

IBM SPSS Categories 29

IBM

Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 45.

Informacje o produkcie

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

© Copyright International Business Machines Corporation .

Spis treści

Rozdział 1. Kategorie.....	1
Wprowadzenie do procedur optymalnego skalowania danych jakościowych.....	1
Co to jest skalowanie optymalne?.....	1
Dlaczego warto używać skalowania optymalnego?.....	1
Poziom skalowania optymalnego i poziom pomiaru.....	2
Która procedura jest najlepsza dla danej aplikacji?.....	5
Współczynnik proporcji w wykresach skalowania optymalnego.....	10
Regresja dla danych jakościowych (CATREG).....	10
Definiowanie skali w regresji dla danych jakościowych.....	11
Dyskretyzacja regresji dla danych jakościowych.....	12
Braki danych w regresji dla danych jakościowych.....	12
Opcje regresji dla danych jakościowych.....	12
Regresja dla danych jakościowych: Regularyzacja.....	13
Wyniki regresji dla danych jakościowych.....	14
Regresja dla danych jakościowych: Zapisz.....	14
Regresja dla danych jakościowych: Wykresy przekształceń.....	15
Dodatkowe właściwości komendy CATREG.....	15
Analiza głównych składowych dla danych jakościowych (CATPCA).....	15
Definiowanie skali i wagi w procedurze CATPCA.....	16
Analiza głównych składowych dla danych jakościowych: Dyskretyzacja.....	17
Analiza głównych składowych dla danych jakościowych: Braki danych.....	17
Opcje analizy głównych składowych dla danych jakościowych.....	18
Analiza głównych składowych dla danych jakościowych: Wyniki.....	19
Analiza głównych składowych dla danych jakościowych: Zapisz.....	20
Analiza głównych składowych dla danych jakościowych: Wykresy obiektów.....	20
Analiza głównych składowych dla danych jakościowych (CATPCA): Wykresy kategorii.....	20
Wykresy ładunków w analizie głównych składowych danych jakościowych.....	21
Analiza głównych składowych dla danych jakościowych: Metoda bootstrap.....	21
Dodatkowe właściwości komendy CATPCA.....	22
Nieliniowa analiza korelacji kanonicznych (OVERALS).....	22
Definiuj zakres i typ skali.....	24
Definiuj zakres.....	24
Nieliniowa analiza korelacji kanonicznych: Opcje.....	24
Dodatkowe właściwości komendy OVERALS.....	25
Analiza korespondencji.....	25
Definiowanie zakresu zmiennej wierszowej w analizie korespondencji.....	26
Definiowanie zakresu kolumn w analizie korespondencji.....	27
Analiza korespondencji: Model.....	27
Analiza korespondencji: Statystyki.....	28
Analiza korespondencji: Wykresy.....	28
Dodatkowe właściwości komendy CORRESPONDENCE.....	29
Wielowymiarowa analiza korespondencji.....	29
Definiowanie wagi zmiennej w wielowymiarowej analizie korespondencji.....	30
Dyskretyzacja wielowymiarowej analizy korespondencji.....	30
Wielowymiarowa analiza korespondencji (MCA) — Braki danych.....	31
Wielowymiarowa analiza korespondencji (MCA) — Opcje.....	31
Wielowymiarowa analiza korespondencji (MCA) — Wynik.....	32
Wielowymiarowa analiza korespondencji (MCA) — Zapisz.....	33
Wielowymiarowa analiza korespondencji (MCA) — Wykresy obiektów.....	33
Wielowymiarowa analiza korespondencji (MCA) — Wykresy zmiennych.....	33
Dodatkowe właściwości komendy MULTIPLE CORRESPONDENCE.....	33

Skalowanie wielowymiarowe (PROXSCAL).....	34
Odległości w macierzach rozłożonych na kolumny.....	35
Bliskości w kolumnach.....	35
Odległości w jednej kolumnie.....	35
Tworzenie bliskości na podstawie danych.....	36
Tworzenie miary na podstawie danych.....	36
Definiowanie modelu skalowania wielowymiarowego.....	36
Ograniczenia skalowania wielowymiarowego.....	37
Skalowanie wielowymiarowe: Opcje.....	38
Wykresy skalowania wielowymiarowego, wersja 1.....	38
Wykresy skalowania wielowymiarowego, wersja 2.....	39
Skalowanie wielowymiarowe: Opcje.....	39
Dodatkowe właściwości komendy PROXSCAL.....	40
Wielowymiarowe skalowanie preferencji (PREFSCAL).....	40
Definiowanie modelu wielowymiarowego skalowania preferencji.....	41
Ograniczenia wielowymiarowego skalowania preferencji.....	41
Wielowymiarowe skalowanie preferencji: Opcje.....	42
Wielowymiarowe skalowanie preferencji: Wykresy.....	43
Wielowymiarowe skalowanie preferencji: Wynik.....	43
Dodatkowe właściwości komendy PREFSCAL.....	44
Uwagi.....	45
Znaki towarowe.....	46
Indeks.....	49

Rozdział 1. Kategorie

Następujące funkcje związane z kategoriami są dostępne w SPSS Statistics produktu Professional Edition oraz w module Kategorie.

Wprowadzenie do procedur optymalnego skalowania danych jakościowych

W procedurach operujących na kategoriach stosuje się techniki optymalnego skalowania, aby analizować dane trudne lub niemożliwe do przeanalizowania przy użyciu standardowych procedur statystycznych. W niniejszym rozdziale opisano zadania realizowane przez poszczególne procedury, sytuacje, w których najbardziej celowe jest zastosowanie poszczególnych procedur, relacje między procedurami oraz relacje między tymi procedurami a ich standardowymi odpowiednikami statystycznymi.

Uwaga: Opisane tutaj procedury i ich implementacje w produkcie IBM SPSS Statistics zostały opracowane przez zespół Data Theory Scaling System Group (DTSS), w skład którego wchodzi pracownicy departamentów edukacji i psychologii Wydziału Nauk Społecznych i Behawioralnych Uniwersytetu w Lejdzie.

Co to jest skalowanie optymalne?

Celem skalowania optymalnego jest przypisanie kategoriom każdej zmiennej kwantyfikacji liczbowych, umożliwiając tym samym wykorzystanie standardowych procedur do otrzymania rozwiązania na podstawie kwantyfikowanych zmiennych.

