementami to iloczyn korelacji między tymi elementami, które są położone między badanymi elementami.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_4\sigma_1\rho_1\rho_2\rho_3$
$(\sigma_2\sigma_1\rho_1$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_2$	$\sigma_4\sigma_2\rho_2\rho_3$
$(\sigma_3\sigma_1\rho_1\rho_2$	$\sigma_3\sigma_2\rho_2$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_3$
$(\sigma_4\sigma_1\rho_1\rho_2\rho_3$	$\sigma_4\sigma_2\rho_2\rho_3$	$\sigma_4\sigma_3\rho_3$	$\sigma_4^2$

**AR(1).** Jest to elementami autoregresywna pierwszego rzędu z wariacjami homogenicznymi. Korelacja między dwoma elementami odpowiada rho dla sąsiednich elementów,  $\rho^2$  dla elementów oddzielonych jednym elementem itd. jest ograniczone, tak że  $-1 < \rho < 1$ .

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3$
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$
$(\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$

**AR(1): Heterogeniczna.** Struktura autoregresyjnej średniej ruchomej z wariacją heterogeniczną. Korelacja między dwoma elementami odpowiada r dla sąsiednich elementów,  $r^2$  dla dwóch elementów oddzielonych jednym elementem itd. i jest ograniczona zakresem od  $-1$  do  $1$ .

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho^2$	$\sigma_4\sigma_1\rho^3$
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho^2$
$(\sigma_3\sigma_1\rho^2$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho$
$(\sigma_4\sigma_1\rho^3$	$\sigma_4\sigma_2\rho^2$	$\sigma_4\sigma_3\rho$	$\sigma_4^2$

**ARMA(1,1).** Struktura autoregresyjnej średniej ruchomej pierwszego rzędu. Jej wariancje są jednorodne. Korelacja między dwoma elementami odpowiada  $\rho$  dla sąsiednich elementów,  $\rho^2$  dla elementów oddzielonych jednym elementem itd. i są parametrami odpowiednio autoregresji i średniej ruchomej, zaś ich wartości są ograniczone do zakresu od  $-1$  do  $1$ , włącznie.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2$	$\sigma^2\rho^3)$
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho^2)$
$(\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho)$
$(\sigma^2\rho^3$	$\sigma^2\rho^2$	$\sigma^2\rho$	$\sigma^2)$

**Symetria złożona.** Struktura o stałej wariancji i stałej kowariancji.

$(\sigma^2 + \sigma_1^2$	$\sigma_1$	$\sigma_1$	$\sigma_1)$
$(\sigma_1$	$\sigma^2 + \sigma_1^2$	$\sigma_1$	$\sigma_1)$
$(\sigma_1$	$\sigma_1$	$\sigma^2 + \sigma_1^2$	$\sigma_1)$
$(\sigma_1$	$\sigma_1$	$\sigma_1$	$\sigma^2 + \sigma_1^2)$

**Symetria złożona: Metryka korelacji.** Struktura kowariancji ma jednorodną wariancję i heterogeniczną korelację między elementami.

$(\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho$	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2$	$\sigma^2\rho)$
$(\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2\rho$	$\sigma^2)$

**Symetria złożona: Heterogeniczna.** Struktura kowariancji zawiera niejednorodne wariancje i stałe korelacje między elementami.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho$	$\sigma_3\sigma_1\rho$	$\sigma_4\sigma_1\rho)$
$(\sigma_2\sigma_1\rho$	$\sigma_2^2$	$\sigma_3\sigma_2\rho$	$\sigma_4\sigma_2\rho)$
$(\sigma_3\sigma_1\rho$	$\sigma_3\sigma_2\rho$	$\sigma_3^2$	$\sigma_4\sigma_3\rho)$
$(\sigma_4\sigma_1\rho$	$\sigma_4\sigma_2\rho$	$\sigma_4\sigma_3\rho$	$\sigma_4^2)$

**Przekątna.** Struktura kowariancji ma niejednorodną wariancję i zerową korelację między elementami.

$(\sigma_1^2$	$0$	$0$	$0)$
$(0$	$\sigma_2^2$	$0$	$0)$
$(0$	$0$	$\sigma_3^2$	$0)$
$(0$	$0$	$0$	$\sigma_4^2)$

**Iloczyn prosty AR1(UN\_AR1).** Określa iloczyn Kroneckera dla jednej macierzy nieustrukturyzowanej oraz drugiej macierzy kowariancji autoregresyjnej pierwszego rzędu. Pierwsza macierz nieustrukturyzowana modeluje obserwacje wielowymiarowe, a druga autoregresyjna struktura kowariancji pierwszego rzędu modeluje kowariancję danych w czasie lub innym czynniku.

**Iloczyn prosty nieustrukturyzowany (UN\_UN).** Określa iloczyn Kroneckera dwóch macierzy nieustrukturyzowanych, z których pierwsza modeluje obserwację wielowymiarową, a druga modeluje kowariancję danych w czasie lub innym czynniku.

**Iloczyn prosty - symetria złożona (UN\_CS).** Określa iloczyn Kroneckera jednej macierzy nieustrukturyzowanej i drugiej macierzy kowariancji o symetrii złożonej ze stałą wariancją i kowariancją.

Pierwsza macierz nieustrukturyzowana modeluje obserwację wielowymiarową, a druga struktura kowariancji o symetrii złożonej modeluje kowariancję danych w czasie lub innym czynniku.

**Analityczna czynnika: Pierwszego rzędu.** Ta struktura kowariancji ma heterogeniczne wariancje składające się ze składnika, który jest heterogeniczny między elementami, i składnika, który jest homogeniczny dla wszystkich elementów. Kowariancja między dowolnymi dwoma elementami to pierwiastek kwadratowy iloczynu ich homogenicznych składników wariancji.

$$\begin{pmatrix}
 \lambda_1^2 + d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\
 \lambda_2\lambda_1 & \lambda_2^2 + d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\
 \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d & \lambda_4\lambda_3 \\
 \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d
 \end{pmatrix}$$

**Analityczna czynnika: Pierwszego rzędu, Heterogeniczna.** Ta struktura kowariancji obejmuje wariancje heterogeniczne składające się z dwu składników heterogenicznych między elementami. Kowariancja między dowolnymi dwoma elementami to pierwiastek kwadratowy iloczynu ich pierwszego składnika wariancji niejednorodnej.

$$\begin{pmatrix}
 \lambda_1^2 + d_1 & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\
 \lambda_2\lambda_1 & \lambda_2^2 + d_2 & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\
 \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d_3 & \lambda_4\lambda_3 \\
 \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d_4
 \end{pmatrix}$$

**Huynh-Feldt.** Jest to macierz „cykliczna”, w której wariancja między dowolnymi dwoma elementami jest równa średniej ich wariancji minus stała. Ani wariancje, ani kowariancje nie są stałe.

$$\begin{pmatrix}
 \sigma_1^2 & [\sigma_1^2 + \sigma_2^2]/2 - \lambda & [\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_1^2 + \sigma_4^2]/2 - \lambda \\
 ([\sigma_1^2 + \sigma_2^2]/2 - \lambda & \sigma_2^2 & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda \\
 ([\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & \sigma_3^2 & [\sigma_3^2 + \sigma_4^2]/2 - \lambda \\
 ([\sigma_1^2 + \sigma_4^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda & [\sigma_3^2 + \sigma_4^2]/2 - \lambda & \sigma_4^2
 \end{pmatrix}$$

**Skalowana jednostkowa.** Struktura o stałej wariancji. Zakłada się, że między żadnymi elementami nie istnieje korelacja.

$$\begin{pmatrix}
 \sigma^2 & 0 & 0 & 0 \\
 0 & \sigma^2 & 0 & 0 \\
 0 & 0 & \sigma^2 & 0 \\
 0 & 0 & 0 & \sigma^2
 \end{pmatrix}$$

**Przestrzena: potęgowa.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  i  $j^{tym}$ .

$$\begin{pmatrix}
 \sigma^2 & \sigma^2 \rho^{d_{12}} & \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{14}} \\
 \sigma^2 \rho^{d_{12}} & \sigma^2 & \sigma^2 \rho^{d_{23}} & \sigma^2 \rho^{d_{24}} \\
 \sigma^2 \rho^{d_{13}} & \sigma^2 \rho^{d_{23}} & \sigma^2 & \sigma^2 \rho^{d_{34}} \\
 \sigma^2 \rho^{d_{14}} & \sigma^2 \rho^{d_{24}} & \sigma^2 \rho^{d_{34}} & \sigma^2
 \end{pmatrix}$$

**Przestrzena: wykładnicza.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  i  $j^{tym}$ .

$$\begin{pmatrix}
 \sigma^2 & \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{14}/\theta\}
 \end{pmatrix}$$

$(\sigma^2 \exp\{-d_{12}/\theta\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{23}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\}$
$(\sigma^2 \exp\{-d_{13}/\theta\}$	$\sigma^2 \exp\{-d_{23}/\theta\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{34}/\theta\}$
$(\sigma^2 \exp\{-d_{14}/\theta\}$	$\sigma^2 \exp\{-d_{24}/\theta\}$	$\sigma^2 \exp\{-d_{34}/\theta\}$	$\sigma^2$

**Przestrzena: Gaussa.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  i  $j^{tym}$ .

$(\sigma^2$	$\sigma^2 \exp\{-d_{12}/\rho^2\}$	$\sigma^2 \exp\{-d_{13}/\rho^2\}$	$\sigma^2 \exp\{-d_{14}/\rho^2\}$
$(\sigma^2 \exp\{-d_{12}/\rho^2\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{23}/\rho^2\}$	$\sigma^2 \exp\{-d_{24}/\rho^2\}$
$(\sigma^2 \exp\{-d_{13}/\rho^2\}$	$\sigma^2 \exp\{-d_{23}/\rho^2\}$	$\sigma^2$	$\sigma^2 \exp\{-d_{34}/\rho^2\}$
$(\sigma^2 \exp\{-d_{14}/\rho^2\}$	$\sigma^2 \exp\{-d_{24}/\rho^2\}$	$\sigma^2 \exp\{-d_{34}/\rho^2\}$	$\sigma^2$

**Przestrzena: liniowa.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  i  $j^{tym}$ , a  $1_{ij}$  to funkcja wskaźnikowa, która jest równa 1 jeśli  $pd_{ij} \leq 0$  i 0 w pozostałych przypadkach.

$(\sigma^2$	$\sigma^2(1 - \rho d_{12}) 1_{12}$	$\sigma^2(1 - \rho d_{13}) 1_{13}$	$\sigma^2(1 - \rho d_{14}) 1_{14}$
$(\sigma^2(1 - \rho d_{12}) 1_{12}$	$\sigma^2$	$\sigma^2(1 - \rho d_{23}) 1_{23}$	$\sigma^2(1 - \rho d_{24}) 1_{24}$
$(\sigma^2(1 - \rho d_{13}) 1_{13}$	$\sigma^2(1 - \rho d_{23}) 1_{23}$	$\sigma^2$	$\sigma^2(1 - \rho d_{34}) 1_{34}$
$(\sigma^2(1 - \rho d_{14}) 1_{14}$	$\sigma^2(1 - \rho d_{24}) 1_{24}$	$\sigma^2(1 - \rho d_{34}) 1_{34}$	$\sigma^2$

**Przestrzena: liniowo-logarytmiczna.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  a pomiarem  $j^{tym}$ , a  $1_{ij}$  to funkcja wskaźnikowa, która jest równa 1, jeśli  $p \log(d_{ij}) \leq 0$  i 0 w pozostałych przypadkach.

$(\sigma^2$	$\sigma^2(1 - \rho \log(d_{12})) 1_{12}$	$\sigma^2(1 - \rho \log(d_{13})) 1_{13}$	$\sigma^2(1 - \rho \log(d_{14})) 1_{14}$
$(\sigma^2(1 - \rho \log(d_{12})) 1_{12}$	$\sigma^2$	$\sigma^2(1 - \rho \log(d_{23})) 1_{23}$	$\sigma^2(1 - \rho \log(d_{24})) 1_{24}$
$(\sigma^2(1 - \rho \log(d_{13})) 1_{13}$	$\sigma^2(1 - \rho \log(d_{23})) 1_{23}$	$\sigma^2$	$\sigma^2(1 - \rho \log(d_{34})) 1_{34}$
$(\sigma^2(1 - \rho \log(d_{14})) 1_{14}$	$\sigma^2(1 - \rho \log(d_{24})) 1_{24}$	$\sigma^2(1 - \rho \log(d_{34})) 1_{34}$	$\sigma^2$

**Przestrzena: sferyczna.** Struktura kowariancji ma jednorodną wariancję i niejednorodną korelację między elementami.  $r_{ij} = d_{ij}/\rho$ , gdzie  $d_{ij}$  to estymowana odległość euklidesowa między pomiarem  $i^{tym}$  a pomiarem  $j^{tym}$ .  $1_{ij}$  to funkcja wskaźnikowa, która jest równa 1, jeśli parametr  $d_{ij}$  jest  $\leq \rho$ , i 0 w pozostałych przypadkach.

$(\sigma^2$	$\sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12}$	$\sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13}$	$\sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14}$
$(\sigma^2(1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12}$	$\sigma^2$	$\sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23}$	$\sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24}$
$(\sigma^2(1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13}$	$\sigma^2(1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23}$	$\sigma^2$	$\sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34}$
$(\sigma^2(1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14}$	$\sigma^2(1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24}$	$\sigma^2(1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34}$	$\sigma^2$

**Toeplitz.** Ta struktura kowariancji ma homogeniczne wariancje i heterogeniczne korelacje między elementami. Korelacja między sąsiadującymi elementami jest jednorodna we wszystkich parach sąsiadujących elementów. Korelacja między elementami rozdzielonymi jednym elementem jest również jednorodna itd.

$(\sigma^2$	$\sigma^2 \rho_1$	$\sigma^2 \rho_2$	$\sigma^2 \rho_3$
-------------	-------------------	-------------------	-------------------



$(\sigma^2\rho_1$	$\sigma^2$	$\sigma^2\rho_1$	$\sigma^2\rho_2$
$(\sigma^2\rho_2$	$\sigma^2\rho_1$	$\sigma^2$	$\sigma^2\rho_1$
$(\sigma^2\rho_3$	$\sigma^2\rho_2$	$\sigma^2\rho_1$	$\sigma^2$

**Toeplitz: Heterogeniczna.** Struktura kowariancji o niejednorodnej wariancji i niejednorodnej korelacji między elementami. Korelacja między sąsiadującymi elementami jest jednorodna we wszystkich parach sąsiadujących elementów. Korelacja między elementami rozdzielonymi jednym elementem jest również jednorodna itd.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_1$	$\sigma_3\sigma_1\rho_2$	$\sigma_4\sigma_1\rho_3$
$(\sigma_2\sigma_1\rho_1$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_1$	$\sigma_4\sigma_2\rho_2$
$(\sigma_3\sigma_1\rho_2$	$\sigma_3\sigma_2\rho_1$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_1$
$(\sigma_4\sigma_1\rho_3$	$\sigma_4\sigma_2\rho_2$	$\sigma_4\sigma_3\rho_1$	$\sigma_4^2$

**Nieustrukturalizowana.** W pełni ogólna macierz kowariancji.

$(\sigma_1^2$	$\sigma_{21}$	$\sigma_{31}$	$\sigma_{41}$
$(\sigma_{21}$	$\sigma_2^2$	$\sigma_{32}$	$\sigma_{42}$
$(\sigma_{31}$	$\sigma_{32}$	$\sigma_3^2$	$\sigma_{43}$
$(\sigma_{41}$	$\sigma_{42}$	$\sigma_{43}$	$\sigma_4^2$

**Nieustrukturalizowana: Metryka korelacji.** Struktura kowariancji o niejednorodnej wariancji i niejednorodnej korelacji między elementami.