Wartości skali optymalnej są przypisywane do kategorii każdej zmiennej na podstawie kryterium optymalizacji używanej procedury. W przeciwieństwie do oryginalnych etykiet zmiennej nominalnej lub porządkowej te wartości skali mają własności metryczne.

W przypadku większości procedur dostępnych w module Kategorie optymalna kwantyfikacja każdej skalowanej zmiennej jest wykonywana za pomocą iteracyjnej metody **przebiegających najmniejszych kwadratów**, w której bieżące kwantyfikacje wykorzystane do odnalezienia rozwiązania są aktualizowane na podstawie tego rozwiązania. Zaktualizowane kwantyfikacje są następnie używane do odnalezienia nowego rozwiązania, na podstawie którego są aktualizowane kwantyfikacje i tak dalej, aż do osiągnięcia kryterium zatrzymującego proces.

Dlaczego warto używać skalowania optymalnego?

Dane jakościowe często występują w badaniach marketingowych, badaniach ankietowych oraz badaniach w naukach społecznych i behawioralnych. W rzeczywistości wielu badaczy pracuje prawie wyłącznie na danych jakościowych.

Adaptacje większości standardowych modeli powstały specjalnie w celu analizy danych jakościowych, jednak nie sprawdzają się one w przypadku zestawów danych, które charakteryzują się następującymi cechami:

- Zbyt mało obserwacji
- Zbyt wiele zmiennych
- Zbyt wiele wartości dla jednej zmiennej

Skorzystanie z technik skalowania optymalnego, które kwantyfikują kategorie, pozwala uniknąć problemów w takich przypadkach. Ponadto są one przydatne, nawet gdy odpowiednie jest skorzystanie z wyspecjalizowanych technik.

Interpretacja wyników skalowania optymalnego często jest oparta na graficznych przedstawieniach danych, a nie na interpretacji oszacowań parametrów. Techniki skalowania optymalnego umożliwiają przeprowadzenie doskonałych badań eksploracyjnych, które dobrze uzupełniają inne modele produktu

IBM SPSS Statistics. Zawężając zakres badania danych, wizualizacja danych poprzez skalowanie optymalnie może stworzyć podstawę do analizy, która skupia się na interpretacji parametrów modelu.

Poziom skalowania optymalnego i poziom pomiaru

Przy pierwszym użyciu procedur kategorii może to zagadnienie może być bardzo kłopotliwe. Podczas określania poziomu nie jest określany poziom *pomiaru* zmiennych, ale poziom, na którym są one *skalowane*. Koncepcja jest taka, że zmienne do skwantyfikowania mogą mieć nieliniowe relacje niezależnie od sposobu ich pomiaru.

Dla procedury kategorii dostępne są trzy podstawowe poziomy pomiaru:

- Poziom **nominalny** oznacza, że wartości zmiennych reprezentują nieuporządkowane kategorie. Przykładami zmiennych, które mogą być zmiennymi nominalnymi są region, obszar kodu pocztowego, przynależność wyznaniowa i kategorie wielokrotnego wyboru.
- Poziom **porządkowy** oznacza, że wartości zmiennych reprezentują kategorie uporządkowane. Przykładami są oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.
- Poziom **liczbowy** oznacza, że wartości zmiennych reprezentują uporządkowane kategorie ze znaczącą metryką, która umożliwi porównywanie odległości między kategoriami. Przykładami są: wiek w latach oraz przychód w tysiącach dolarów.

Założmy na przykład, że zmienne *region*, *zadanie* i *wiek* są zakodowane w sposób przedstawiony w poniższej tabeli.

Kod regionu	Wartość regionu	Kod zadania	Wartość zadania	Wiek
1	Północ	1	wewn.	20
2	Południe	2	przedstawiciel handl.	22
3	Wschód	3	menedżer	25
4	Zachód			27

Przedstawione wartości reprezentują kategorie poszczególnych zmiennych. *Region* będzie zmienną nominalną. Dla zmiennej *region* istnieją cztery kategorie bez wewnętrznego uporządkowania. Wartości od 1 do 4 reprezentują te cztery kategorie; schemat kodowania jest całkowicie dowolny. Z kolei dla zmiennej *zadanie* można założyć, że będzie to zmienna porządkowa. Kategorie porządkowe tworzą ciąg pomiędzy wartościami wewn. i menedżer. Wyższe kody reprezentują zadanie dla stanowisk znajdujących się wyżej na drabinie kariery. Znane są jednak tylko informacje o kolejności — nie można nic powiedzieć o odległości pomiędzy sąsiadującymi kategoriami. Z kolei dla zmiennej *wiek* można założyć, że jest to zmienna liczbowa. W przypadku *wieku* odległości pomiędzy wartościami są wewnątrznie istotne. Odległość pomiędzy wartościami od 20 do 22 jest taka sama, jak odległość pomiędzy wartościami 25 i 27, przy czym odległość pomiędzy wartościami 22 i 25 jest większa od pozostałych.

Wybór poziomu skalowania optymalnego

Ważne jest, aby zrozumieć, że nie są dostępne żadne wewnętrzne właściwości zmiennych, które automatycznie predefiniują, jaki poziom skalowania optymalnego należy dla nich określić. Dane można eksplorować w dowolny sposób, który ma sens i ułatwia interpretację. Przykładowo, użycie podczas analizy zmiennej z poziomu numerycznego na poziomie porządkowym przekształcenia nieliniowego umożliwi rozwiązanie z mniejszą liczbą wymiarów.

Poniżej przedstawiono dwa przykłady, które pokazują, że "oczywisty" poziom pomiaru może nie być najbardziej optymalnym poziomem skalowania optymalnego. Założmy, że zmienna sortuje obiekty, dzieląc je na grupy wiekowe. Mimo że wiek może być skalowany jako zmienna liczbowa, może być prawdą, że bezpieczeństwo osób mających mniej niż 25 ma pozytywną relację z wiekiem, natomiast bezpieczeństwo osób w wieku powyżej 60 lat ma negatywną relację z wiekiem. W takim przypadku bardziej zasadne jest traktowanie wieku jako zmiennej nominalnej.

I kolejny przykład: wydaje się, że zmienna, która sortuje osoby według poglądów politycznych, powinna być zasadniczo nominalna. Jeśli jednak uszeregujemy partie od lewej do prawej strony politycznej, kwantyfikacja partii powinna respektować ten porządek, co umożliwi zastosowanie porządkowego poziomu analizy.