$(\sigma_1^2$	$\sigma_2\sigma_1\rho_{21}$	$\sigma_3\sigma_1\rho_{31}$	$\sigma_4\sigma_1\rho_{41}$
$(\sigma_2\sigma_1\rho_{21}$	$\sigma_2^2$	$\sigma_3\sigma_2\rho_{32}$	$\sigma_4\sigma_2\rho_{42}$
$(\sigma_3\sigma_1\rho_{31}$	$\sigma_3\sigma_2\rho_{32}$	$\sigma_3^2$	$\sigma_4\sigma_3\rho_{43}$
$(\sigma_4\sigma_1\rho_{41}$	$\sigma_4\sigma_2\rho_{42}$	$\sigma_4\sigma_3\rho_{43}$	$\sigma_4^2$

**Składowe wariancji.** Struktura przypisująca strukturę Skalowana jednostkowa do każdego z określonych efektów losowych.

## Statystyki Bayesa

Produkt IBM SPSS Statistics zapewnia obsługę następujących statystyk Bayesa.

### Testy t dla jednej próby i pary prób

Procedura wnioskowania bayesowskiego dotyczącego testów t dla jednej próby udostępnia opcje wnioskowania bayesowskiego dla testów t dla prób zależnych z jednej i dwóch prób poprzez charakteryzację rozkładów a posteriori. Gdy dostępne są dane o rozkładzie normalnym, można użyć rozkładu normalnego a priori w celu uzyskania normalnego a posteriori.

### Testy proporcji dwumianowej

Procedura Wnioskowanie bayesowskie dotyczące testów t dla jednej próby: Dwumianowy udostępnia opcje wykonywania Bayesowskich testów t dla jednej próby w rozkładzie dwumianowym. Parametrem badanym jest  $\pi$ , który oznacza prawdopodobieństwo sukcesu w stałej liczbie prób, w których może wystąpić sukces lub niepowodzenie. Próby są od siebie niezależne, a prawdopodobieństwo  $\pi$  pozostaje takie samo w każdej próbie. Dwumianowa zmienna losowa może być traktowana jako suma stałej liczby niezależnych prób Bernoulliego.

### Analiza rozkładu Poissona

Procedura Wnioskowanie bayesowskie dotyczące testów t dla jednej próby: Poissona udostępnia opcje wykonywania bayesowskich testów t dla jednej próby w rozkładzie Poissona. Rozkład Poissona — użyteczny model dla rzadkich zdarzeń — zakłada, że w niewielkich przedziałach czasu

prawdopodobieństwo wystąpienia zdarzenia jest proporcjonalne do długości czasu oczekiwania. Do kreślenia statystyki bayesowskiej dla rozkładu Poissona wykorzystywane jest wstępne prawdopodobieństwo w rodzinie sprzężonych rozkładów Gamma.

### **Próby zależne**

Projekt wnioskowania bayesowskiego dotyczący testów dla prób zależnych jest podobny do bayesowskich testów dla jednej próby pod względem sposobów postępowania z próbami zależnymi. Nazwy zmiennych można podawać parami, a następnie uruchomić analizę bayesowską na różnicy między średnimi.

### **Testy t dla prób niezależnych**

Procedura wnioskowania bayesowskiego z prób niezależnych zapewnia opcje umożliwiające użycie zmiennej grupującej w celu zdefiniowania dwóch niepowiązanych grup oraz przeprowadzenia wnioskowania bayesowskiego względem różnicy dwóch średnich grupowych. Estymację czynników Bayesa można przeprowadzić, używając różnych metodologii, a także scharakteryzować żądany rozkład a posteriori zakładając, że wariancje są znane albo nieznanne.

### **Korelacja parami (Pearsona)**

Wnioskowanie bayesowskie dotyczące współczynnika korelacji Pearsona mierzy zależność liniową między dwiema zmiennymi ilościowymi, których połączenie jest zgodne z rozkładem normalnym dla pary zmiennych. Konwencjonalne wnioskowanie statystyczne dotyczące współczynnika korelacji było wielokrotnie omawiane, a jego realizacja była od dawna oferowana w IBM SPSS Statistics. Projekt wnioskowania bayesowskiego dotyczącego współczynnika korelacji Pearsona umożliwia takie wnioskowanie w oparciu o estymację czynników Bayesa oraz charakteryzację rozkładów a posteriori.

### **regresja liniowa**

Wnioskowanie Bayesowskie dotyczące regresji liniowej to metoda statystyczna popularna w modelowaniu ilościowym. Regresja liniowa jest podstawowym i standardowym podejściem, w którym badacze używają wartości kilku zmiennych w celu wyjaśnienia lub przewidzenia wartości wyniku ilościowego. Bayesowska regresja liniowa jednej zmiennej to podejście do regresji liniowej, w którym analiza statystyczna jest podejmowana w kontekście wnioskowania bayesowskiego.

### **Jednoczynnikowa ANOVA**

Bayesowska Procedura Jednoczynnikowa ANOVA generuje jednoczynnikową analizę wariancji dla ilościowej zmiennej zależnej i pojedynczego czynnika (niezależnego). Analiza wariancji jest wykorzystywana do testowania hipotezy o równości kilku średnich. Produkt SPSS Statistics obsługuje czynnika Bayesa, sprzężone rozkłady a priori oraz nieinformacyjne rozkłady a priori.

### **Modele regresji logarytmiczno-liniowej**

Plan przeznaczony do testowania niezależności dwóch czynników wymaga dwóch zmiennych jakościowych do utworzenia tabeli przypadkowej, a następnie przeprowadza wnioskowanie Bayesowskie na powiązaniu wiersz-kolumna. Estymację czynników Bayesa można przeprowadzić, przyjmując założenie różnych modeli i charakteryzując żądany rozkład a posteriori poprzez symulację jednoczesnego przedziału wiarygodności dla składników interakcji.

### **Jednoczynnikowa ANOVA z powtarzanymi pomiarami**

Procedura Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami mierzy jeden czynnik z tego samego obiektu w każdym odrębnym punkcie czasowym lub stanie i umożliwiając krzyżowanie obiektów w ramach poziomów. Zakłada się, że każdy obiekt ma jedną obserwację w każdym punkcie czasowym lub stanie (przy czym nie uwzględnia się np. indywidualnej interakcji pacjent-terapia).

## **Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Normalny**

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Normalny udostępni opcje wnioskowania bayesowskiego dla testów t dla prób zależnych z jednej i dwóch prób poprzez charakteryzację rozkładów a posteriori. Gdy dostępne są dane o rozkładzie normalnym, można użyć rozkładu normalnego a priori w celu uzyskania normalnego a posteriori.

1. Z menu wybierz:

**Analizuj > Statystyki Bayesa > Jedna próba Rozkład normalny**

2. Wybierz odpowiednie **Zmienne testowane** z listy **Dostępne zmienne**. Należy wybrać co najmniej jedną zmienną.

**Uwaga:** Lista dostępnych zmiennych udostępnia wszystkie zmienne oprócz zmiennych daty i zmiennych łańcuchowych.

3. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 2. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

1

2

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.
4. Wybierz i/lub wprowadź odpowiednie ustawienia **Wariancja danych i wartości hipotezy**. Tabela odzwierciedla zmienne, które aktualnie znajdują się na liście **Zmienne testowane**. W miarę dodawania zmiennych lub usuwania ich z listy **Zmienne testowane** w tabeli następuje automatycznie dodawanie lub usuwanie tych samych zmiennych z kolumn zmiennych.
- Gdy na liście **Zmienne testowane** znajduje się co najmniej jedna zmienna, aktywne są kolumny **Zmienna znana** i **Wartość wariancji**.

#### **Znana wariancja**

Wybierz tę opcję dla każdej zmiennej, jeśli wariancja jest znana.

#### **Wartości wariancji**

Opcjonalny parametr, który określa wartość wariancji, jeśli jest znana, dla danych obserwacji.

<sup>1</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>2</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- Gdy na liście **Zmienne testowane** znajduje się co najmniej jedna zmienna, a nie wybrano opcji **Charakteryzuj rozkład a posteriori**, aktywne są kolumny **Testowana wartość zerowa** oraz **Wartość g**.

#### **Zerowa wartość testowana**

Wymagany parametr, który określa wartość zerową w estymacji czynnika Bayesa. Dozwolona jest tylko jedna wartość, a 0 jest wartością domyślną.

#### **Wartość g**

Określa wartość w celu zdefiniowania  $\psi^2 = g\sigma^2_x$  w estymacji czynnika Bayesa. Gdy zostanie podana **Wartość wariancji**, **Wartość g** domyślnie przyjmuje wartość 1. Gdy **Wartość wariancji** nie jest podana, można określić stałe g albo pominać wartość w celu scałkowania.

5. Opcjonalnie można kliknąć opcję **Kryteria**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dotyczące testów dla jednej próby: kryteria” na stronie 110 (procent przedziału wiarygodności, opcje braków danych oraz ustawienia metody numerycznej) albo kliknąć opcję **A priori**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dotyczące testów dla jednej próby: A priori Normalny” na stronie 111 (typ analizy a priori, np. parametry wnioskowania, średniej zależne od wariancji lub precyzji).

## **Wnioskowanie bayesowskie dotyczące testów dla jednej próby: kryteria**

Na potrzeby wnioskowania bayesowskiego dotyczącego testów dla jednej próby można podać następujące kryteria analizy:

### **Procentowy przedział wiarygodności %**

Określ poziomy istotności dla obliczenia przedziałów wiarygodności. Poziomem domyślnym jest 95%.

### **Braki danych**

Podaj metodę kontroli braków danych.

### **Wyłączanie obserwacji parami**

Jest to ustawienie domyślne, które wyklucza rekordy z brakami danych analiza po analizie.

Rekordy z brakującymi wartościami dla zmiennej, która jest używana do określonego testu, są pomijane w tym teście.

### **Wyklucz wszystkie obserwacje z brakami**

To ustawienie wyklucza obserwacjami rekordy, które zawierają braki danych. Z analizy wykluczane są rekordy zawierające braki danych dla dowolnego pola nazwanego w jakiegokolwiek opcji komendy.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

### **Metoda numeryczna**

Określ metodę numeryczną, która będzie używana do oszacowania całki.

### **Adaptacyjna kwadratura Gaussa-Lobatto**

Jest to ustawienie domyślne, które wymaga podejścia z adaptacyjną kwadraturą Gaussa-Lobatto.

#### **Tolerancja**

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

Ta opcja jest dostępna tylko wtedy, gdy wybrano ustawienie **Adaptacyjna kwadratura Gaussa-Lobatto**.

#### **Maksymalna liczba iteracji**

Podaj maksymalną liczbę iteracji metody Adaptacyjna kwadratura Gaussa-Lobatto. Wartość musi być dodatnią liczbą całkowitą. Ustawieniem domyślnym jest 2000. Ta opcja jest dostępna tylko wtedy, gdy wybrano ustawienie **Adaptacyjna kwadratura Gaussa-Lobatto**.

### **Aproksymacja Monte Carlo**

Ta opcja wymaga podejścia z aproksymacją Monte Carlo.

#### **Ustaw własną wartość startową**

W przypadku wyboru tej opcji można określić własną wartość startową w polu **Wartość startowa**.

### **Wartość startowa**

Podaj zestaw wartości pseudolosowych dla metody aproksymacji Monte Carlo. Wartość musi być dodatnią liczbą całkowitą. Domyślnie przypisana zostanie wartość pseudolosowa.

### **Liczba prób Monte Carlo**

Określ liczbę punktów, które są próbkowane na potrzeby aproksymacji Monte Carlo. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 1000000. Ta opcja jest dostępna tylko wtedy, gdy wybrano ustawienie **Aproksymacja Monte Carlo**.

## **Wnioskowanie bayesowskie dotyczące testów dla jednej próby: A priori Normalny**

Na potrzeby wnioskowania bayesowskiego dotyczącego testów dla jednej próby można podać następujące kryteria rozkładu a priori:

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### **Prawdopodobieństwo wstępne zależne od wariancji/precyzji**

Udostępnia opcje definiowania wartości wariancji i precyzji.

#### **Wariancja**

Wybierz, aby określić rozkład a priori dla parametru wariancji. Gdy ta opcja jest wybrana, na liście **Rozkład a priori** dostępne są następujące opcje:

**Uwaga:** Gdy wariancja danych jest określona dla niektórych zmiennych, następujące ustawienia są dla tych zmiennych ignorowane.

- **Rozproszenie** — ustawienie domyślne. Określa rozproszenie a priori.
- **Odwrotny chi-kwadrat** — określa rozkład i parametry dla odwrotnego  $\chi^2(v_0, \sigma^2_0)$ , gdzie  $v_0 > 0$  jest stopniem swobody, a  $\sigma^2_0 > 0$  jest parametrem skali.
- **Odwrotna Gamma** — określa rozkład i parametry dla odwrotności-Gamma  $(\alpha_0, \beta_0)$ , gdzie  $\alpha_0 > 0$  jest parametrem kształtu, a  $\beta_0 > 0$  jest parametrem skali.
- **Jeffreya S2** - określa nieinformacyjny rozkład a priori  $1/\sigma^2_0$ .
- **Jeffreys S4** - określa nieinformacyjny rozkład a priori  $1/\sigma^4_0$ .

#### **Precyzja**

Wybierz, aby określić rozkład a priori dla parametru precyzji. Gdy ta opcja jest wybrana, na liście **Rozkład a priori** dostępne są następujące opcje:

- **Gamma** — określa rozkład i parametry dla Gamma  $(\alpha_0, \beta_0)$ , gdzie  $\alpha_0 > 0$  jest parametrem kształtu, a  $\beta_0 > 0$  jest parametrem skali.
- **Chi-kwadrat** — określa rozkład i parametry dla  $\chi^2(v_0)$ , gdzie  $v_0 > 0$  jest stopniem swobody.

#### **Parametr kształtu**

Określ parametr kształtu  $\alpha_0$  dla rozkładu odwrotnego-Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0.

#### **Parametr skali**

Określ parametr skali  $b_0$  dla rozkładu odwrotnego Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0. Im większy parametr skali, tym bardziej rozprzestrzenia się rozkład.

### **A priori dla średniej zależne od wariancji/precyzji**

Określ rozkład a priori dla parametru średniej, który jest zależny od parametru wariancji lub precyzji.

## Normalny

Określa rozkład i parametry dla Normalny( $\mu_0, K^{-1}_0\sigma^2_0$ ) przy wariacji lub Normalny( $\mu_0, K_0/\sigma^2_0$ ) przy precyzji, gdzie  $\mu_0 \in (-\infty, \infty)$ , a  $\sigma^2 > 0$ .

### Parametr lokalizacji

Wprowadź wartość liczbową, która określa parametr lokalizacji dla rozkładu.

### Parametr skali

Określ parametr skali  $b_0$  dla rozkładu odwrotnego Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0.

### Kappa

Określa wartość  $K_0$  w Normalny( $\mu_0, K^{-1}_0\sigma^2_0$ ) lub Normalny( $\mu_0, K_0/\sigma^2_0$ ). Należy wprowadzić pojedynczą wartość, która jest większa od 0 (wartością domyślną jest 1).

## Rozproszenie

Domyślne ustawienie, które określa rozproszenie a priori 1.

## Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Dwumianowy

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Dwumianowy udostępnia opcje wykonywania Bayesowskich testów dla jednej próby w rozkładzie dwumianowym. Parametrem badanym jest  $\pi$ , który oznacza prawdopodobieństwo sukcesu w stałej liczbie prób, w których może wystąpić sukces lub niepowodzenie. Próby są od siebie niezależne, a prawdopodobieństwo  $\pi$  pozostaje takie samo w każdej próbie. Dwumianowa zmienna losowa może być traktowana jako suma stałej liczby niezależnych prób Bernoulliego.

Mimo że nie jest to konieczne, w przypadku estymacji parametru o rozkładzie dwumianowym standardowo wybierany jest rozkład a priori z rodziny rozkładów Beta. Rodzina Beta jest sprzężeniem dla rodziny rozkładów dwumianowych i jako taka prowadzi do rozkładu a posteriori, który w ostatecznej postaci nadal należy do rodziny rozkładów Beta.

1. Z menu wybierz:

**Analizuj > Statystyka bayesowska > Dwumianowy dla jednej próby**

2. Wybierz odpowiednie **Zmienne testowane** z listy **Dostępne zmienne**. Należy wybrać co najmniej jedną zmienną.

**Uwaga:** Lista dostępnych zmiennych udostępnia wszystkie zmienne oprócz zmiennych daty i zmiennych łańcuchowych.

3. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 3. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

3

4

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.
4. Wybierz i/lub wprowadź odpowiednie ustawienia **Kategoria sukcesu i wartości hipotezy**. Tabela odzwierciedla zmienne, które aktualnie znajdują się na liście **Zmienne testowane**. W miarę dodawania zmiennych lub usuwania ich z tabeli **Zmienne testowane** w tabeli następuje automatycznie dodawanie lub usuwanie tych samych zmiennych z kolumn par zmiennych.
- Gdy opcja **Charakteryzuj rozkład a posteriori** zostanie wybrana jako **Analiza bayesowska**, następuje aktywacja kolumny **Kategoria sukcesu**.
  - Gdy opcja **Estymuj czynnik Bayesa** lub **Użyj obu metod** zostanie wybrana jako **Analiza bayesowska**, wszystkie dostępne do edycji kolumny zostaną aktywowane.

#### **Zerowa Punkt**

Włącza i wyłącza opcję **Zerowa Proporcja**. Gdy to ustawienie jest włączone, opcje **Zerowa A priori Kształt** i **Zerowa A priori Skala** są wyłączone.

#### **Zerowa A priori Kształt**

Określa parametr kształtu  $a_0$  pod hipotezą zerową wnioskowania z rozkładem dwumianowym.

#### **Zerowa A priori Skala**

Określa parametr skali  $b_0$  pod hipotezą zerową wnioskowania z rozkładem dwumianowym.

#### **Zerowa Proporcja**

Określa parametr kształtu  $a_0$  oraz parametr skali  $b_0$  dla hipotezy zerowej sprzężonego rozkładu a priori (w celu uwzględnienia wartości beta i priori Haldane). Poprawny zakres to wartości liczbowe z zakresu od 0 do 1.

#### **Alternatywna A priori Kształt**

Parametr wymagany do określenia  $a_0$  pod hipotezą alternatywną wnioskowania z rozkładem dwumianowym, jeśli ma być estymowany czynnik Bayesa.

<sup>3</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>4</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### **Alternatywna A priori Skala**

Parametr wymagany do określenia  $b_0$  pod hipotezą alternatywną wnioskowania z rozkładem dwumianowym, jeśli ma być estymowany czynnik Bayesa.

### **Kategorie sukcesu**

Udostępnia opcje definiowania sprzężonych rozkładów a priori. Udostępnione opcje określają definicję sukcesu dla zmiennych liczbowych i łańcuchowych, gdy wartości danych są testowane względem wartości testowanej.

### **Ostatnia kategoria**

Ustawienie domyślne, które wykonuje test dwumianowy z użyciem ostatniej wartości liczbowej znalezionej w kategorii po jej posortowaniu w kolejności rosnącej.

### **Pierwsza kategoria**

Wykonuje test dwumianowy z użyciem pierwszej wartości liczbowej znalezionej w kategorii po jej posortowaniu w kolejności rosnącej.

### **Punkt środkowy**

Używa jako obserwacji wartości liczbowych  $\geq$  punkt środkowy. Wartość punktu środkowego jest średnią z minimum i maksimum danych próby.

### **Punkt podziału**

Używa jako obserwacji wartości liczbowych  $\geq$  określonej wartości odcięcia. To ustawienie musi być pojedynczą wartością liczbową.

### **Poziom**

Traktuje określone przez użytkownika wartości łańcuchowe (może ich być więcej niż 1) jako obserwacje. W celu rozdzielenia różnych wartości użyj przecinków.

5. Opcjonalnie można kliknąć opcję **Kryteria**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dotyczące testów dla jednej próby: kryteria” na stronie 110 (procent przedziału wiarygodności, opcje braków danych oraz ustawienia metody numerycznej) albo kliknąć opcję **A priori**, aby określić ustawienia procedury “Bayesowskie testy dla jednej próby: A priori Dwumianowy/Poissona” na stronie 114 (rozkład sprzężony lub niestandardowy rozkład a priori).

## **Bayesowskie testy dla jednej próby: A priori Dwumianowy/Poissona**

Na potrzeby wnioskowania bayesowskiego dotyczącego testów dla jednej próby można podać następujące kryteria rozkładu a priori:

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### **Parametr kształtu**

W przypadku a priori dwumianowego określ parametr kształtu  $a_0$  do dystrybucji Beta.

W przypadku a priori Poissona określ parametr kształtu  $a_0$  do dystrybucji Gamma.

Należy wprowadzić pojedynczą wartość, która jest większa od 0.

### **Parametr skali**

W przypadku a priori dwumianowego określ parametr skali  $b_0$  do dystrybucji Beta.

W przypadku rozkładu a priori Poissona określ parametr skali  $b_0$  dla rozkładu Gamma.

Należy wprowadzić pojedynczą wartość, która jest większa od 0.

## **Wnioskowanie bayesowskie dla jednej próby: rozkład Poissona**

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura Wnioskowanie bayesowskie dotyczące testów dla jednej próby: Poissona udostępnia opcje wykonywania bayesowskich testów dla jednej próby w rozkładzie Poissona. Rozkład Poissona — użyteczny



model dla rzadkich zdarzeń — zakłada, że w niewielkich przedziałach czasu prawdopodobieństwo wystąpienia zdarzenia jest proporcjonalne do długości czasu oczekiwania. Do kreślenia statystyki bayesowskiej dla rozkładu Poissona wykorzystywane jest wstępne prawdopodobieństwo w rodzinie sprzężonych rozkładów Gamma.

1. Z menu wybierz:

**Analizuj > Statystyka bayesowska > Jedna próbka Poissona**

2. Wybierz odpowiednie **Zmienne testowane** z listy **Dostępne zmienne**. Należy wybrać co najmniej jedną zmienną.

**Uwaga:** Lista dostępnych zmiennych udostępnia wszystkie zmienne oprócz zmiennych daty i zmiennych łańcuchowych.

3. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 4. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

5

6

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

4. Wybierz i/lub wprowadź odpowiednie ustawienia **Wartości hipotezy**. Tabela odzwierciedla zmienne, które aktualnie znajdują się na liście **Zmienne testowane**. W miarę dodawania zmiennych lub usuwania ich z tabeli **Zmienne testowane** w tabeli następuje automatycznie dodawanie lub usuwanie tych samych zmiennych z kolumn par zmiennych.

<sup>5</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>6</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- Gdy opcja **Charakteryzuj rozkład a posteriori** zostanie wybrana jako **Analiza bayesowska**, żadna z tych kolumn nie zostanie aktywowana.
- Gdy opcja **Estymuj czynnik Bayesa** lub **Użyj obu metod** zostanie wybrana jako **Analiza bayesowska**, wszystkie dostępne do edycji kolumny zostaną aktywowane.

#### Zerowa Punkt

Włącza i wyłącza opcję **Zerowa Intensywność**. Gdy to ustawienie jest włączone, opcje **Zerowa A priori Kształt** i **Zerowa A priori Skala** są wyłączone.

#### Zerowa A priori Kształt

Określa parametr kształtu  $a_0$  pod hipotezą zerową wnioskowania z rozkładem Poissona.

#### Zerowa A priori Skala

Określa parametr skali  $b_0$  pod hipotezą zerową wnioskowania z rozkładem Poissona.

#### Zerowa Intensywność

Określa parametr kształtu  $a_0$  oraz parametr skali  $b_0$  dla hipotezy zerowej sprzężonego rozkładu a priori (w celu uwzględnienia relacji Poissona-Gamma). Wartość minimalna musi być wartością liczbową większą niż 0. Wartość maksymalna musi być maksymalną wartością typu double.

#### Alternatywna A priori Kształt

Parametr wymagany do określenia  $a_1$  pod hipotezą alternatywną wnioskowania z rozkładem Poissona, jeśli ma być estymowany czynnik Bayesa.

#### Alternatywna A priori Skala

Parametr wymagany do określenia  $b_1$  pod hipotezą alternatywną wnioskowania z rozkładem Poissona, jeśli ma być estymowany czynnik Bayesa.

5. Opcjonalnie można kliknąć opcję **Kryteria**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dotyczące testów dla jednej próby: kryteria” na stronie 110 (procent przedziału wiarygodności, opcje braków danych oraz ustawienia metody numerycznej) albo kliknąć opcję **A priori**, aby określić ustawienia procedury “Bayesowskie testy dla jednej próby: A priori Dwumianowy/Poissona” na stronie 114 (rozkład sprzężony lub niestandardowy rozkład a priori).

## Wnioskowanie bayesowskie dotyczące testów dla prób zależnych: Normalny

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura Wnioskowanie bayesowskie dotyczące testów dla prób zależnych: Normalny udostępni opcje Bayesowskich testów dla prób zależnych. Nazwy zmiennych można podawać parami, a następnie uruchomić analizę bayesowską na różnicy między średnimi.

1. Wybierz z menu następujące opcje:

**Analizuj > Statystyka bayesowska > Próby zależne Rozkład normalny**

2. Wybierz odpowiednie **Pary zmiennych** z listy **Dostępne zmienne**. Należy wybrać co najmniej jedną parę zmiennych źródłowych, ale dla dowolnego zestawu par nie można wybrać więcej niż dwóch zmiennych źródłowych.

**Uwaga:** Lista dostępnych zmiennych udostępni wszystkie zmienne oprócz zmiennych łańcuchowych.

3. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 5. Progi powszechnie wykorzystywane do określenia istotności dowodów*

<b>Czynnik Bayesa</b>	<b>Kategoria dowodu</b>	<b>Czynnik Bayesa</b>	<b>Kategoria dowodu</b>	<b>Czynnik Bayesa</b>	<b>Kategoria dowodu</b>
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

7

8

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.
4. Wybierz i/lub wprowadź odpowiednie ustawienia **Wariancja danych i wartości hipotezy**. Tabela odzwierciedla pary zmiennych, które aktualnie znajdują się na liście **Pary zmiennych**. W miarę dodawania par zmiennych lub usuwania ich z tabeli **Pary zmiennych** w tabeli następuje automatycznie dodawanie lub usuwanie tych samych par zmiennych z kolumn par zmiennych.
- Gdy na liście **Pary zmiennych** znajduje się co najmniej jedna para zmiennych, aktywne są kolumny **Wariancja znana** i **Wartość wariancji**.
 

**Znana wariancja**  
Wybierz tę opcję dla każdej zmiennej, jeśli wariancja jest znana.

**Wartość wariancji**  
Opcjonalny parametr, który określa wartość wariancji, jeśli jest znana, dla danych obserwacji.
  - Gdy na liście **Pary zmiennych** znajduje się co najmniej jedna para zmiennych, a nie wybrano opcji **Charakteryzuj rozkład a posteriori**, aktywne są kolumny **Testowana wartość zerowa** oraz **Wartość g**.
 

**Zerowa wartość testowa**  
Wymagany parametr, który określa wartość zerową w estymacji czynnika Bayesa. Dozwolona jest tylko jedna wartość, a 0 jest wartością domyślną.

**Wartość g**  
Określa wartość w celu zdefiniowania  $\psi^2 = g\sigma^2_x$  w estymacji czynnika Bayesa. Gdy zostanie podana **Wartość wariancji**, **Wartość g** domyślnie przyjmuje wartość 1. Gdy **Wartość wariancji** nie jest podana, można określić stałą  $g$  albo pominąć wartość w celu scałkowania.
5. Opcjonalnie można kliknąć opcję **Kryteria**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dotyczące testów dla jednej próby: kryteria” na stronie 110 (procent przedziału wiarygodności, opcje braków danych oraz ustawienia metody numerycznej) albo kliknąć opcję **A priori**,

<sup>7</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>8</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

aby określić ustawienia procedury “Bayesowskie testy dla jednej próby: A priori Dwumianowy/Poissona” na stronie 114 (rozkład sprzężony lub niestandardowy rozkład a priori).

## Bayesowskie testy dla prób niezależnych

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura wnioskowania bayesowskiego z prób niezależnych zapewnia opcje umożliwiające użycie zmiennej grupującej w celu zdefiniowania dwóch niepowiązanych grup oraz przeprowadzenia wnioskowania bayesowskiego względem różnicy dwóch średnich grupowych. Estymację czynników Bayesa można przeprowadzić, używając różnych metodologii, a także scharakteryzować żądany rozkład a posteriori zakładając, że wariancje są znane albo nieznanne.

1. Z menu wybierz:

**Analizuj > Statystyka bayesowska > Testy dla prób niezależnych**

2. Wybierz odpowiednie **Zmienne testowane** z listy zmiennych źródłowych. Należy wybrać co najmniej jedną zmienną źródłową.

3. Z listy **Dostępne zmienne** wybierz odpowiednią **Zmienną grupującą**. Zmienna grupująca definiuje dwie grupy dla testu *t* bez odpowiedników. Wybrana zmienna grupująca może być zmienną liczbową lub łańcuchową.

4. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori**: gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa**: w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0*: hipoteza zerowa

*H1*: hipoteza alternatywna

9

10

<sup>9</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.
5. Użyj opcji **Definiuj grupy**, aby zdefiniować dwie grupy dla testu  $t$ , określając dwie wartości (dla zmiennych łańcuchowych), albo dwie wartości, punkt centralny lub punkt podziału (dla zmiennych numerycznych).

**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.

W przypadku zmiennych numerycznych:

- **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.

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.

6. Opcjonalnie można kliknąć opcję **Kryteria**, aby określić ustawienia procedury “Bayesowskie testy dla prób niezależnych: Kryteria” na stronie 120 (procent przedziału wiarygodności, opcje braków danych oraz ustawienia adaptacyjnej metody kwadraturowej), kliknąć opcję **Prawdopodobieństwa a priori**, aby określić ustawienia procedury “Wnioskowanie bayesowskie dla prób niezależnych: rozkład a priori” na stronie 120 (wariancja danych, rozkład a priori zależny od wariancji oraz a priori dla średniej zależnej od wariancji) lub kliknąć opcję **Estymuj czynnik Bayesa**, aby określić ustawienia procedury “Bayesowskie testy dla prób niezależnych: Czynnik Bayesa” na stronie 121.

## Bayesowskie testy dla prób niezależnych - Definiuj grupy (liczbowe)

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.

## Bayesowskie testy dla prób niezależnych - Definiuj grupy (łańcuchowe)

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.

<sup>10</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**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.

## Bayesowskie testy dla prób niezależnych: Kryteria

Na potrzeby wnioskowania bayesowskiego dotyczącego prób niezależnych można podać następujące kryteria analizy:

### Procentowy przedział wiarygodności %

Określ poziomy istotności dla obliczenia przedziałów wiarygodności. Poziomem domyślnym jest 95%.

### Braki danych

Podaj metodę kontroli braków danych.

#### Wyłączanie obserwacji parami

Jest to ustawienie domyślne, które wyklucza rekordy z brakami danych analiza po analizie.

Rekordy z brakującymi wartościami dla zmiennej, która jest używana do określonego testu, są pomijane w tym teście.

#### Wyklucz wszystkie obserwacje z brakami

To ustawienie wyklucza obserwacjami rekordy, które zawierają braki danych. Z analizy wykluczane są rekordy zawierające braki danych dla dowolnego pola nazwanego w jakiegokolwiek opcji komendy.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

### Adaptacyjna metoda kwadraturowa

Określ wartość tolerancji i maksymalną liczbę iteracji dla adaptacyjnej metody kwadraturowej.