Nawet jeśli nie istnieją wstępnie zdefiniowane właściwości zmiennych, które przyporządkują je wyłącznie do jednego poziomu, istnieje kilka ogólnych wytycznych, które ułatwią pracę początkującemu użytkownikowi. W przypadku kwantyfikacji nominalnej pojedynczej zwykle nie wiadomo, jaka ma być kolejność kategorii, ale zgodnie z oczekiwaniami podczas analizy powinna zostać jakaś narzucona. Jeśli kolejność kategorii jest znana, należy spróbować zastosować kwantyfikację porządkową. Jeśli kolejność kategorii nie jest możliwa do określenia, można użyć kwantyfikacji wielokrotnej nominalnej.

Wykresy przekształceń

Ze względu na różnice w poziomach, na jakich mogą być skalowane poszczególne zmienne, występują różne ograniczenia w kwantyfikacji. Wykresy przekształceń obrazują relację między kwantyfikacjami a pierwotnymi kategoriami będącymi wynikiem wybranego optymalnego poziomu skalowania. Na przykład wykres przekształcenia liniowego uzyskuje się, gdy zmienna jest traktowana jako liczbowa. Zmienne traktowane jako porządkowe powodują wygenerowanie niemalejącego wykresu przekształceń. Wykresy przekształceń zmiennych traktowanych jako nominalne, o kształcie U (lub jego odwrotności), obrazują relację kwadratową. Na podstawie zmiennych nominalnych mogą także powstawać wykresy przekształceń bez wyraźnych trendów, jeśli kolejność kategorii zostanie całkowicie zmieniona. Na poniższej ilustracji przedstawiono przykładowy wykres przekształceń.

Wykresy przekształceń szczególnie dobrze nadają się do oceny wyboru optymalnego poziomu skalowania. Jeśli dla kilku kategorii kwantyfikacje będą podobne, uzasadnione może być połączenie ich w jedną kategorię. Z kolei jeśli kwantyfikacja zmiennej traktowanej jako nominalna będzie wykazywać trend rosnący, przekształcenie porządkowe może przynieść podobne dopasowanie. Jeśli ten trend jest liniowy, odpowiednie może być traktowanie zmiennej jako liczbowej. Jeśli jednak uzasadnione jest połączenie kategorii lub zmiana poziomów skalowania, analiza nie ulegnie istotnej zmianie.

Kody kategorii

Należy zachować ostrożność podczas kodowania zmiennych jakościowych, ponieważ niektóre schematy kodowania mogą spowodować otrzymanie nieprawidłowego wyniku lub niekompletnych analiz. W poniższej tabeli przedstawiono dostępne schematy kodowania *zadania*.

Tabela 2. Alternatywne schematy kodowania zadania

Kategoria	A	B	C	D
wewn.	1	1	5	1
przedstawiciel handl.	2	2	6	5
menedżer	3	7	7	3

Niektóre procedury kategorii wymagają zdefiniowania zakresu wszystkich używanych zmiennych. Każda wartość spoza tego zakresu jest traktowana jako brak wartości. Minimalną wartością kategorii jest zawsze 1. Maksymalna wartość kategorii jest podawana przez użytkownika. Ta wartość nie określa *liczby* kategorii dla zmiennej — jest to *największa* wartość dla kategorii. Przykładowo, w tabeli dla schematu A maksymalna wartość kategorii wynosi 3, a dla schematu B wynosi 7, chociaż oba kody schematów dotyczą tych samych trzech kategorii.

Zakres zmiennych określa, które kategorie zostaną pominięte podczas analizy. Wszystkie kategorie z kodami spoza zdefiniowanego zakresu zostają wykluczone z analizy. Jest to prosta metoda pomijania kategorii, ale może spowodować wykonanie niepożądanego wyniku. Nieprawidłowo zdefiniowana wartość maksymalna kategorii może spowodować wykluczenie *prawidłowych* kategorii z analizy. Przykładowo, zdefiniowanie dla schematu B maksymalnej wartości kategorii na 3 oznacza, że *zadanie* ma kategorie zakodowane w zakresie od 1 do 3; kategoria *menedżera* nie jest brana pod uwagę. Ponieważ żadna kategoria nie została zakodowana jako 3, trzecia kategoria w analizie nie będzie zawierała

żadnych obserwacji. Ta analiza będzie odpowiednia, jeśli konieczne jest pominięcie wszystkich kategorii menedżera. Jeśli jednak menedżerowie mają być uwzględniani, maksymalna wartość dla kategorii musi być zdefiniowana jako 7, a wartości, które nie będą uwzględniane, muszą być zakodowane jako wartości powyżej 7 lub poniżej 1.

W przypadku zmiennych traktowanych jako nominalne lub porządkowe zakres kategorii nie wpływa na wyniki. W przypadku zmiennych nominalnych znaczenie ma wyłącznie etykieta, a wartość powiązana z tą etykietą jest nieistotna. W przypadku zmiennych porządkowych kolejność kategorii jest zachowywana w kwantyfikacjach; wartości kategorii są nieistotne. Wszystkie schematy kodowania wyświetlane dla samej kolejności kategorii będą zapewniały takie same wyniki. Przykładowo, pierwsze trzy schematy w tabeli są funkcjonalnie równoważne, jeśli *zadanie* zostaje poddane analizie na poziomie porządkowym. Kolejność kategorii jest w tych schematach identyczna. Z kolei schemat D odwraca drugą i trzecią kategorię i zwraca wyniki inne niż w pozostałych schematach.

Chociaż wiele schematów kodowania dla zmiennych jest funkcjonalnie równoważnych, preferowane są schematy z niewielkimi różnicami pomiędzy kodami, ponieważ kody mają wpływ na wynik przetwarzany w trakcie wykonywania procedury. Wszystkie kategorie zakodowane z użyciem wartości od 1 do maksimum zdefiniowanego przez użytkownika są prawidłowe. Jeśli któraś z tych kategorii jest pusta, odpowiednie kwantyfikacje zostaną potraktowane jako systemowe braki danych lub jako wartość 0, w zależności od procedury. Chociaż żadne z tych przypisań nie wpływa na analizę, wynik dla tych kategorii zostanie wygenerowany. Dlatego w przypadku schematu B *zadanie* obejmuje cztery kategorie, których wartości są traktowane jako systemowe braki danych. W przypadku schematu C również cztery kategorie otrzymają wskaźniki systemowych braków danych. Natomiast w przypadku schematu A żadne kwantyfikacje nie są traktowane jako systemowe braki danych. Użycie kolejnych liczb całkowitych jako kodów dla zmiennych traktowanych jako nominalne lub porządkowe spowoduje, że znacznie mniej wyników będzie wpływało na wyniki.