#### Tolerancja

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

#### Maksymalna liczba iteracji

Podaj maksymalną liczbę iteracji metody Adaptacyjna metoda kwadraturowa. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 500.

## Wnioskowanie bayesowskie dla prób niezależnych: rozkład a priori

Na potrzeby wnioskowania bayesowskiego dotyczącego prób niezależnych można podać następujące kryteria rozkładu a priori:

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### Wariancja danych

Udostępnia opcje definiowania ustawień wariancji danych.

#### Znana wariancja

W przypadku wyboru tej opcji można wprowadzić dwie znane wariancje grupowe. Obie wartości muszą być  $> 0$ .

#### Wariancja grupy 1

Wprowadź wartość pierwszej znanej wariancji grupy.

#### Wariancja grupy 2

Wprowadź wartość drugiej znanej wariancji grupy.

#### Założenie o równości wariancji

Kontroluje, czy przyjmowane jest założenie, czy nie, że dwie wariancje grupowe są równe.

Domyślnie przyjmowane jest założenie, że wariancje grupowe są nierówne. To ustawienie jest ignorowane, gdy wartości są wprowadzane dla dwóch wariancji grupowych.

### Założenie o nierówności wariancji

Kontroluje, czy przyjmowane jest założenie, czy nie, że dwie wariancje grupowe są nierówne. Domyślnie przyjmowane jest założenie, że wariancje grupowe są nierówne. To ustawienie jest ignorowane, gdy wartości są wprowadzane dla dwóch wariancji grupowych.

### Prawdopodobieństwo wstępne zależne od wariancji

Określ rozkład a priori dla dwóch różnych wariancji.

#### Jeffreysa

W przypadku wybrania tej opcji używany jest nieinformacyjny (obiektywny) rozkład a priori dla przestrzeni parametrów.

#### Odwrotny chi-kwadrat

Określa ciągły rozkład prawdopodobieństwa zmiennej losowej o wartościach dodatnich, a także parametry dla rozkładu odwrotnego  $\chi^2(\nu_0, \sigma^2_0)$ , gdzie  $\nu_0 > 0$  jest stopniem swobody, a  $\sigma^2_0 > 0$  jest parametrem skali.

#### Stopnie swobody

Określ liczbę wartości w ostatecznym obliczeniu, które mogą się wahać.

#### Parametr skali

Określ parametr skali  $\sigma^2_0 > 0$  dla odwrotności- $\chi^2(\nu_0, \sigma^2_0)$ . Należy wprowadzić pojedynczą wartość, która jest większa od 0. Im większy parametr skali, tym bardziej rozprzestrzenia się rozkład.

### A priori dla średniej zależne od wariancji

Udostępnia opcje określania rozkładu a priori dla dwóch średnich grupowych.

**Uwaga:** Opcje **Rozproszenie** i **Normalny** są dostępne tylko po wybraniu opcji **Znana wariancja**.

#### Rozproszenie

Ustawienie domyślne. Określa rozproszenie a priori.

#### Normalny

W przypadku wybrania tej opcji należy podać parametry lokalizacji i skali dla zdefiniowanych średnich grupowych.

#### Parametr lokalizacji

Wprowadź wartości liczbowe określające parametry położenia dla dystrybucji grup.

#### Parametr skali

Określ parametr skali  $\sigma^2_0 > 0$  dla odwrotności- $\chi^2(\nu_0, \sigma^2_0)$ . Dla każdej grupy należy wprowadzić pojedynczą wartość większą niż 0. Im większy parametr skali, tym bardziej rozprzestrzenia się rozkład.

## Bayesowskie testy dla prób niezależnych: Czynniki Bayesa

Możesz określić metodę estymacji czynnika Bayesa.

### Metoda Roudera

W przypadku wybrania tej opcji wywoływane jest podejście Roudera. Jest to ustawienie domyślne.

### Metoda Gonena

W przypadku wybrania tej opcji wywoływana jest metoda Gonena i należy podać następujące ustawienia wielkości efektu:

#### Średnia dla wielkości efektu

Wprowadź wartość, która określa różnicę średnich między tymi dwiema grupami.

#### Wariancja dla wielkości efektu

Wprowadź wartość, która określa wariancję dla tych dwóch grup. Wartość ta musi być  $> 0$ .

### Metoda hiperapriori

W przypadku wybrania tej opcji wywoływane jest podejście hyper-g, w którym konieczne jest podanie pojedynczej wartości. Wprowadź wartość z zakresu od -1 do -0,5 w polu **Parametr kształtu**. Wartością domyślną jest -0,75.

## Wnioskowanie bayesowskie dotyczące korelacji Pearsona

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Korelacja Pearsona mierzy zależność liniową między dwiema zmiennymi ilościowymi, których połączenie jest zgodne z rozkładem normalnym dla pary zmiennych. Konwencjonalne wnioskowanie statystyczne dotyczące współczynnika korelacji było wielokrotnie omawiane, a jego realizacja była od dawna oferowana w IBM SPSS Statistics. Projekt wnioskowania bayesowskiego dotyczącego współczynnika korelacji Pearsona umożliwia użytkownikom takie wnioskowanie w oparciu o estymację czynników Bayesa oraz charakteryzację rozkładów a posteriori.

1. Z menu wybierz:

**Analizuj > Statystyka bayesowska > Korelacja Pearsona**

2. Wybierz odpowiednie **Zmienne testowane** z listy **Dostępne zmienne**, które będą używane na potrzeby wnioskowania o korelacje parami. Należy wybrać co najmniej dwie zmienne źródłowe. W przypadku wybrania więcej niż dwóch zmiennych analiza będzie wykonywana parami na wszystkich kombinacjach wybranych zmiennych.

3. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 7. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

11

12

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

<sup>11</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>12</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.



4. Określ wartość **Maksymalna liczba wykresów**, jaka znajdzie się w danych wynikowych. Zestaw wykresów może zawierać 3 wykresy na tym samym panelu. Wykresy są generowane w kolejności od pierwszej zmiennej w odniesieniu do pozostałych zmiennych, następnie druga zmienna w odniesieniu do pozostałych zmiennych itd. Zdefiniowana wartość całkowita musi należeć do przedziału od 0 do 50. Domyślnie powstaje 10 zestawów wykresów, które przedstawiają pięć zmiennych. Ta opcja jest niedostępna w przypadku wyboru ustawienia **Estymuj czynnik Bayesa**.
5. Opcjonalnie kliknij **Kryteria**, aby podać ustawienia "Bayesowska korelacja Pearsona: Kryteria" na stronie 123 (procent okresu wiarygodności, opcje brakujących wartości i ustawienia metody numerycznej), kliknij **Poprzednie**, aby podać ustawienia "Bayesowska korelacja Pearsona: Rozkład a priori" na stronie 123 (wartość  $c$  dla wcześniejszego  $p(\rho) \propto (1 - \rho^2)^c$  lub kliknij **Współczynnik Bayesa**, aby podać ustawienia "Bayesowskie testy dla prób niezależnych: Czynnik Bayesa" na stronie 121.

## Bayesowska korelacja Pearsona: Kryteria

Na potrzeby wnioskowania bayesowskiego dotyczącego korelacji Pearsona można (parami) podać poniższe kryteria analizy.

### Procentowy przedział wiarygodności %

Określ poziom istotności dla obliczenia przedziałów wiarygodności. Poziomym domyślnym jest 95%.

### Braki danych

Podaj metodę kontroli braków danych.

### Wyłączanie obserwacji parami

To ustawienie wyklucza parami rekordy, które zawierają braki danych.

### Wyłącz wszystkie obserwacje z brakami

To ustawienie wyklucza obserwacjami rekordy, które zawierają braki danych. Z analizy wykluczane są rekordy zawierające braki danych dla dowolnego pola nazwanego w jakiegokolwiek opcji komendy.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

### Metoda numeryczna

Określ metodę numeryczną, która będzie używana do oszacowania całki.

### Wartość startowa generatora liczb losowych

W przypadku wyboru tej opcji można określić własną wartość startową w polu **Wartość startowa**.

### Tolerancja

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

### Maksymalna liczba iteracji

Podaj maksymalną liczbę iteracji metody. Wartość musi być dodatnią liczbą całkowitą. Ustawieniem domyślnym jest 2000.

### Liczba prób Monte Carlo

Określ liczbę punktów, które są próbkowane na potrzeby aproksymacji Monte Carlo. Wartość musi być dodatnią liczbą całkowitą. Wartością domyślną jest 10000.

### Symulowane próby dla uzyskania rozkładu a posteriori

Podaj liczbę prób, które zostaną użyte w celu uzyskania żądanego rozkładu a posteriori. Wartością domyślną jest 10000.

## Bayesowska korelacja Pearsona: Rozkład a priori

Możesz określić wartość  $c$  dla rozkładu a priori  $p(\rho) \propto (1 - \rho^2)^c$ .

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### **Jednostajny (c = 0)**

W przypadku wyboru tej opcji używany jest jednostajny rozkład a priori.

### **Jeffreys (c = -1.5)**

W przypadku wybrania tej opcji używany jest nieinformacyjny rozkład a priori.

### **Ustaw niestandardową wartość c**

W przypadku wyboru tej opcji można określić własną **Wartość c**. Dozwolona jest dowolna pojedyncza liczba rzeczywista.

## **Korelacja Pearsona w ujęciu bayesowskim: współczynnik Bayesa**

Możesz określić metodę estymacji czynnika Bayesa. Następujące opcje są dostępne tylko wtedy, gdy wybrano analizę Bayesowską **Estymuj czynnik Bayesa** lub opcję **Użyj obu metod**.

### **Czynnik Bayesa JZS**

W przypadku wybrania tej opcji wywoływane jest podejście Zellnera-Siowa. To jest ustawienie domyślne.

### **Ułamkowy czynnik Bayesa**

W przypadku wyboru tej opcji można określić ułamkowy czynnik Bayesa i wartość hipotezy zerowej. W przypadku ułamkowego czynnika Bayesa należy określić wartość (0, 1). Wartością domyślną jest 0,5.

## **Wnioskowanie bayesowskie dotyczące modeli regresji liniowej**

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Regresja to metoda statystyczna popularna w modelowaniu ilościowym. Regresja liniowa jest podstawowym i standardowym podejściem, w którym badacze używają wartości kilku zmiennych w celu wyjaśnienia lub przewidzenia wartości wyniku ilościowego. Bayesowska regresja liniowa jednej zmiennej to podejście do regresji liniowej, w którym analiza statystyczna jest podejmowana w kontekście wnioskowania bayesowskiego.

Procedurę regresji można wywołać i zdefiniować pełny model.

1. Z menu wybierz:

**Analiza > Statystyka bayesowska > Regresja liniowa**

2. Wybierz pojedynczą, nietańczuchową zmienną zależną z listy **Dostępne zmienne**. Musisz wybrać jedną zmienną nietańczuchową.
3. Wybierz dla modelu co najmniej jedną zmienną czynnika jakościowego z listy **Dostępne zmienne**.
4. Wybierz z listy **Dostępne zmienne** co najmniej jedną nietańczuchową współzmienną ilościową.

**Uwaga:** Obie listy **Czynniki** i **Współzmienne** nie mogą być puste. Musisz wybrać co najmniej jeden **Czynnik** lub **Współzmienną** zmienną.

5. Opcjonalnie wybierz z listy **Dostępne zmienne** pojedynczą zmienną nietańczuchową, która będzie pełnić rolę wagi regresji.
6. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 8. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

13

14

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

Opcjonalnie można wykonać następujące czynności:

- Kliknij opcję **Kryteria**, aby określić procentowy przedział wiarygodności oraz ustawienia metody numerycznej.
- Kliknij opcję **A priori**, aby zdefiniować odniesienie i ustawienia rozkładu sprzężonego a priori.
- Kliknij opcję **Czynnik Bayesa**, aby określić ustawienia czynnika Bayesa.
- Kliknij przycisk **Zapisz**, aby zidentyfikować elementy do zapisania, a następnie zapisz informacje o modelu w pliku XML.
- Kliknij opcję **Predykcja**, aby określić zmienne niezależne dla predykcji bayesowskiej.
- Kliknij opcję **Wykresy**, aby wykreślić rozkłady a posteriori parametrów regresji, wariację warunków błędów i wartości przewidywane.
- Kliknij opcję **Testy F**, aby porównać modele statystyczne w celu zidentyfikowania modelu najlepiej pasującego do populacji, z której go wylosowano.

## Bayesowskie modele regresji liniowej: Kryteria

Na potrzeby modeli bayesowskich regresji liniowej można podać poniższe kryteria analizy.

### Procentowy przedział wiarygodności %

Określ poziomy istotności dla obliczenia przedziałów wiarygodności. Poziomym domyślnym jest 95%.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

### Metoda numeryczna

Określ metodę numeryczną, która będzie używana do oszacowania całki.

<sup>13</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>14</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Tolerancja

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

### Maksymalna liczba iteracji

Podaj maksymalną liczbę iteracji metody. Wartość musi być dodatnią liczbą całkowitą. Ustawieniem domyślnym jest 2000.

## Bayesowskie modele regresji liniowej: Rozkłady a priori

Następujące ustawienia rozkładu a priori można określić dla parametrów regresji oraz wariancji błędów. Następujące opcje są dostępne tylko wtedy, gdy wybrano opcję **Charakteryzuj rozkład a posteriori** na potrzeby **Analizy bayesowskiej**.

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### Odniesienie

Jeśli ta opcja jest wybrana, analiza referencyjna zwraca obiektywne wnioski bayesowskie. Instrukcje wnioskujące są zależne tylko od przyjętego modelu oraz od dostępnych danych, a rozkład a priori, który jest używany w celu wnioskowania, jest najmniej informacyjny. To jest ustawienie domyślne.

### Sprzężony

Udostępnia opcje definiowania sprzężonych rozkładów a priori. Sprzężone rozkłady a priori przyjmują rozkład normalny-odwrotny-Gamma-tączny. Mimo że sprzężone rozkłady a priori nie są wymagane w przypadku wykonywania aktualizacji Bayesowskich, ułatwiają one procesy obliczeń.

**Uwaga:** W celu określenia prawdopodobieństwa a priori koniugatów dla modelu regresji liniowej, należy ustawić oczekiwaną wartość średnią parametrów regresji w tabeli **Prawdopodobieństwa a priori wariancji błędu**. Można też użyć ustawień **Wariancja macierzy kowariancji**, aby określić wstępną wariancję-kowariancję.

### A priori zależne od wariancji błędów

#### Parametr kształtu

Określ parametr kształtu  $a_0$  dla rozkładu odwrotnego-Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0.

#### Parametr skali

Określ parametr skali  $b_0$  dla rozkładu odwrotnego Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0. Im większy parametr skali, tym bardziej rozprzestrzenia się rozkład.

W tabeli wyświetlana jest średnia liczba parametrów regresji (wraz z wyrazem wolnym), która określa średni wektor  $\theta_0$  dla zdefiniowanych parametrów regresji. Liczba wartości musi być zgodna z liczbą parametrów regresji z uwzględnieniem wyrazu wolnego.

Pierwsza zmienna ma zawsze nazwę INTERCEPT. Z drugiego wiersza kolumna **Zmienne** jest automatycznie wypełniana zmiennymi, które są określone przez Czynniki i Współzmienne. Kolumna **Średniej** nie zawiera żadnych wartości domyślnych.

W celu skasowania wartości kliknij opcję **Resetuj**.

### Macierz wariancji-kowariancji: $\sigma^2x$

Określ wartości  $V_0$  w dolnym trójkącie macierzy wariancji-kowariancji dla rozkładu normalnego a priori wielu zmiennych. Należy zwrócić uwagę na to, że  $V_0$  musi być semi-dodatnią wartością skończoną. Ostatnia wartość każdego wiersza musi być dodatnia. Następny wiersz powinien zawierać o jedną wartość więcej niż poprzedni wiersz. W przypadku kategorii odniesienia (jeśli takie istnieją) nie są określane żadne wartości.

W celu skasowania wartości kliknij opcję **Resetuj**.

### **Użyj macierzy jednostkowej**

W przypadku wyboru tej opcji używana jest skalowana macierz jednostkowa. W przypadku normalnego rozkład a priori nie można określać wartości  $V_0$  w dolnym trójkącie macierzy wariancji-kowariancji.

## **Bayesowskie modele regresji liniowej: Czynniki Bayesa**

W tym przypadku można określić plan modelu dla analizy, który będzie obejmował podejście stosowane do estymacji czynnika Bayesa dla Bayesowskich modeli regresji liniowej. Następujące opcje są dostępne tylko wtedy, gdy wybrano analizę Bayesowską **Estymuj czynniki Bayesa** lub opcję **Użyj obu metod**.

### **Model zerowy**

W przypadku wybrania tej opcji szacowane czynniki Bayesa są oparte na modelu zerowym. To jest ustawienie domyślne.

### **Model pełny**

W przypadku wybrania tej opcji szacowane czynniki Bayesa są oparte na modelu pełnym i możliwe jest wybranie zmiennych i dodatkowych czynników oraz współzmiennych.

### **Zmienne**

Powoduje wyświetlenie listy wszystkich zmiennych dostępnych dla modelu pełnego.

### **Dodatkowe czynniki**

Wybierz zmienne z listy **Zmienne**, które będą używane jako dodatkowe czynniki.

### **Dodatkowe współzmiennie**

Wybierz zmienne z listy **Zmienne**, które będą używane jako dodatkowe współzmiennie.

### **Obliczenie**

Określ podejście estymacji czynników Bayesa. Ustawieniem domyślnym jest metoda JZS.

### **Metoda JZS**

W przypadku wybrania tej opcji wywoływane jest podejście Zellnera-Siowa. To jest ustawienie domyślne.

### **Metoda Zellnera**

Jeśli ta opcja jest wybrana, wywołuje metodę Zellnera i wymagane jest podanie pojedynczej wartości  $g$  rozkładu a priori  $> 0$  (nie istnieje wartość domyślna).

### **Metoda hiperapriori**

W przypadku wybrania tej opcji wywoływane jest podejście hiper- $g$  i wymagane jest podanie parametru kształtu  $\alpha_0$  dla odwrotnego rozkładu Gamma. Należy podać pojedynczą wartość  $> 0$  (wartością domyślną jest 3).

### **Metoda Roudera**

W przypadku wybrania tej opcji wywoływane jest podejście Roudera i wymagane jest podanie skali kształtu  $b_0$  dla odwrotnego rozkładu Gamma. Należy podać pojedynczą wartość  $> 0$  (wartością domyślną jest 1).

## **Bayesowskie modele regresji liniowej: Zapisz**

W tym oknie dialogowym można określić statystyki oceniane dla rozkładu predykcji bayesowskiej oraz wyeksportować wyniki modelu do pliku XML.

### **Statystyka predykcyjna a posteriori**

W przypadku tej opcji możliwe jest wyznaczenie następujących statystyk, które są oparte na predykcjach bayesowskich.

### **Średnie**

Średnia przewidywanego rozkładu a posteriori.

### **Wariancje**

Wariancja przewidywanego rozkładu a posteriori.

### **Dominanty**

Dominanta przewidywanego rozkładu a posteriori.

### **Dolna granica przedziału wiarygodności**

Dolna granica przedziału wiarygodności przewidywanego rozkładu a posteriori.

### **Górna granica przedziału wiarygodności**

Górna granica przedziału wiarygodności przewidywanego rozkładu a posteriori.

**Uwaga:** Do każdej statystyki można przypisać nazwy odpowiadających zmiennych.

### **Eksportuj informacje o modelu do pliku XML**

Wprowadź nazwę i lokalizację pliku XML, do którego zostanie wyeksportowana macierz wariancji-kowariancji ocenianego parametru.

## **Bayesowskie modele regresji liniowej: Predykcja**

Możliwe jest określenie zmiennych niezależnych w celu wygenerowania rozkładów przewidywanych.

### **Zmienne niezależne dla predykcji bayesowskiej**

Ta tabela zawiera listę wszystkich dostępnych zmiennych niezależnych. Kolumna **Zmienne niezależne** jest automatycznie wypełniana niektórymi zmiennymi czynnikowymi oraz współzmiennymi. Określ wektor wartości obserwowanych z wartościami dla zmiennych niezależnych. Każdej zmiennej niezależnej można przypisać jedną wartość lub jeden łańcuch, a następnie zezwolić na predykcję tylko jednej obserwacji. W przypadku czynników dozwolone są wartości oraz łańcuchy.

W celu przeprowadzenia predykcji (poprzez kliknięcie opcji **Uruchom analizę**) musisz podać wszystkie wartości zmiennych niezależnych albo nie podawać żadnej z nich.

Gdy zmienna czynnikowa lub współzmienna zostanie usunięta, wiersz odpowiadającej zmiennej niezależnej zostanie usunięty z tabeli.

W przypadku współzmiennych można podawać tylko wartości liczbowe. W przypadku czynników dozwolone są wartości liczbowe oraz łańcuchy.

**Uwaga:** W celu skasowania zdefiniowanych wartości kliknij opcję **Resetuj**.

## **Bayesowskie modele regresji liniowej: Wykresy**

Możliwe jest kontrolowanie wykresów, które będą generowane.

### **Współzmiennie**

Powoduje wyświetlenie listy aktualnie zdefiniowanych współzmiennych.

### **Wykres współzmiennych**

Wybierz współzmiennie przeznaczone do umieszczenia na wykresie (z listy **Współzmiennie**), a następnie dodaj je do listy **Wykres współzmiennych**.

### **Czynniki**

Powoduje wyświetlenie listy aktualnie zdefiniowanych czynników.

### **Wykres czynników**

Wybierz czynniki przeznaczone do umieszczenia na wykresie (z listy **Czynniki**), a następnie dodaj je do listy **Wykres czynników**.

### **Maksymalna liczba kategorii na wykresie**

Wybierz maksymalną liczbę kategorii do umieszczenia na wykresie (pojedyncza, dodatnia wartość całkowita). To ustawienie dotyczy wszystkich czynników. Domyślnie w przypadku każdego czynnika na wykresie umieszczane są 2 pierwsze poziomy.

### **Dołącz wykresy**

#### **Wyraz wolny**

W przypadku wybrania tej opcji na wykresie umieszczany jest wyraz wolny. Domyślnie to ustawienie nie jest wybrane.

#### **Wariancja składników błędu**

W przypadku wybrania tej opcji tworzony jest wykres wariancji błędów. Domyślnie to ustawienie nie jest wybrane.

### Przewidywany rozkład Bayesa

W przypadku wybrania tej opcji tworzony jest wykres przewidywanego rozkładu. Domyślnie to ustawienie nie jest wybrane. To ustawienie można wybrać tylko wtedy, gdy wybrano poprawne wartości zmiennych niezależnych.

## Bayesowskie modele regresji liniowej: Testy F

Możliwe jest utworzenie co najmniej jednego częściowego testu F. Test F jest testem statystycznym, w którym statystyka testowa ma rozkład F w ramach hipotezy zerowej. Testy F są zwykle stosowane w przypadku porównywania modeli statystycznych dopasowanych do zestawu danych, aby zidentyfikować model najlepiej pasujący do populacji, z której losowano dane.

### Dostępne zmienne

Ta opcja powoduje wyświetlenie listy zmiennych czynnikowych i współzmiennych, które zostały wybrane z okna dialogowego Bayesowskiej regresji liniowej. Gdy zmienne czynnikowe i współzmiennne są dodawane albo usuwane z głównego okna dialogowego, lista jest odpowiednio aktualizowana.

### Zmienne testowane

Z listy **Dostępne zmienne** wybierz zmienne czynnikowe/współzmiennne do testowania, a następnie dodaj je do listy **Zmienne testujące**.

**Uwaga:** Jeśli nie wybrano żadnych zmiennych testujących ani współzmiennych, należy wybrać opcję **Uwzględnij wyraz wolny**.

### Zmienne testujące i wartości

Określ wartości do testowania. Liczba wartości musi być zgodna z liczbą parametrów w modelu oryginalnym. Gdy wartości są określone, należy podać pierwszą wartość dla wyrazu wolnego (przy założeniu, że wszystkie wartości wynoszą 0, jeśli nie zostaną jawnie zdefiniowane).

### Uwzględnij wyraz wolny

W przypadku wybrania tej opcji w teście uwzględniane są wyrazy wolne. Domyślnie to ustawienie nie jest wybrane.

W przypadku włączenia tej opcji należy użyć pola **Wartość testująca**, aby określić wartość.

### Etykieta testu (opcjonalna)

Opcjonalnie można określić etykietę dla każdego testu. Można określić wartość łańcuchową o maksymalnej długości 255 bajtów. Dozwolona jest tylko jedna etykieta na każdy test F.

## Bayesowska jednoczynnikowa ANOVA

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Procedura Jednoczynnikowa ANOVA generuje jednoczynnikową analizę wariancji dla ilościowej zmiennej zależnej i pojedynczego czynnika (niezależnego). Analizę wariancji wykorzystuje się do testowania hipotezy o równości kilku średnich. Produkt SPSS Statistics obsługuje czynnika Bayesa, sprzężone rozkłady apriori oraz nieinformacyjne rozkłady apriori. Analiza wariancji jest używana do testowania hipotezy, że kilka średnich jest równe.

1. Z menu wybierz:

**Analizuj > Statystyka Bayesa > Jednoczynnikowa ANOVA**

2. Wybierz pojedynczą, liczbową zmienną **Zależną** z listy **Dostępne zmienne**. Musisz wybrać co najmniej jedną zmienną.

3. Wybierz dla modelu pojedynczą zmienną **Czynnik** z listy **Dostępne zmienne**. Musisz wybrać co najmniej jedną zmienną **Czynnik**.

4. Wybierz z listy **Dostępne zmienne** pojedynczą, nielańcuchową zmienną, która będzie pełnić rolę **Wagi** regresji. Pole zmiennej **Waga** może być puste.

5. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy

rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.

- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 9. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

15

16

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

Opcjonalnie można wykonać następujące czynności:

- Kliknij opcję **Kryteria**, aby określić procentowy przedział wiarygodności oraz ustawienia metody numerycznej.
- Kliknij opcję **A priori**, aby zdefiniować odniesienie i ustawienia rozkładu sprzężonego a priori.
- Kliknij opcję **Czynnik Bayesa**, aby określić ustawienia czynnika Bayesa.
- Kliknij opcję **Wykresy**, aby kontrolować wykresy, które będą generowane.

## Bayesowska jednoczynnikowa ANOVA: Kryteria

Na potrzeby bayesowskiej analizy jednoczynnikowej ANOVA można podać poniższe kryteria analizy.

### Procentowy przedział wiarygodności %

Określ poziomy istotności dla obliczenia przedziałów wiarygodności. Poziomem domyślnym jest 95%.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

### Metoda numeryczna

Określ metodę numeryczną, która będzie używana do oszacowania całki.

<sup>15</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>16</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.



### Tolerancja

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

### Maksymalna liczba iteracji

Podaj maksymalną liczbę iteracji metody. Wartość musi być dodatnią liczbą całkowitą. Ustawieniem domyślnym jest 2000.

## Bayesowska jednoczynnikowa ANOVA: A priori

Następujące ustawienia rozkładu a priori można określić dla parametrów regresji oraz wariancji błędów. Następujące opcje są dostępne tylko wtedy, gdy wybrano opcję **Charakteryzuj rozkład a posteriori** na potrzeby **Analizy bayesowskiej**.

**Uwaga:** Wielu badaczy może zakwestionować konieczność określenia a priori. Referencyjne a priori minimalizują problemy wynikające z przekroczenia a priori przy wzroście danych. Po określeniu informacyjnych wartości a priori podczas zastosowania metod bayesowskich można skutecznie korzystać z tych informacji. Wymogi dotyczące określenia a priori nie powinny działać jak czynnik zniechęcający do używania analizy bayesowskiej.

### Odniesienie

Jeśli ta opcja jest wybrana, analiza referencyjna zwraca obiektywne wnioskowanie Bayesowskie. Instrukcje wnioskujące są zależne tylko od przyjętego modelu oraz od dostępnych danych, a rozkład a priori, który jest używany w celu wnioskowania, jest najmniej informacyjny. To jest ustawienie domyślne.

### Sprzężenie

Udostępnia opcje definiowania sprzężonych rozkładów a priori. Sprzężone rozkłady a priori przyjmują rozkład normalny-odwrotny-Gamma-łączny. Mimo że sprzężone rozkłady a priori nie są wymagane w przypadku wykonywania aktualizacji Bayesowskich, ułatwiają one procesy obliczeń.

### A priori zależne od wariancji błędów

#### Parametr kształtu

Określ parametr kształtu  $a_0$  dla rozkładu odwrotnego-Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0.

#### Parametr skali

Określ parametr skali  $b_0$  dla rozkładu odwrotnego Gamma. Należy wprowadzić pojedynczą wartość, która jest większa od 0. Im większy parametr skali, tym bardziej rozprzestrzenia się rozkład.

### A priori wg parametrów regresji

Określ wektor  $\beta_0$  średnich dla średnich grupowych. Liczba wartości musi być zgodna z liczbą parametrów regresji z uwzględnieniem wyrazu wolnego.

Kolumna **Zmienne** jest automatycznie wypełniana poziomami czynnika. Kolumna **Średniej** nie zawiera żadnych wartości domyślnych.

W celu skasowania wartości kliknij opcję **Resetuj**.

### Macierz wariancji-kowariancji: $\sigma^2x$

Określ wartości  $V_0$  w dolnym trójkącie macierzy wariancji-kowariancji dla rozkładu normalnego a priori wielu zmiennych. Należy zwrócić uwagę na to, że  $V_0$  musi być semi-dodatnią wartością skończoną. Należy określić tylko dolny trójkąt tabeli.

Wiersze i kolumny są automatycznie wypełniane poziomami czynnika. Wszystkie wartości na przekątnych są równe 1; wszystkie wartości poza przekątnymi są równe 0.

W celu skasowania wartości kliknij opcję **Resetuj**.

### Użyj macierzy jednostkowej

W przypadku wyboru tej opcji używana jest macierz jednostkowa. W przypadku normalnego rozkład a priori nie można określać wartości  $V_0$  w dolnym trójkącie macierzy wariancji-kowariancji.

## Bayesowska jednoczynnikowa ANOVA: Czynniki Bayesa

W tym przypadku można określić podejście stosowane do estymacji czynnika Bayesa dla Bayesowskich modeli jednoczynnikowych ANOVA. Następujące opcje są dostępne tylko wtedy, gdy wybrano analizę Bayesowską **Estymuj czynniki Bayesa** lub opcję **Użyj obu metod**.

### Obliczenie

Określ podejście estymacji czynników Bayesa. Ustawieniem domyślnym jest metoda JZS.

#### Metoda JZS

W przypadku wybrania tej opcji wywoływane jest podejście Zellnera-Siowa. To jest ustawienie domyślne.

#### Metoda Zellnera

Jeśli ta opcja jest wybrana, wywołuje metodę Zellnera i wymagane jest podanie pojedynczej wartości  $g$  rozkładu a priori  $> 0$  (nie istnieje wartość domyślna).

#### Metoda hiperapriori

W przypadku wybrania tej opcji wywoływane jest podejście hiper- $g$  i wymagane jest podanie parametru kształtu  $a_0$  dla odwrotnego rozkładu Gamma. Należy podać pojedynczą wartość  $> 0$  (wartością domyślną jest 3).

#### Metoda Roudera

W przypadku wybrania tej opcji wywoływane jest podejście Roudera i wymagane jest podanie skali kształtu  $b_0$  dla odwrotnego rozkładu Gamma. Należy podać pojedynczą wartość  $> 0$  (wartością domyślną jest 1).

## Bayesowska jednoczynnikowa ANOVA: Wykresy

Możliwe jest kontrolowanie wykresów, które będą generowane.