Schematy kodowania dla zmiennych traktowanych jako liczbowe są bardziej ograniczone niż te dla zmiennych traktowanych jako porządkowe. W przypadku tych zmiennych różnice pomiędzy kolejnymi kategoriami nie są istotne. w poniższej tabeli przedstawiono schematy kodowania dla *wieku*.

Kategoria	A	B	C
20	20	1	1
22	22	3	2
25	25	6	3
27	27	8	4

Podczas rekodowania zmiennych liczbowych konieczne jest zachowanie różnic pomiędzy kategoriami. Użycie oryginalnych wartości jest jedyną metodą zapewniającą zachowanie różnic. Może to jednak spowodować, że wiele kategorii będzie miało wskaźniki systemowych braków danych. Przykładowo, schemat A obejmuje oryginalnie obserwowane wartości. W przypadku wszystkich procedur dotyczących kategorii z wyjątkiem analizy korespondencji, maksymalna wartość kategorii wynosi 27, a minimalna wartość kategorii jest ustawiona na 1. Pierwsze 19 kategorii jest puste i otrzymuje wskaźniki systemowego braku. Wynik może dość szybko stać się nieefektywny, jeśli maksymalna wartość dla kategorii jest dużo większa od 1 i pomiędzy wartością 1 i wartością maksymalną istnieje wiele pustych kategorii.

Aby zmniejszyć liczbę wyników, można wykonać operację rekodowania. Jednak w przypadku wartości liczbowych nie należy używać funkcji automatycznego rekodowania. Kodowanie z użyciem kolejnych liczb całkowitych spowoduje różnicę wynoszącą 1 pomiędzy wszystkimi kolejnymi kategoriami, a w wyniku tego wszystkie kwantyfikacje zostaną równomiernie rozmieszczone. Po przeprowadzeniu rekodowania na kolejne liczby całkowite charakterystyki metryk istotne w przypadku traktowania zmiennych jako wartości liczbowe zostają zniszczone. Przykładowo, schemat C w tabeli dotyczy automatycznego rekodowania dla *wieku*. Różnica pomiędzy kategoriami 22 i 25 zmieniła się z trzech na jeden, a kwantyfikacje będą odzwierciedlać dalsze różnice.

Alternatywnym schematem rekodowania, który zachowuje różnice pomiędzy kategoriami, jest odjęcie najmniejszej wartości kategorii od każdej kategorii i dodanie 1 do każdej różnicy. W wyniku tego przekształcenia tworzony jest schemat B. Najmniejsza wartość dla kategorii, 20, została odjęta od każdej kategorii, a wartość 1 została dodana do każdego wyniku. Przekształcone kody mają minimalną wartość 1, a wszystkie różnice są identyczne z oryginalnymi danymi. Maksymalna wartość dla kategorii wynosi teraz 8, a wszystkie zerowe kwantyfikacje znajdujące się przed pierwszą kwantyfikacją niezerową zostaną wyeliminowane. Jednak kwantyfikacje niezerowe odpowiadające poszczególnym kategoriom wynikającym ze schematu B są takie same jak kwantyfikacje ze schematu A.

Która procedura jest najlepsza dla danej aplikacji?

Techniki uwzględnione w czterech procedurach (analiza korespondencji, wielowymiarowa analiza korespondencji, analiza głównych składowych dla danych jakościowych i nieliniowa analiza korelacji kanonicznych) należą do ogólnego obszaru analizy danych wielowymiarowych określanego jako **redukcja wymiarów**. Oznacza to, że relacje pomiędzy zmiennymi są reprezentowane w kilku wymiarach — na przykład w dwóch lub trzech, o ile to możliwe. Pozwala to na opisanie struktur lub wzorców w relacjach, których oszacowanie byłoby zbyt przy ich oryginalnej wielkości i złożoności. W aplikacjach do badań rynku techniki te mogą mieć formę **mapowania percepcyjnego**. Największą zaletą tych procedur jest możliwość uwzględnienia danych o różnych poziomach skalowania optymalnego.

Regresja dla danych jakościowych opisuje relacje pomiędzy zmienną odpowiedzi jakościowej a kombinacją zmiennych predyktorów jakościowych. Wpływ poszczególnych zmiennych predyktorów na zmienną odpowiedzi jest opisywany przez odpowiednią wagę regresji. Podobnie jak w innych procedurach, dane można analizować na różnych poziomach skalowania optymalnego.

Skalowanie wielowymiarowe i wielowymiarowe skalowanie preferencji opisują relacje pomiędzy obiektami w przestrzeni niskowymiarowej z użyciem odległości pomiędzy obiektami.

Poniżej przedstawiono krótkie wytyczne dotyczące poszczególnych procedur:

- Regresja dla danych jakościowych umożliwia prognozowanie wartości jakościowych zmiennych zależnych na podstawie kombinacji jakościowych zmiennych niezależnych.
- Analiza głównych składowych dla danych jakościowych umożliwia uwzględnienie wzorców zmiennych w jednym zestawie zmiennych na różnych poziomach skalowania optymalnego.
- Nieliniowa analiza korelacji kanonicznych pozwala ocenić zakres korelacji dla co najmniej dwóch zestawów zmiennych na różnych poziomach skalowania optymalnego.
- Analizy korespondencji można użyć do analizy dwustronnych tabel kontyngencji lub danych, które mogą być wyrażone w postaci tabeli dwustronnej, takich jak preferencje dotyczące marki lub dane socjometryczne.
- Wielowymiarowa analiza korespondencji umożliwia analizowanie macierzy jakościowych danych wielowymiarowych, jeśli założenie nie ma być silniejsze niż takie, że wszystkie zmienne są analizowane na poziomie nominalnym.
- Skalowanie wielowymiarowe pozwala na analizowanie danych odległości w celu znalezienia reprezentacji najmniejszych kwadratów dla pojedynczego zestawu obiektów w przestrzeni niskowymiarowej.
- Wielowymiarowe skalowanie preferencji umożliwia analizowanie danych odległości w celu znalezienia reprezentacji najmniejszych kwadratów dla dwóch zestawów obiektów w przestrzeni niskowymiarowej.

Regresja dla danych jakościowych

Regresja dla danych jakościowych jest najbardziej odpowiednią procedurą, jeśli celem analizy jest predykcja zmiennej zależnej (odpowieź) z zestawu zmiennych niezależnych (predyktor). Podobnie jak we wszystkich procedurach skalowania optymalnego, wartości skali są przypisywane do poszczególnych kategorii każdej zmiennej w taki sposób, że te wartości są optymalne w odniesieniu do regresji. Rozwiązanie wynikające z przeprowadzenia regresji dla danych jakościowych maksymalizuje korelację kwadratową pomiędzy przekształconą odpowiedzią a ważoną kombinacją przekształconych predyktorów.

Relacja z innymi procedurami operującymi na kategoriach. Regresję dla danych jakościowych ze skalowaniem optymalnym można porównać do analizy korelacji kanonicznych ze skalowaniem optymalnym obejmującej dwa zestawy, z których każdy zawiera tylko zmienną zależną. W przypadku tej drugiej techniki podobieństwo zestawów uzyskuje się przez porównanie każdego zestawu z nieznaną zmienną, która znajduje się w obszarze pomiędzy wszystkimi zestawami. W przypadku regresji dla danych jakościowych podobieństwo przekształconej odpowiedzi i liniowej kombinacji przekształconych odpowiedzi jest oceniane bezpośrednio.

Relacja z technikami standardowymi. W standardowej regresji liniowej zmienne jakościowe mogą być rekodowane jako zmienne wskaźnikowe lub mogą być traktowane w taki sam sposób, jak zmienne z poziomu przedziału ufności. W pierwszym przypadku model zawiera osobną stałą i nachylenie dla każdej kombinacji poziomów zmiennych jakościowych. W wyniku tego wiele parametrów wymaga interpretacji. W drugim przypadku dla każdej zmiennej szacowany jest tylko jeden parametr. Jednak dowolność kodowania kategorii sprawia, że uogólnienie jest niemożliwe.

Jeśli niektóre zmienne nie są ciągłe, można skorzystać z alternatywnej analizy. Jeśli odpowiedź jest ciągła, a predyktory są jakościowe, często stosowana jest analiza wariancji. Jeśli odpowiedź jest jakościowa, a predyktory są ciągłe, odpowiednia może być analiza regresji logistycznej lub analiza dyskryminacyjna. Modele logliniowe są często używane, jeśli zarówno odpowiedź jak i predyktory są jakościowe.

Regresja ze skalowaniem optymalnym oferuje trzy poziomy skalowania dla każdej zmiennej. Kombinacje tych poziomów pozwalają uwzględnić szeroki zakres relacji nieliniowych, dla których pojedyncze "standardowe" metody są nieodpowiednie. W rezultacie skalowanie optymalne oferuje większą elastyczność niż standardowe rozwiązania przy niewiele większym poziomie złożoności.

Ponadto nieliniowe przekształcenia predyktorów zwykle redukują zależności pomiędzy predyktorami. W przypadku porównania wartości własnych macierzy korelacji dla predyktorów z wartościami własnymi macierzy korelacji dla predyktorów poddanych skalowaniu optymalnym ten drugi zestaw zwykle będzie mniej zmienny od pierwszego. Innymi słowy, w przypadku regresji dla danych jakościowych skalowanie optymalne powoduje zmniejszenie większych wartości własnych macierzy korelacji predyktora i zwiększenie mniejszych wartości własnych.

Analiza głównych składowych dla danych jakościowych

Analiza głównych składowych dla danych jakościowych jest najbardziej odpowiednia, jeśli konieczne jest uwzględnienie wzorców zmiennych w jednym zestawie zmiennych na różnych poziomach skalowania optymalnego. Ta technika umożliwia przeprowadzenie próby zredukowania wymiarowości zestawu zmiennych z jednoczesnym uwzględnieniem jak największej ilości zmiennych. Wartości skali są przypisywane do poszczególnych kategorii każdej zmiennej w taki sposób, że wartości te są optymalne w odniesieniu do rozwiązania głównych składowych. Obiekty uwzględniane w analizie uzyskują wartości składowych na podstawie danych kwantyfikowanych. Wykresy wartości składowych przedstawiają wzorce dla obiektów poddawanych analizie i mogą ujawnić nietypowe obiekty w danych. Rozwiązanie wynikające z analizy głównych składowych dla danych jakościowych maksymalizuje korelacje wartości obiektów z każdą zmienną skwantyfikowaną dla określonej liczby składowych (wymiarów).

Istotnym zadaniem głównych składowych dla danych jakościowych jest sprawdzenie preferowanych danych, podczas którego respondenci określają rangę lub ustalają liczbę pozycji z uwzględnieniem preferencji. W przypadku typowej konfiguracji danych programu IBM SPSS Statistics wiersze odpowiadają pojedynczym pomiarom, kolumny odpowiadają pomiarom dla pozycji, a wartości pomiędzy wierszami odpowiadają preferowanym ocenom (na przykład w skali od 0 do 10), tworząc dane warunkowe wiersza. W przypadku preferowanych danych można potraktować pojedyncze pomiary jako zmienne. Korzystając z procedury transpozycji można dane transponować. Oceniający stają się zmiennymi, a wszystkie zmienne są deklarowane jako porządkowe. Dopuszczalne jest użycie większej liczby zmiennych niż obiektów w CATPCA.

Relacja z innymi procedurami operującymi na kategoriach. Jeśli wszystkie zmienne są zadeklarowane jako nominalne wielokrotne, analiza głównych składowych dla danych jakościowych (CATPCA) spowoduje przeprowadzenie analizy równoważnej do wielowymiarowej analizy korespondencji dla tych samych zmiennych. Dlatego analiza głównych składowych dla danych jakościowych może być postrzegana jako typ wielokrotnej analizy korespondencji, w której niektóre zmienne są zadeklarowane jako porządkowe lub liczbowe.

Relacja z technikami standardowymi. Jeśli wszystkie zmienne są skalowane na poziomie liczbowym, analiza głównych składowych dla danych jakościowych jest równoważna z analizą głównych składowych dla danych standardowych.