### Grupy wykresów

Określ grupy, które będą umieszczane na wykresie. Wygeneruj wykresy wiarygodności, rozkładu a priori oraz a posteriori dla średnich z podanych grup. Lista **Grupy** jest podzbiorem kategorii zmiennej czynnikowej, dlatego jej format powinien być zgodny z typem danych czynnika oraz wartościami rzeczywistymi.

### Wariancja składników błędów

W przypadku wybrania tej opcji tworzony jest wykres wariancji błędów. Domyślnie to ustawienie nie jest wybrane. Ta opcja jest niedostępna w przypadku wyboru opcji **Estymuj czynniki Bayesa** jako analizy bayesowskiej.

## Bayesowskie modele logliniowe

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

Plan przeznaczony do testowania niezależności dwóch czynników wymaga dwóch zmiennych jakościowych do utworzenia tabeli przypadkowej, a następnie przeprowadza wnioskowanie Bayesowskie na powiązaniu wiersz-kolumna. Estymację czynników Bayesa można przeprowadzić, przyjmując założenie różnych modeli i charakteryzując żądany rozkład a posteriori poprzez symulację jednoczesnego przedziału wiarygodności dla składników interakcji.

1. Z menu wybierz:

#### **Analizuj > Statystyka bayesowska > Modele log-liniowe**

2. Wybierz z listy **Dostępne zmienne** pojedynczą, nieskalową zmienną wierszową. Musisz wybrać co najmniej jedną zmienną nieskalową.

3. Wybierz z listy **Dostępne zmienne** pojedynczą, nieskalową zmienną kolumnową. Musisz wybrać co najmniej jedną zmienną nieskalową.

4. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori**: gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy

rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.

- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 10. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

17

18

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

Opcjonalnie można wykonać następujące czynności:

- Kliknij opcję **Kryteria**, aby określić procentowy przedział wiarygodności oraz ustawienia metody numerycznej.
- Kliknij opcję **Czynnik Bayesa**, aby określić ustawienia czynnika Bayesa.
- Kliknij opcję **Wyświetlaj**, aby określić sposób wyświetlania zawartości w tabelach wynikowych.

## Bayesowskie modele logliniowe: Kryteria

Na potrzeby bayesowskich modeli logliniowych można podać poniższe kryteria analizy.

### Procentowy przedział wiarygodności %

Określ poziomy istotności dla obliczenia przedziałów wiarygodności. Poziomem domyślnym jest 95%.

### Metoda numeryczna

Określ metodę numeryczną, która będzie używana do oszacowania całki.

### Klucz początkowy generatora liczb losowych

W przypadku wyboru tej opcji można określić własną wartość startową w polu **Wartość startowa**. Podaj ustaloną wartość pseudolosową. Wartość musi być dodatnią liczbą całkowitą. Domyślnie przypisana zostanie wartość pseudolosowa.

<sup>17</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>18</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Uwaga:** Następujące opcje są dostępne tylko wtedy, gdy jako analizę Bayesowską (opcja **Analiza bayesowska**) wybrano opcję **Estymuj czynnik Bayesa** lub **Użyj obu metod**.

#### **Tolerancja**

Określ wartość tolerancji dla metod numerycznych. Ustawieniem domyślnym jest 0,000001.

#### **Maksymalna liczba iteracji**

Podaj maksymalną liczbę iteracji metody. Wartość musi być dodatnią liczbą całkowitą. Ustawieniem domyślnym jest 2000.

#### **Symulowane próby dla uzyskania rozkładu a posteriori**

Podaj liczbę prób, które zostaną użyte w celu uzyskania żądanego rozkładu a posteriori. Wartością domyślną jest 10000.

#### **Formatuj**

Wybierz, czy kategorie będą wyświetlane w kolejności **Rosnąco** czy **Malejąco**. Ustawieniem domyślnym jest **Rosnąco**.

## **Bayesowskie modele logliniowe: Czynnik Bayesa**

Możesz określić model przyjmowany dla danych obserwacji (Poissona, Wielomianowy lub Nieparametryczny). Ustawieniem domyślnym jest rozkład wielomianowy. Następujące opcje są dostępne tylko wtedy, gdy wybrano analizę Bayesowską **Estymuj czynnik Bayesa** lub opcję **Użyj obu metod**.

#### **Model Poissona**

Po wybraniu tej opcji dla danych obserwowanych przyjmowany jest model Poissona.

#### **Model wielomianowy**

Po wybraniu tej opcji dla danych obserwowanych przyjmowany jest model wielomianowy. To jest ustawienie domyślne.

#### **Stałe marginesy**

Wybierz opcję **Podsumowanie całości**, **Suma wiersza** lub **Suma kolumny**, aby określić stałe sumy brzegowe dla tabeli przypadkowej. Ustawieniem domyślnym jest **Podsumowanie całości**.

#### **Rozkład a priori**

Określ typ rozkładu a priori w przypadku estymacji czynnika Bayesa.

#### **Sprzężony**

Wybierz tę opcję, aby określić sprzężony rozkład a priori. Tabela **Parametry kształtu** umożliwi określenie parametrów kształtu  $a_{rs}$  dla rozkładu Gamma. Jeśli jako typ rozkładu a priori wybrany jest **Sprzężony**, należy określić parametry kształtu.

Jeśli zostanie podana pojedyncza wartość, przyjmuje się, że wszystkie wartości  $a_{rs}$  są równe tej wartości. Ustawieniem domyślnym jest  $a_{rs} = 1$ . Jeśli wymagane jest podanie więcej niż jednej wartości, można rozdzielić wartości spacjami.

Liczba wartości liczbowych podanych w każdym wierszu i każdej kolumnie musi być zgodna z wymiarem tabeli przypadkowej. Wszystkie podane wartości muszą być  $> 0$ .

W celu skasowania wartości kliknij opcję **Resetuj**.

#### **Parametr skali**

Określ parametr skali  $b$  dla rozkładu Gamma. Należy podać pojedynczą wartość  $> 0$ .

#### **Mieszany Dirichleta**

Wybierz tę opcję, aby określić rozkład a priori mieszany Dirichleta.

#### **Własny**

Wybierz tę opcję, aby określić rozkład a priori własny.

#### **Model nieparametryczny**

Po wybraniu tej opcji dla danych obserwowanych przyjmowany jest model nieparametryczny.

#### **Stałe marginesy**

Wybierz opcję **Suma wiersza** lub **Suma kolumny**, aby określić stałe sumy brzegowe dla tabeli przypadkowej. Ustawieniem domyślnym jest **Suma wiersza**.

### Rozkład a priori

Określ parametry dla rozkładu a priori Dirichleta. Jeśli wybrany jest **Model nieparametryczny**, należy określić parametry **Rozkładu a priori**. Jeśli zostanie podana pojedyncza wartość, przyjmuje się, że wszystkie wartości  $\lambda_s$  są równe tej wartości. Ustawieniem domyślnym jest  $\lambda_s = 1$ . Jeśli wymagane jest podanie więcej niż jednej wartości, można rozdzielić wartości spacjami. Wszystkie podane wartości muszą być  $> 0$ . Liczba określonych, liczbowych wartości musi być zgodna z wymiarem wiersza lub kolumny, która nie jest ustalona dla tabeli kontyngencji.

W celu skasowania wartości kliknij opcję **Resetuj**.

## Bayesowskie modele logliniowe: Drukowanie

Możesz określić sposób wyświetlania zawartości w tabelach wynikowych.

### Projekt tabeli

#### Ukryj tabelę

W przypadku wyboru tej opcji tabela przypadkowa nie jest dołączana do danych wynikowych. Domyślnie to ustawienie nie jest włączone.

**Uwaga:** Następujące ustawienia nie mają wpływu, gdy włączone jest ustawienie **Ukryj tabelę**.

### Statystyki

Określ statystyki w celu testowania niezależności.

#### Chi-kwadrat

Wybierz, aby obliczyć statystykę chi-kwadrat Pearsona, stopnie swobody oraz dwustronną istotność asymptotyczną. W przypadku tabeli przypadkowej  $2 \times 2$  to ustawienie powoduje także obliczenie poprawki Yatesa w kierunku ciągłości, stopni swobody oraz skojarzonej dwustronnej istotności asymptotycznej. W przypadku tabeli przypadkowej  $2$  na  $2$ , która zawiera co najmniej jedną oczekiwaną liczebnością w komórce  $< 5$ , to ustawienie powoduje także obliczenie dokładnej istotności (test 2-stronny i 1-stronny) dokładnego testu Fishera.

#### Iloraz wiarygodności

Wybierz, aby obliczyć statystykę testu ilorazu wiarygodności, stopnie swobody oraz skojarzoną dwustronną istotność asymptotyczną.

### Liczebności

Określ typy liczebności uwzględnione w tabeli przypadkowej.

#### Obserwowane

Wybierz tę opcję, aby w tabeli przypadkowej uwzględnić obserwowane liczebności komórek.

#### Oczekiwana

Wybierz tę opcję, aby w tabeli przypadkowej uwzględnić oczekiwane liczebności komórek.

### Wartość procentowa

Określ typy procentów uwzględnione w tabeli przypadkowej.

#### Wiersz

Wybierz tę opcję, aby w tabeli przypadkowej uwzględnić procenty w wierszach.

#### Kolumna

Wybierz tę opcję, aby w tabeli przypadkowej uwzględnić procenty w kolumnach.

#### Łącznie

Wybierz tę opcję, aby w tabeli przypadkowej uwzględnić procenty łączne.

## Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami: Modele

Ta funkcja wymaga SPSS Statistics produktu Standard Edition lub opcji Advanced Statistics.

W modelach Bayesowskiej jednoczynnikowej analizie wariancji (ANOVA) zakłada się, że istnieje jeden pomiar dla każdego obiektu. Jednak założenie to nie zawsze jest prawdziwe. Stosunkowo często zdarza się, że projekt badania służy do zbadania średniej odpowiedzi w wielu punktach czasowych lub stanach. Procedura Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami mierzy jeden czynnik z tego

samego obiektu w każdym odrębnym punkcie czasowym lub stanie i umożliwiając krzyżowanie obiektów w ramach poziomów. Zakłada się, że każdy obiekt ma jedną obserwację w każdym punkcie czasowym lub stanie (przy czym nie uwzględnia się np. indywidualnej interakcji pacjent-terapia).

1. Z menu wybierz:

**Analizuj > Statystyka bayesowska > Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami**

2. Wybierz co najmniej dwie zmienne **Powtórzone pomiary** z listy **Dostępne zmienne**.

3. Opcjonalnie wybierz z listy **Dostępne zmienne** pojedynczą, zmienną, która będzie pełnić rolę **Wagi** regresji. Pole zmiennej **Waga** może być puste.

**Uwaga:** Lista dostępnych zmiennych udostępni wszystkie zmienne oprócz zmiennych łańcuchowych.

4. Ustaw żadaną **Analizę bayesowską**:

- **Charakteryzuj rozkład a posteriori:** gdy ta opcja jest wybrana, wnioskowanie bayesowskie jest wykonywane z perspektywy uzyskiwanej poprzez charakteryzację rozkładów a posteriori. Brzegowy rozkład posteriori parametrów badanych można sprawdzić poprzez scałkowanie parametrów zakłócających, a następnie opracowanie wiarygodnych przedziałów ufności w celu wnioskowania bezpośredniego. Jest to ustawienie domyślne.
- **Estymuj czynnik Bayesa:** w przypadku wybrania tej opcji estymacja czynnika Bayesa (jednej z istotnych metodologii wnioskowania bayesowskiego) obejmuje współczynnik naturalny w celu porównania prawdopodobieństw brzegowych między hipotezą zerową a alternatywną.

*Tabela 11. Progi powszechnie wykorzystywane do określenia istotności dowodów*

Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu	Czynnik Bayesa	Kategoria dowodu
>100	Dowód dotyczący ekstremum dla H1	1-3	Anegdotyczny dowód na H1	1/30-1/10	Silny dowód na H0
30-100	Bardzo silny dowód dla H1	1	Brak dowodu	1/100-1/30	Bardzo silny dowód dla H0
10-30	Silny dowód na H1	1/3-1	Anegdotyczny dowód na H0	1/100	Skrajnie silny dowód na H0
3-10	Umiarkowanie silny dowód na H1	1/10-1/3	Umiarkowanie silny dowód na H0		

*H0: hipoteza zerowa*

*H1: hipoteza alternatywna*

19

20

- **Użyj obu metod:** w przypadku wyboru tej opcji używane są obie metody wnioskowania **Charakteryzuj rozkład a posteriori** oraz **Estymuj czynnik Bayesa**.

Opcjonalnie można wykonać następujące czynności:

- Kliknij opcję **Kryteria**, aby określić procentowy przedział wiarygodności oraz ustawienia metody numerycznej.

<sup>19</sup> Lee, M.D., and Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>20</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- Kliknij opcję **Czynnik Bayesa**, aby określić ustawienia czynnika Bayesa.
- Kliknij opcję **Wykresy**, aby wykreślić rozkłady a posteriori średnich grupowych.

## Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami: kryteria

Na potrzeby bayesowskiej analizy jednoczynnikowej ANOVA z powtarzanymi pomiarami można podać poniższe kryteria analizy.

### Procentowy przedział wiarygodności %

Określ poziom istotności dla obliczenia przedziałów wiarygodności. Poziomym domyślnym jest 95%.

### Metoda numeryczna

Określ metodę numeryczną, która będzie używana do oszacowania całki.

### Ustaw własną wartość startową

W przypadku wyboru tej opcji można określić własną wartość startową w polu **Wartość startowa**. Wartością domyślną jest 2 000 000. Wartość musi być liczbą całkowitą w przedziale od 1 do 2 147 483 647. Domyślnie przypisana zostanie wartość pseudolosowa.

### Liczba prób Monte Carlo

Określ liczbę punktów, które są próbkowane na potrzeby aproksymacji Monte Carlo. Wartość musi być dodatnią liczbą całkowitą w przedziale od  $10^3$  do  $10^6$ . Wartością domyślną jest 30 000.

## Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami: czynnik Bayesa

W tym przypadku można określić podejście stosowane do estymacji czynnika Bayesa dla Bayesowskich modeli jednoczynnikowych ANOVA z powtarzanymi pomiarami. Następujące opcje są dostępne tylko wtedy, gdy wybrano analizę Bayesowską **Estymuj czynnik Bayesa** lub opcję **Użyj obu metod**.

### Kryterium informacyjne Bayesa (BIC)

Używa rozszerzenia aproksymacji BIC schematu powtarzanych pomiarów do estymacji czynników Bayesa. Ustawienie to wylicza efektywną wielkość próby z uwzględnieniem korelacji powtarzanych pomiarów i sugeruje ulepszony składnik kary, dokonując oszacowania wartości BIC na potrzeby wyboru między dwoma konkurującymi ze sobą modelami. To jest ustawienie domyślne.

### Schemat mieszany Roudera

Używa wielowymiarowych uogólnień rozkładu Cauchy'ego jako prawdopodobieństwa wstępnego dla standaryzowanego rozmiaru efektu oraz jako nieinformacyjnego prawdopodobieństwa wstępnego dla wariancji.

**Uwaga:** Gdy ta opcja jest wybrana, ignorowane jest globalne ustawienie wagi liczebności i wagi regresji.

## Bayesowska jednoczynnikowa ANOVA z powtarzanymi pomiarami: wykresy

Można wpływać na generowane wykresy, aby uwidocznić rozkłady a posteriori średnich grupowych. W tabeli wymienione są wszystkie zmienne, które zostały wybrane jako powtarzane pomiary w oknie dialogowym Zmienne. Wybierz zmienne powtórzonych pomiarów do wykreślenia.