W ujęciu bardziej ogólnym, analiza głównych składowych dla danych jakościowych (CATPCA) jest alternatywnym rozwiązaniem do obliczania korelacji pomiędzy skalami nieliczbowymi i analizowania ich z użyciem głównych składowych dla danych standardowych lub analizy czynnikowej. Nieprzemysłane użycie zwykłego współczynnika korelacji Pearsona jako miary powiązania dla danych porządkowych może spowodować ogromne obciążenie podczas estymacji korelacji.

Nieliniowa analiza korelacji kanonicznych

Nieliniowa analiza korelacji kanonicznych jest bardzo ogólną procedurą o wielu różnych zastosowaniach. Celem nieliniowej analizy korelacji kanonicznych jest analizowanie relacji pomiędzy co najmniej dwoma zestawami zmiennych zamiast pomiędzy samymi zmiennymi, jak w przypadku analizy głównych składowych. Przykładowo, dostępne są dwa zestawy zmiennych: jeden zestaw zmiennych może obejmować elementy danych demograficznych dla zestawu respondentów, a drugi odpowiedzi dla zbioru elementów dotyczących położenia. Poziomy skalowania w tej analizie mogą być dowolną kombinacją poziomów nominalnych, porządkowych i liczbowych. Analiza korelacji kanonicznych ze skalowaniem optymalnym określa podobieństwo pomiędzy zestawami poprzez jednoczesne porównanie zmiennych kanonicznych z poszczególnymi zestawów ocen przypisanych do obiektów.

Relacja z innymi procedurami operującymi na kategoriach. Jeśli dostępne są co najmniej dwa zestawy zmiennych zawierające tylko jedną zmienną w zestawie, analiza korelacji kanonicznych ze skalowaniem optymalnym jest równoważna z analizą głównych składowych ze skalowaniem optymalnym. Jeśli wszystkie zmienne w objętej analizą dla jednej zmiennej w zestawie są wielokrotnie nominalne, analiza korelacji kanonicznych ze skalowaniem optymalnym jest równoważna wielowymiarowej analizie korespondencji. Jeśli dostępne są dwa zestawy zmiennych, z których jeden zawiera tylko jedną zmienną, analiza korelacji kanonicznych ze skalowaniem optymalnym jest równoważna regresji dla danych jakościowych ze skalowaniem optymalnym.

Relacja z technikami standardowymi. Standardowa analiza korelacji kanonicznych to technika statystyczna, która wyszukuje w jednym i w drugim zestawie zmiennych maksymalnie skorelowane kombinacje liniowe. W tym zestawie kombinacji liniowych analiza korelacji kanonicznych umożliwia wyszukanie kolejnych niezależnych zestawów kombinacji liniowych, zwanych zmiennymi kanonicznymi, aż do uzyskania maksymalnej wartości równej licznie zmiennych w najmniejszym zestawie.

Jeśli analizą objęte są dwa zestawy zmiennych i wszystkie zmienne są zdefiniowane jako liczbowe, analiza korelacji kanonicznych ze skalowaniem optymalnym jest równoważna standardowej analizie korelacji kanonicznych. Chociaż program IBM SPSS Statistics nie oferuje procedury analizy korelacji kanonicznych, wiele istotnych statystyk można uzyskać na podstawie wielowymiarowej analizy wariancji.

Analiza korelacji kanonicznych ze skalowaniem optymalnym ma różne inne zastosowania. Jeśli dostępne są dwa zestawy zmiennych i jeden z tych zestawów zawiera zmienną nominalną zadeklarowaną jako nominalna pojedyncza, wyniki analizy korelacji kanonicznych ze skalowaniem optymalnym mogą być interpretowane w podobny sposób, jak w przypadku analizy regresji. Jeśli z założenia zmienna ma być zmienną nominalną wielokrotną, analiza ze skalowaniem optymalnym jest alternatywnym rozwiązaniem dla analizy dyskryminacyjnej. Pogrupowanie zmiennych w więcej niż dwóch zestawach zapewnia różne możliwości analizowania danych.

Analiza korespondencji

Celem analizy korespondencji jest utworzenie wykresów łącznych (biplot) dla tabel korespondencji. W tabeli korespondencji zakłada się, że zmienne wiersza i kolumny reprezentują nieuporządkowane kategorie; dlatego, zawsze używany jest nominalny poziom optymalnego skalowania. W przypadku obu zmiennych sprawdzane są tylko informacje dotyczące nominalności. Oznacza to, że uwzględniany jest tylko fakt, że niektóre obiekty należą do tej samej kategorii, a inne nie. Nie są dokonywane żadne założenia dotyczące odległości lub kolejności pomiędzy kategoriami tej samej zmiennej.

Jednym z zastosowań analizy korespondencji jest analiza dwustronnych tabel kontyngencji. Jeśli tabela zawiera aktywne wiersze r i aktywne kolumny c , liczba wymiarów w rozwiązaniu analizy korespondencji

to minimalna wartość r minus 1 lub c minus 1, w zależności od tego, która wartość jest mniejsza. Innymi słowy, możliwa jest dokładna reprezentacja kategorii wiersza lub kolumny w tabeli kontyngencji w przestrzeni wymiarów. Praktycznie jednak bardziej odpowiednia byłaby reprezentacja kategorii wiersza i kolumny tabeli dwustronnej w przestrzeni niskowymiarowej, na przykład obejmującej dwa wymiary, ponieważ wykresy dwuwymiarowe są bardziej zrozumiałe niż wielowymiarowe reprezentacje przestrzenne.

Jeśli używanych jest mniej wymiarów niż ich maksymalna możliwa liczba, statystyki utworzone podczas analizy opisują, jak dobrze kategorie wiersza i kolumny są reprezentowane w reprezentacji niskowymiarowej. Przy założeniu, że jakość reprezentacji rozwiązania dwuwymiarowego jest dobra, można zbadać wykresy punktów wierszowych i punktów kolumnowych w celu sprawdzenia, które kategorie zmiennej wiersza są podobne, które kategorie zmiennej kolumny są podobne i które kategorie wiersza i kolumny są do siebie podobne.

Relacja z innymi procedurami operującymi na kategoriach. Prosta analiza korespondencji jest ograniczona do tabel dwustronnych. Jeśli konieczne jest zbadanie więcej niż dwóch zmiennych, można zmienne połączyć, tworząc zmienne interakcji. Przykładowo, dla zmiennych *region*, *zadanie* i *wiek* można połączyć zmienne *region* i *zadanie*, tworząc nową zmienną *regzad* z 12 kategoriami przedstawionymi w poniższej tabeli. Ta nowa zmienna tworzy tabelę dwustronną ze zmienną *wiek* (12 wierszy, 4 kolumny), którą można poddać analizie z użyciem analizy korespondencji.