## Regresja grzbietowa jądra

Regresja grzbietowa jądra jest procedurą rozszerzenia, która wykorzystuje klasę Python **sklearn.kernel\_ridge.KernelRidge** do oszacowania modeli regresji grzbietowej jądra. Modele regresji grzbietowej jądra są nieparametrycznymi modelami regresji, które są zdolne do modelowania liniowych i nieliniowych relacji między zmiennymi predyktora i wynikami. Wyniki mogą być bardzo wrażliwe na wybór hiperparametrów modelu. Regresja grzbietowa jądra ułatwia wybór wartości hiperparametrów poprzez k-krotną walidację krzyżową w określonych siatkach wartości przy użyciu klasy **sklearn.model\_selection.GridSearchCV**.

### Przykład

## Statystyka

addytywne CHI2, CHI2, cosinus, laplasjan, liniowy, wielomianowy, RBF, sigmoid, alfa, gamma, Coef0, stopień, walidacja krzyżowa, obserwacje a przewidywania, reszty a przewidywania, podwójne współczynniki wagowe, współczynniki wagowe przestrzeni jądra.

## Zagadnienia dotyczące danych

### Dane

- Możliwe jest określenie dowolnej lub wszystkich ośmiu różnych funkcji jądra.
- Wybrana funkcja jądra określa, które hiperparametry są aktywne.
- Hiperparametry obejmują alfa regularyzację grzbietu, które są wspólne dla wszystkich jąder, plus trzy inne hiperparametry dla każdej konkretnej funkcji jądra.
- Jeśli określono wiele podkomend jądra lub określono więcej niż jedną wartość dla dowolnego parametru, wykonywane jest wyszukiwanie w siatce z walidacją krzyżową w celu oceny modeli, a następnie wybierany jest najlepiej dopasowany model oparty na wstrzymanych danych.
- Rozszerzenie akceptuje zmienne dzielone z procedury Split File (podział pliku) i wagi z wykorzystaniem procedury Weight Cases (ważenie obserwacji).
- Podczas uwzględniania wag są one używane do tworzenia dopasowanych wartości we wszystkich analizach. Ze względu na ograniczenia w metodzie oceny w klasie **sklearn.model\_selection.GridSearchCV** oceny walidacji krzyżowej, które są używane do wyboru modelu, nie są ważone.

### Założenia

## Uzyskiwanie regresji grzbietowej jądra

1. Wybierz z menu następującą opcję:

**Analizuj > Regresja > Grzbiet jądra...**

2. Wybierz zmienną **zależną**.

3. Wybierz przynajmniej jedną zmienną **niezależną**.

4. Domyślne ustawienie **Pojedynczy model** jest używane tylko wtedy, gdy określono tylko jedną wartość dla każdego parametru funkcji jądra. Jeśli wybrano ustawienie **Pojedynczy model**, nie można określić dodatkowych funkcji i wag **jądra**, które są w pełni stosowane w całej analizie, wartościowaniu i ocenianiu wyników. Można również użyć przycisków strzałek w górę i w dół, aby zmienić kolejność funkcji jądra.

Opcjonalnie wybierz opcję **Wybór modelu** z listy **Tryb**.

Po wybraniu opcji **Wybór modelu** z listy **Tryb** można dodać wiele funkcji jądra do listy **Jądro**.

a. Kliknij ikonę dodawania (+), aby dołączyć dodatkowe funkcje jądra.

b. Aby wybrać funkcję jądra, należy kliknąć pustą komórkę w kolumnie **Jądro**.

c. Dwukrotnie kliknij dowolną komórkę wiersza funkcji jądra, aby określić wartości parametrów funkcji jądra dla odpowiedniej kolumny (**Alfa**, **Gamma**, **Coef0**, **Stopień**). Więcej informacji na ten temat zawiera sekcja "Parametry jądra" na stronie 139. Poniżej wymieniono domyślne parametry dostrajania funkcji jądra.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

#### **Cosinus**

ALPHA=1

#### **Laplasjan**

ALPHA=1 GAMMA=1/p



**Liniowe**

Domyślna funkcja jądra. ALPHA=1

**Wielomianowe**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

**RBF**

ALPHA=1 GAMMA=1/p

**Sigmoid**

ALPHA=1 GAMMA=1/p COEF0=1

**Uwaga:** Jeśli określono więcej niż jedną wartość dla dowolnego parametru funkcji jądra, przeprowadzane jest wyszukiwanie w siatce z walidacją krzyżową w celu oceny modeli, a następnie wybierany jest najlepszy model dopasowania oparty na wstrzymanych danych.

5. Opcjonalnie można kliknąć opcję **Opcje**, aby określić krotność walidacji krzyżowej, opcje wyświetlania, ustawienia wykresu i elementy do zapisania. Więcej informacji na ten temat zawiera sekcja [“Regresja grzbietowa jądra: Opcje”](#) na stronie 139.
6. Kliknij **OK**.

## Parametry jądra

Okno dialogowe **Parametry jądra** udostępnia opcje służące do określania wartości parametrów funkcji pojedynczego jądra oraz do określania, że wybór modelu jest wykonywany przy użyciu funkcji wyszukiwania siatki w kombinacjach jądra i określonych wartości parametrów siatki.

**Określ pojedyncze parametry**

Włącz to ustawienie, aby określić wartości dla wybranego parametru funkcji jądra.

- Wprowadź wartość i kliknij przycisk **Dodaj**, aby dołączyć wartość do parametru funkcji jądra.
- Wybierz wartość parametru i kliknij przycisk **Zmień**, aby zaktualizować wartość.
- Wybierz wartość parametru i kliknij przycisk **Usuń**, aby usunąć wartość.

**Określ parametry siatki**

Włącz to ustawienie, aby określić, że wybór modelu jest wykonywany przy użyciu wyszukiwania siatki w kombinacjach jądra i określonych wartości parametrów siatki.

## Regresja grzbietowa jądra: Opcje

W oknie dialogowym **Wykresy** dostępne są opcje służące do określania krotności walidacji krzyżowej, opcji wyświetlania, ustawień wykresu oraz elementów do zapisania.

**Wartość krotności walidacji krzyżowej**

Liczba podziałów lub krotności w walidacji krzyżowej z wyszukiwaniem siatki w celu wyboru modelu. Wprowadź liczbę całkowitą większą od 1. Wartością domyślną jest 5. Ustawienie jest dostępne tylko wtedy, gdy **Wybór modelu** jest wybierany jako **Tryb** w podstawowym oknie dialogowym **Regresja grzbietowej jądra**.

**Wyświetl**

Udostępnia opcje służące do określania, które dane wyjściowe mają być wyświetlane w przypadku, gdy walidacja krzyżowa jest aktywna.

**Najlepiej**

Ustawienie domyślne wyświetla tylko podstawowe wyniki dla wybranego najlepszego modelu.

**Porównaj**

Wyświetla podstawowe wyniki dla wszystkich ocenionych modeli.

**Porównywanie modeli i krotności**

Wyświetla pełne wyniki dla każdego podziału lub krotności dla każdego ocenianego modelu.

**Wykres**

Udostępnia opcje określające wykresy wartości obserwowanych lub rezydualnych w porównaniu z wartościami przewidywanymi.

**Obserwowane a przewidywane**

Wyświetla wykres rozrzutu obserwowanych i przewidywanych wartości dla określonego lub najlepszego modelu.

**Reszty a przewidywane**

Wyświetla wykres rozrzutu reszt w porównaniu z wartościami przewidywanymi dla określonego lub najlepszego modelu.

**Zapisz**

Tabela zawiera opcje umożliwiające określenie zmiennych, które mają zostać zapisane w aktywnym zbiorze danych.

**Wartości przewidywane**

Zapisuje przewidywane wartości z określonego lub najlepszego modelu w aktywnym zbiorze danych. Można dołączyć opcjonalną nazwę zmiennej.

**Reszty**

Umożliwia zapisanie reszt z określonych lub najlepszych predykcji modelu do aktywnego zbioru danych. Można dołączyć opcjonalną nazwę zmiennej.

**Podwójne współczynniki**

Umożliwia zapisanie podwójnych współczynników lub współczynników wagowych przestrzeni jądra z określonego modelu do aktywnego zbioru danych. Można dołączyć opcjonalną nazwę zmiennej. Ustawienie jest dostępne tylko wtedy, gdy określono opcję **Wybór modelu** dla opcji **Tryb** w podstawowym oknie dialogowym **Regresja grzbietowa jądra**.

## 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 przysłanych 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](http://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

- analiza kowariancji
  - w procedurze OML wielu zmiennych [1](#)
- analiza logliniowa
  - Analiza logliniowa Logit [71](#)
  - Ogólna analiza log-liniowa [68](#)
  - w uogólnionych liniowych modelach mieszanych [51](#)
- Analiza logliniowa Logit
  - czynniki [71](#)
  - kontrasty [71](#)
  - kryteria [73](#)
  - określanie modelu [72](#)
  - opcje wyświetlania [73](#)
  - przedziały ufności [73](#)
  - reszty [73](#)
  - rozkład liczebności komórek [71](#)
  - struktury komórek [71](#)
  - wartości przewidywane [73](#)
  - współzienne komórek [71](#)
  - wykresy [73](#)
  - zapisywanie zmiennych [73](#)
- Analiza logliniowa: Wybór modelu
  - definiowanie zakresów czynników [67](#)
  - dodatkowe właściwości komendy [68](#)
  - modele [67](#)
  - opcje [68](#)
- analiza probit
  - uogólnione liniowe modele mieszane [51](#)
- analiza przeżycia
  - regresja Coxa zależna od czasu [99](#)
  - w metodzie Kaplana-Meiera [93](#)
  - w regresji Coxa [95](#)
  - w regresji grzbietowej jądra [137](#)
  - w tabelach trwania życia [74](#)
- analiza wariancji
  - w składowych wariancji [20](#)
  - w uogólnionych liniowych modelach mieszanych [51](#)
- ANOVA
  - w procedurze OML powtarzanych pomiarów [9](#)
  - w procedurze OML wielu zmiennych [1](#)
- ANOVA wielu zmiennych [1](#)

## B

- błąd standardowy
  - w OML [7](#)
  - w procedurze OML powtarzanych pomiarów [16](#)
- Bonferroni
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- budowanie składników [3](#), [12](#), [20](#), [67](#), [70](#), [73](#)

## C

- C Dunnetta
  - w procedurze OML powtarzanych pomiarów [15](#)

- C Dunnetta (*kontynuacja*)
  - w procedurze OML wielu zmiennych [5](#)
- częstości
  - w analizie logliniowej: wybór modelu [68](#)
- czynniki
  - w procedurze OML powtarzanych pomiarów [12](#)
- czynniki i poziomy
  - w liniowych modelach mieszanych [27](#)

## D

- dobroć dopasowania
  - w ogólnych równaniach estymujących [48](#)
  - w uogólnionych modelach liniowych [35](#)

## E

- efekty losowe
  - w liniowych modelach mieszanych [25](#), [29](#)
- efekty losowe a priori
  - w składowych wariancji [20](#)
- efekty stałe
  - w liniowych modelach mieszanych [24](#)
- eliminacja wsteczna
  - w analizie logliniowej: wybór modelu [66](#)
- estymacja metodą największej wiarygodności
  - w składowych wariancji [20](#)
- estymacja metodą warunkowej największej wiarygodności
  - w składowych wariancji [20](#)

## F

- F R-E-G-W
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- funkcja łączenia
  - uogólnione liniowe modele mieszane [54](#)
- funkcja łączenia komplementarny log-log
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia logarytmiczna dopełnienia
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia logit
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia nieparzysty potęgowy
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia probit
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia skumulowany Cauchit
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)
- funkcja łączenia skumulowany komplementarny log-log

- funkcja łączenia skumulowany komplementarny log-log (*kontynuacja*)
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja łączenia skumulowany probit
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja łączenia skumulowany ujemny log-log
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja łączenia ujemny dwumianowy
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja łączenia ujemny log-log
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja łączenia ujemny skumulowany logit
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - funkcja przeżycia
    - w tabelach trwania życia [74](#)
- G**
- GLOR
    - w ogólnej analizie logliniowej [68](#)
  - Grzbiet jądra
    - alfa [137](#)
    - coef0 [137](#)
    - gamma [137](#)
    - pojedynczy model [137](#)
    - stopień [137](#)
    - wybór modelu [137](#)
  - GT2 Hochberga
    - w procedurze OML powtarzanych pomiarów [15](#)
    - w procedurze OML wielu zmiennych [5](#)
- H**
- hierarchiczna dekompozycja
    - w składowych wariancji [21](#)
  - hierarchiczne modele
    - uogólnione liniowe modele mieszane [51](#)
  - hierarchiczne modele nieliniowe [66](#)
- I**
- iloraz szans
    - w ogólnej analizie logliniowej [68](#)
  - informacja o analizowanych danych
    - w ogólnych równaniach estymujących [48](#)
    - w uogólnionych modelach liniowych [35](#)
  - informacje o modelu
    - w ogólnych równaniach estymujących [48](#)
    - w uogólnionych modelach liniowych [35](#)
  - iteracje
    - w analizie logliniowej: wybór modelu [68](#)
    - w ogólnych równaniach estymujących [46](#)
    - w uogólnionych modelach liniowych [34](#)
- K**
- Kaplan-Meier
    - liniowy trend poziomów czynnika [94](#)
  - Kaplan-Meier (*kontynuacja*)
    - porównywanie poziomów czynników [94](#)
  - Kaplan-Meier (KM)
    - definiowanie zdarzeń [94](#)
    - dodatkowe właściwości komendy [95](#)
    - kwartyle [95](#)
    - przykład [93](#)
    - statystyki [93, 95](#)
    - średnia i mediana czasu przeżycia [95](#)
    - tabele przeżycia [95](#)
    - wykresy [95](#)
    - zapisywanie zmiennych wynikowych [95](#)
    - zmiennie statusu przeżycia [94](#)
  - kategoria odniesienia
    - w ogólnych równaniach estymujących [44, 45](#)
    - w uogólnionych modelach liniowych [32](#)
  - klasa tworząca
    - w analizie logliniowej: wybór modelu [67](#)
  - kontrasty
    - w analizie logliniowej Logit [71](#)
    - w ogólnej analizie logliniowej [68](#)
    - w regresji Coxa [96](#)
  - kroki połowienia
    - w liniowych modelach mieszanych [26](#)
    - w ogólnych równaniach estymujących [46](#)
    - w uogólnionych modelach liniowych [34](#)
- L**
- Liniowe modele mieszane
    - budowanie składników [24, 25](#)
    - dodatkowe właściwości komendy [29](#)
    - efekty losowe [25, 29](#)
    - efekty stałe [24](#)
    - kryteria oszacowania [26](#)
    - model [27](#)
    - składniki interakcji [24](#)
    - struktura kowariancji [103](#)
    - szacowane średnie brzegowe [28](#)
    - zapisywanie zmiennych [28](#)
  - logarytmiczna funkcja łączenia
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
- M**
- macierz korelacji
    - w liniowych modelach mieszanych [27](#)
    - w ogólnych równaniach estymujących [48](#)
    - w uogólnionych modelach liniowych [35](#)
  - macierz kowariancji
    - w liniowych modelach mieszanych [27](#)
    - w ogólnych równaniach estymujących [46, 48](#)
    - w OML [7](#)
    - w uogólnionych modelach liniowych [34, 35](#)
  - macierz kowariancji efektów losowych
    - w liniowych modelach mieszanych [27](#)
  - macierz kowariancji parametrów
    - w liniowych modelach mieszanych [27](#)
  - macierz kowariancji reszt
    - w liniowych modelach mieszanych [27](#)
  - macierz L
    - w uogólnionych modelach liniowych [35](#)