Tabela 4. Kombinacje zmiennych region i zadanie

Kod kategorii	Definicja kategorii	Kod kategorii	Definicja kategorii
1	Północ, wewn.	7	Wschód, wewn.
2	Północ, przedstawiciel handl.	8	Wschód, przedstawiciel handl.
3	Północ, menedżer	9	Wschód, menedżer
4	Południe, wewn.	10	Zachód, wewn.
5	Południe, przedstawiciel handl.	11	Zachód, przedstawiciel handl.
6	Południe, menedżer	12	Zachód, menedżer

Jedynym mankamentem tego rozwiązania jest to, że każda para zmiennych może być połączona. Można połączyć *zadanie* i *wiek*, uzyskując kolejną zmienną jakościową z 12 kategoriami. Lub można połączyć *region* i *wiek*, co spowoduje powstanie nowej zmiennej jakościowej z 16 kategoriami. Każda z tych zmiennych interakcji tworzy tabelę dwustronną dla pozostałych zmiennych. Analiza korespondencji tych trzech tabel nie zapewni identycznych wyników, chociaż każde rozwiązanie jest poprawne. Ponadto, jeśli istnieją cztery zmienne lub więcej zmiennych, mogą zostać utworzone tabele dwustronne porównujące zmienną interakcji z inną zmienną iteracji. Liczba możliwych do przeanalizowania tabel może stać się bardzo duża, nawet w przypadku kilku zmiennych. Można wybrać jedną z tych tabel do przeanalizowania lub przeprowadzić analizę wszystkich. Alternatywnie procedura wielowymiarowej analizy korespondencji może być używana do zbadania wszystkich zmiennych jednocześnie, bez konieczności tworzenia zmiennych interakcji.

Relacja z technikami standardowymi. Procedura tabel krzyżowych może być również używana do analizy tabel kontyngencji, z niezależnością stanowiącą główny cel analizy. Jednak nawet w przypadku małych tabel wykrycie przyczyny odchylenia od niezależności może być trudne. Użyteczność analizy korespondencji polega na wyświetlaniu wzorców dla tabel dwustronnych dowolnej wielkości. Jeśli istnieje powiązanie pomiędzy zmiennymi wiersza i kolumny — czyli, jeśli wartość chi-kwadrat jest istotna — analiza korespondencji może pomóc w określeniu charakteru relacji.

Wielowymiarowa analiza korespondencji

Wielowymiarowa analiza korespondencji próbuje utworzyć rozwiązanie, w którym obiekty z tej samej kategorii są rysowane na wykresie blisko siebie, a obiekty z różnych kategorii są wykreślane daleko od

siebie. Każdy obiekt znajduje się najbliżej jak to możliwe względem punktów kategorii dla kategorii, która dotyczy obiektu. W ten sposób kategorie dzielą obiekty na jednorodne podgrupy. Zmienne są uznawane za jednorodne, jeśli klasyfikują obiekty w tych samych kategoriach do tych samych podgrup.

W przypadku rozwiązania jednowymiarowego wielowymiarowa analiza korespondencji przypisuje optymalne wartości skali (kwantyfikacje kategorii) do każdej kategorii poszczególnych zmiennych w taki sposób, że ogólnie, średnio, kategorie mają maksymalny rozrzut. W przypadku rozwiązania dwuwymiarowego wielowymiarowa analiza korespondencji wyszukuje drugi zbiór kwantyfikacji kategorii poszczególnych zmiennych niepowiązanych z pierwszym zbiorem, próbując ponownie zmaksymalizować rozrzut itd. Ponieważ kategorie zmiennych uzyskują tyle ocen, ile jest wymiarów, w analizie zakłada się, że zmienne są zmiennymi wielokrotnymi nominalnymi na poziomie skalowania optymalnego.

Wielowymiarowa analiza korespondencji przypisuje również oceny do obiektów poddawanych analizie w taki sposób, że kwantyfikacje kategorii są średnimi lub środkami ciężkości ocen obiektów dla obiektów znajdujących się w danej kategorii.

Relacja z innymi procedurami operującymi na kategoriach. Wielowymiarowa analiza korespondencji jest również znana jako analiza jednorodności lub podwójne skalowanie. Zapewnia możliwe do porównania, ale nie identyczne, wyniki na potrzeby analizy korespondencji, jeśli istnieją tylko dwie zmienne. Analiza korespondencji generuje unikalny wynik podsumowujący dopasowanie i jakość reprezentacji rozwiązania, z uwzględnieniem informacji o stabilności. Dlatego analiza korespondencji jest zalecana zwykle do wielowymiarowej analizy korespondencji w przypadku dwóch zmiennych. Kolejna różnica pomiędzy dwoma procedurami polega na tym, że dane wejściowe dla wielowymiarowej analizy korespondencji są w postaci macierzy danych, w której wiersze są obiektami, a kolumny zmiennymi, natomiast dane wejściowe dla analizy korespondencji mogą być taką samą macierzą danych, ogólną macierzą odległości lub tabelą kontyngencji połączenia, która jest zagregowaną macierzą, w której wiersze i kolumny reprezentują kategorie zmiennych. Wielowymiarową analizę korespondencji można również traktować jako analizę głównych składowych danych skalowanych na poziomie wielokrotnie nominalnym.

Relacja z technikami standardowymi. Wielowymiarowa analiza korespondencji może być również uważana za analizę wielozmienną tabeli kontyngencji. Wielozmienną tabelę kontyngencji można również analizować z zastosowaniem procedury tabel krzyżowych, jednak ta procedura zapewnia osobne statystyki podsumowania dla każdej kategorii poszczególnych zmiennych kontrolnych. Wielowymiarowa analiza korespondencji często umożliwia podsumowanie relacji pomiędzy zmiennymi za pomocą jednego wykresu dwuwymiarowego. Zaawansowana wielowymiarowa analiza korespondencji umożliwia zastępowanie oryginalnych wartości kategorii wartościami optymalnego skalowania z pierwszego wymiaru i wykonanie dodatkowej analizy wielu zmiennych. Ponieważ wielowymiarowa analiza korespondencji zastępuje etykiety kategorii liczbowymi wartościami skali, po wykonaniu tej analizy możliwe jest zastosowanie różnych procedur, które wymagają danych liczbowych. Przykładowo, procedura analizy czynnikowej tworzy pierwszy główny element składowy, który jest odpowiednikiem pierwszego wymiaru w wielowymiarowej analizie korespondencji. Oceny składników w pierwszym wymiarze są takie same, jak oceny obiektów, a kwadraty ładunków składowych są takie same, jak miary dyskryminacji. Drugi wymiar wielowymiarowej analizy korespondencji nie jest jednak taki sam, jak drugi wymiar analizy czynnikowej.