Macierz L  
w ogólnych równaniach estymujących [48](#)  
macierz współczynników kontrastów  
w ogólnych równaniach estymujących [48](#)  
w uogólnionych modelach liniowych [35](#)  
metoda Newtona-Raphsona  
w analizie logliniowej Logit [71](#)  
w ogólnej analizie logliniowej [68](#)  
MINQUE  
w składowych wariancji [20](#)  
model ryzyk proporcjonalnych  
w regresji Coxa [95](#)  
modele mieszane  
liniowe [22](#)  
uogólnione liniowe modele mieszane [51](#)  
modele nasycone  
w analizie logliniowej: wybór modelu [67](#)  
Modele parametryczne Frailty  
zmienne statusu przeżycia [89](#)  
Modele parametrycznego czasu przyspieszonego  
niepowodzenia  
behawioralne [76](#)  
Dokonaj predykcji [80](#)  
drukowanie [79](#)  
eksport [81](#)  
kryteria [77](#)  
Oszacowanie [78](#)  
usługowym [77](#)  
wykres [81](#)  
modele użytkownika  
w analizie logliniowej: wybór modelu [67](#)  
w procedurze OML powtarzanych pomiarów [12](#)  
w składowych wariancji [19](#)

## N

najmniejsza istotna różnica  
w procedurze OML powtarzanych pomiarów [15](#)  
w procedurze OML wielu zmiennych [5](#)  
Newman-Keuls  
w procedurze OML powtarzanych pomiarów [15](#)  
w procedurze OML wielu zmiennych [5](#)  
NIR Fishera  
w procedurze OML powtarzanych pomiarów [15](#)  
w procedurze OML wielu zmiennych [5](#)  
normalne wykresy prawdopodobieństwa  
w analizie logliniowej: wybór modelu [68](#)

## O

obserwacje ocenzone  
w metodzie Kaplana-Meiera [93](#)  
w regresji Coxa [95](#)  
w tabelach trwania życia [74](#)  
Ocena Fishera  
w liniowych modelach mieszanych [26](#)  
ocena hazardu  
w tabelach trwania życia [74](#)  
ocenie  
w liniowych modelach mieszanych [26](#)  
Odległość Cooka  
w OML [7](#)  
w procedurze OML powtarzanych pomiarów [16](#)

Odległość Cooka (*kontynuacja*)  
w uogólnionych modelach liniowych [38](#)  
odwrócony rozkład Gaussa  
w ogólnych równaniach estymujących [42](#)  
w uogólnionych modelach liniowych [29](#)  
Ogólna analiza log-liniowa  
czynniki [68](#)  
dodatkowe właściwości komendy [71](#)  
kontrasty [68](#)  
kryteria [70](#)  
określanie modelu [69](#)  
opcje wyświetlania [70](#)  
przedziały ufności [70](#)  
reszty [70](#)  
rozkład liczebności komórek [68](#)  
struktury komórek [68](#)  
współzienne komórek [68](#)  
wykresy [70](#)  
zapisywanie wartości przewidywanych [70](#)  
zapisywanie zmiennych [70](#)  
ogólna funkcja estymowalna  
w ogólnych równaniach estymujących [48](#)  
ogólne funkcje estymowalne  
w uogólnionych modelach liniowych [35](#)  
Ogólne równania estymujące  
eksport modelu [50](#)  
kategoria odniesienia dla odpowiedzi binarnej [44](#)  
odpowiedź [44](#)  
określanie modelu [45](#)  
opcje czynników jakościowych [45](#)  
statystyki [48](#)  
szacowane średnie brzegowe [49](#)  
typ modelu [42](#)  
wartości początkowe [47](#)  
zapisywanie zmiennych w aktywnym zbiorze danych [50](#)  
OML powtarzanych pomiarów  
definiowanie czynników [12](#)  
dodatkowe właściwości komendy [18](#)  
model [12](#)  
testy post hoc [15](#)  
wykresy profili [14](#)  
zapisywanie zmiennych [16](#)  
OML wielu zmiennych  
czynniki [1](#)  
testy post hoc [5](#)  
współzienne [1](#)  
wykresy profili [5](#)  
zmienna zależna [1](#)  
oszacowania parametrów  
w analizie logliniowej Logit [71](#)  
w analizie logliniowej: wybór modelu [68](#)  
w liniowych modelach mieszanych [27](#)  
w ogólnej analizie logliniowej [68](#)  
w ogólnych równaniach estymujących [48](#)  
w uogólnionych modelach liniowych [35](#)

## P

parametr skali  
w ogólnych równaniach estymujących [46](#)  
w uogólnionych modelach liniowych [34](#)  
Parametryczne współużytkowane modele Frailty  
behawioralne [83](#)  
Dokonaj predykcji [87](#)

- Parametryczne współzależne modele Frailty (*kontynuacja*)
    - drukowanie [86](#)
    - eksport [88](#)
    - kryteria [83](#)
    - Oszacowanie [85](#)
    - usługowym [84](#)
    - wykres [88](#)
  - pełne modele czynnikowe
    - w procedurze OML powtarzanych pomiarów [12](#)
    - w składowych wariancji [19](#)
  - posegmentowane współzależne zmienne od czasu
    - w regresji Coxa [99](#)
  - potęgowa funkcja łączenia
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - przebieg iteracji
    - w liniowych modelach mieszanych [26](#)
    - w ogólnych równaniach estymujących [48](#)
    - w uogólnionych modelach liniowych [35](#)
  - przedziały ufności
    - w analizie logliniowej Logit [73](#)
    - w liniowych modelach mieszanych [27](#)
    - w ogólnej analizie logliniowej [70](#)
  - przewidywane wartości stałe
    - w liniowych modelach mieszanych [28](#)
  - Przeżycie AFT
    - okno dialogowe przeżycia-zmienne kategorii [82](#)
- Q**
- Q R-E-G-W
    - w procedurze OML powtarzanych pomiarów [15](#)
    - w procedurze OML wielu zmiennych [5](#)
- R**
- Regresja Coxa
    - definiowanie zdarzenia [98](#)
    - DfBety [97](#)
    - dodatkowe właściwości komendy [98](#)
    - funkcja hazardu [97](#)
    - funkcja przeżycia [97](#)
    - funkcje bazowe [98](#)
    - iteracje [98](#)
    - kontrasty [96](#)
    - krokowe wprowadzanie i usuwanie [98](#)
    - przykład [95](#)
    - reszty cząstkowe [97](#)
    - statystyki [95, 98](#)
    - współzależne [95](#)
    - współzależne jakościowe [96](#)
    - współzależne łańcuchowe [96](#)
    - współzależne zależne od czasu [99](#)
    - wykresy [97](#)
    - zapisywanie zmiennych wyników [97](#)
    - zmienna statusu przeżycia [98](#)
  - Regresja grzbietowa jądra
    - krotność walidacji krzyżowej [139](#)
    - parametry [139](#)
    - parametry siatki [139](#)
    - wykresy [139](#)
    - wyświetlanie [139](#)
  - Regresja grzbietowa jądra (*kontynuacja*)
    - zapisywanie [139](#)
  - regresja logistyczna
    - uogólnione liniowe modele mieszane [51](#)
  - regresja Poissona
    - uogólnione liniowe modele mieszane [51](#)
    - w ogólnej analizie logliniowej [68](#)
  - regresja wielu zmiennych [1](#)
  - Reszta Pearsona
    - w ogólnych równaniach estymujących [50](#)
    - w uogólnionych modelach liniowych [38](#)
  - reszty
    - w analizie logliniowej Logit [73](#)
    - w analizie logliniowej: wybór modelu [68](#)
    - w liniowych modelach mieszanych [28](#)
    - w ogólnej analizie logliniowej [70](#)
    - w ogólnych równaniach estymujących [50](#)
    - w uogólnionych modelach liniowych [38](#)
  - reszty niestandardyzowane
    - w OML [7](#)
    - w procedurze OML powtarzanych pomiarów [16](#)
  - reszty odchylenia
    - w uogólnionych modelach liniowych [38](#)
  - reszty standaryzowane
    - w OML [7](#)
    - w procedurze OML powtarzanych pomiarów [16](#)
  - reszty usuniętych
    - w OML [7](#)
    - w procedurze OML powtarzanych pomiarów [16](#)
  - reszty wiarygodności
    - w uogólnionych modelach liniowych [38](#)
  - rozkład dwumianowy
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład dwumianowy ujemny
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład gamma
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład normalny
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład Poissona
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład Tweediego
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rozkład wielomianowy
    - w ogólnych równaniach estymujących [42](#)
    - w uogólnionych modelach liniowych [29](#)
  - rzeczywiście istotna różnica Tukey'a
    - w procedurze OML powtarzanych pomiarów [15](#)
    - w procedurze OML wielu zmiennych [5](#)
- S**
- separacja
    - w ogólnych równaniach estymujących [46](#)
    - w uogólnionych modelach liniowych [34](#)
  - składniki interakcji
    - w liniowych modelach mieszanych [24](#)
  - składniki zagnieżdżone

- składniki zagnieżdżone (*kontynuacja*)
  - w liniowych modelach mieszanych [25](#)
  - w ogólnych równaniach estymujących [45](#)
  - w uogólnionych modelach liniowych [33](#)
- Składowe wariancje
  - dodatkowe właściwości komendy [22](#)
  - model [19](#)
  - opcje [20](#)
  - zapisywanie wyników [21](#)
- statystyka Walda
  - w analizie logliniowej Logit [71](#)
  - w ogólnej analizie logliniowej [68](#)
- statystyki opisowe
  - w liniowych modelach mieszanych [27](#)
  - w ogólnych równaniach estymujących [48](#)
  - w uogólnionych modelach liniowych [35](#)
- struktury kowariancji
  - w liniowych modelach mieszanych [103](#)
- Student-Newman-Keuls
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- suma kwadratów
  - w składowych wariancji [21](#)
- sumy kwadratów
  - w liniowych modelach mieszanych [25](#)
- szacowane średnie brzegowe
  - w liniowych modelach mieszanych [28](#)
  - w ogólnych równaniach estymujących [49](#)
  - w uogólnionych modelach liniowych [37](#)

## T

- T2 Tamhane'a
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- T3 Dunnetta
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- tabela krzyżowa
  - w analizie logliniowej: wybór modelu [66](#)
- tabele kontyngencji
  - w ogólnej analizie logliniowej [68](#)
- Tabele trwania życia
  - czynniki [75](#)
  - dodatkowe właściwości komendy [76](#)
  - funkcja przeżycia [74](#)
  - ocena hazardu [74](#)
  - porównywanie poziomów czynników [75](#)
  - przykład [74](#)
  - statystyki [74](#)
  - test Wilcoxon (Gehana) [75](#)
  - ukrywanie tabel [75](#)
  - wykresy [75](#)
  - zmiennie statusu przeżycia [82](#)
- test b Tukey'a
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test Breslowa
  - w metodzie Kaplana-Meiera [94](#)
- test Dunnetta
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test Gabriela porównań parami
  - w procedurze OML powtarzanych pomiarów [15](#)

- test Gabriela porównań parami (*kontynuacja*)
  - w procedurze OML wielu zmiennych [5](#)
- test Gamesa i Howella porównań parami
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test Gehana
  - w tabelach trwania życia [75](#)
- test mnożnika Lagrange'a
  - w uogólnionych modelach liniowych [35](#)
- test parametrów kowariancji
  - w liniowych modelach mieszanych [27](#)
- test rang logarytmicznych
  - w metodzie Kaplana-Meiera [94](#)
- test Scheffégo
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test t Sidaka
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test t Wallera-Duncana
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test Tarone'a-Ware'a
  - w metodzie Kaplana-Meiera [94](#)
- test wielokrotnych rozstępów Duncana
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- test Wilcoxon
  - w tabelach trwania życia [75](#)
- tolerancja osobliwości
  - w liniowych modelach mieszanych [26](#)
- tożsamościowa funkcja łączenia
  - w ogólnych równaniach estymujących [42](#)
  - w uogólnionych modelach liniowych [29](#)

## U

- uogólnione liniowe modele mieszane
  - blok efektów losowych [58](#)
  - efekty losowe [57](#)
  - efekty stałe [56](#), [63](#)
  - eksport modelu [61](#), [62](#)
  - funkcja łączenia [54](#)
  - kowariancje efektów losowych [64](#)
  - oszacowane średnie [65](#)
  - parametry kowariancji [64](#)
  - podsumowanie modelu [62](#)
  - Przesunięcie [59](#)
  - przewidywane przez obserwowane [63](#)
  - rozkład zmiennej przewidywanej [54](#)
  - składniki zdefiniowane przez użytkownika [57](#)
  - struktura danych [62](#)
  - szacowane średnie brzegowe [60](#)
  - tabela klasyfikacji [63](#)
  - waga analizy [59](#)
  - widok modelu [62](#)
  - współczynniki stałe [64](#)
  - zapisywanie pól [61](#), [62](#)
- uogólnione modele liniowe
  - eksport modelu [39](#)
  - kategoria odniesienia dla odpowiedzi binarnej [32](#)
  - kryteria oszacowania [34](#)
  - opcje czynników jakościowych [33](#)
  - statystyki [35](#)

- uogólnione modele liniowe (*kontynuacja*)
  - szacowane średnie brzegowe [37](#)
- Uogólnione modele liniowe
  - funkcja łączenia [29](#)
  - odpowiedź [32](#)
  - określanie modelu [33](#)
  - predyktory [33](#)
  - rozkład [29](#)
  - typy modelu [29](#)
  - Wartości początkowe [35](#)
  - zapisywanie zmiennych w aktywnym zbiorze danych [38](#)
- Uogólnione równania estymujące
  - kryteria oszacowania [46](#)
  - predyktory [45](#)
- uogólniony iloraz logarytmów szans
  - w ogólnej analizie logliniowej [68](#)
- uogólniony model liniowy
  - uogólnione liniowe modele mieszane [51](#)
  - w uogólnionych liniowych modelach mieszanych [51](#)
- Uogólniony model liniowy (OLM)
  - zapisywanie macierzy [7](#)
  - zapisywanie zmiennych [7](#)

## W

- wartości przewidywane
  - w analizie logliniowej Logit [73](#)
  - w liniowych modelach mieszanych [28](#)
  - w ogólnej analizie logliniowej [70](#)
- wartości wpływu
  - w OML [7](#)
  - w procedurze OML powtarzanych pomiarów [16](#)
  - w uogólnionych modelach liniowych [38](#)
- ważone wartości przewidywane
  - w OML [7](#)
  - w procedurze OML powtarzanych pomiarów [16](#)
- widok modelu
  - w uogólnionych liniowych modelach mieszanych [62](#)
- wielokrotne F Ryana-Einota-Gabriela-Welscha
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- wielokrotny rozstęp Ryana-Einota-Gabriela-Welscha
  - w procedurze OML powtarzanych pomiarów [15](#)
  - w procedurze OML wielu zmiennych [5](#)
- wielomianowa regresja logistyczna
  - uogólnione liniowe modele mieszane [51](#)
- wielomianowe modele logit [71](#)
- wielopoziomowe modele
  - uogólnione liniowe modele mieszane [51](#)
- współzmiennie
  - w regresji Coxa [96](#)
- współzmiennie łańcuchowe
  - w regresji Coxa [96](#)
- wykresy
  - w analizie logliniowej Logit [73](#)
  - w ogólnej analizie logliniowej [70](#)
- wykresy profili
  - w procedurze OML powtarzanych pomiarów [14](#)
  - w procedurze OML wielu zmiennych [5](#)
- wzdłużne modele
  - uogólnione liniowe modele mieszane [51](#)

## Z

- Zbieżność Hessego
  - w ogólnych równaniach estymujących [46](#)
  - w uogólnionych modelach liniowych [34](#)
- zbieżność logarytmu wiarygodności
  - w liniowych modelach mieszanych [26](#)
  - w ogólnych równaniach estymujących [46](#)
  - w uogólnionych modelach liniowych [34](#)
- zbieżność parametru
  - w liniowych modelach mieszanych [26](#)
  - w ogólnych równaniach estymujących [46](#)
  - w uogólnionych modelach liniowych [34](#)
- zmiennie obiektowe
  - w liniowych modelach mieszanych [23](#)
- zmiennie powtórzonych pomiarów
  - w liniowych modelach mieszanych [23](#)