Skalowanie wielowymiarowe

Skalowanie wielowymiarowe jest najbardziej odpowiednią procedurą w sytuacji, gdy celem analizy jest znalezienie struktury w zbiorze miar odległości w pojedynczym zbiorze obiektów lub obserwacji. Realizuje się to poprzez przypisywanie obserwacji do poszczególnych miejsc w niskowymiarowej przestrzeni pojęciowej w taki sposób, że odległości między punktami w przestrzeni możliwie blisko odpowiadają danym miarom (nie)podobieństwa. Wynikiem jest reprezentacja najmniejszych kwadratów obiektów w przestrzeni niskowymiarowej, co (w wielu przypadkach) pomaga lepiej zrozumieć dane.

Relacja z innymi procedurami operującymi na kategoriach. Gdy dostępne są dane z wieloma zmiennymi, na podstawie których można utworzyć odległości i które następnie analizuje się metodą skalowania wielowymiarowego, uzyskane wyniki są podobne do wyników analizy danych przy użyciu metody głównych składowych jakościowych CATPCA z główną normalizacją obiektu. Ten rodzaj analizy PCA nazywany jest także analizą składowych głównych.

Relacja z technikami standardowymi. Procedura Wielowymiarowe skalowanie preferencji (PROXSCAL) oferuje kilka udoskonaleń w stosunku do procedury skalowania dostępnej w opcji Statistics Base (ALSCAL). PROXSCAL w przypadku niektórych modeli korzysta z szybszego algorytmu i umożliwia nałożenie ograniczeń na przestrzeń wspólną. Ponadto PROXSCAL próbuje zminimalizować surową wartość STRESS, a nie S-STRESS (tzw. ang. **strain**). Znormalizowana surowa wartość STRESS jest z reguły preferowana, ponieważ jest miarą bazującą na odległościach, podczas gdy S-STRESS bazuje na kwadratach odległości.

Wielowymiarowe skalowanie preferencji (PREFSCAL)

Wielowymiarowe skalowanie preferencji jest najbardziej odpowiednią procedurą w sytuacji, gdy celem analizy jest znalezienie struktury w zbiorze miar odległości między dwoma zbiorami obiektów (tzw. obiektami w rzędach i obiektami w kolumnach). Realizuje się to poprzez przypisywanie obserwacji do poszczególnych miejsc w niskowymiarowej przestrzeni pojęciowej w taki sposób, że odległości między punktami w przestrzeni możliwie blisko odpowiadają danym miarom (nie)podobieństwa. Wynikiem jest reprezentacja najmniejszych kwadratów obiektów wierszy i kolumn w przestrzeni niskowymiarowej, co (w wielu przypadkach) pomaga lepiej zrozumieć dane.

Relacja z innymi procedurami operującymi na kategoriach. Jeśli dane zawierają odległości między obiektami z jednego zbioru (kwadratowa macierz symetryczna), należy użyć Skalowania wielowymiarowego.

Relacja z technikami standardowymi. Procedura Wielowymiarowe skalowanie preferencji (PREFSCAL) oferuje kilka udoskonaleń w stosunku do procedury skalowania preferencji dostępnej w opcji Statistics Base (ALSCAL). PREFSCAL pozwala nałożyć ograniczenia na przestrzeń wspólną; ponadto PREFSCAL próbuje zminimalizować ukaraną wartość STRESS, co pomaga uniknąć rozwiązań zdegenerowanych (na co podatne są starsze algorytmy).

Współczynnik proporcji w wykresach skalowania optymalnego

Współczynnik proporcji w wykresach skalowania optymalnego jest izotropowy. W przypadku wykresu dwuwymiarowego odległość reprezentująca jedną jednostkę w wymiarze 1 jest równa odległości reprezentującej jedną jednostkę w wymiarze 2. W przypadku zmiany zakresu wymiaru na wykresie dwuwymiarowym system zmienia wielkość drugiego wymiaru w taki sposób, aby odległości fizyczne były równe. W procedurach skalowania optymalnego izotropowego współczynnika proporcji nie można zastąpić.

Regresja dla danych jakościowych (CATREG)

Regresja dla danych jakościowych służy do ilościowej oceny danych jakościowych i polega na przypisaniu kategoriom wartości liczbowych, co pozwala na uzyskanie optymalnego równania regresji liniowej dla transformowanych zmiennych. Regresja dla danych jakościowych jest również określana akronimem CATREG (z ang. *categorical regression*).

Standardowa analiza regresji liniowej polega na zminimalizowaniu sumy kwadratów różnic pomiędzy zmienną odpowiedzi (zależną) a ważoną kombinacją zmiennych predyktora (niezależnych). Zmienne są zwykle ilościowe, przy czym (nominalne) dane jakościowe przekodowane są na zmienne binarne lub kontrasty. W rezultacie zmienne jakościowe separują grupy obserwacji, a opisywana technika estymuje odrębne zestawy parametrów dla każdej grupy. Oszacowane współczynniki odzwierciedlają sposób, w jaki zmiany w predyktorach oddziałują na odpowiedź. Przewidywanie odpowiedzi jest możliwe dla każdej kombinacji wartości predyktorów.

Alternatywna metoda polega na regresji odpowiedzi na wartości predyktorów. Dlatego dla każdej zmiennej szacowany jest tylko jeden współczynnik. Jednak w przypadku zmiennych jakościowych wartości kategorii są przypisywane arbitralnie. Zmiana kodowania kategorii powoduje uzyskanie innych współczynników, co utrudnia porównywanie analiz tych samych zmiennych.

Technika CATREG uzupełnia standardową metodę o jednoczesne skalowanie zmiennych nominalnych, porządkowych i liczbowych. Procedura kwantyfikuje zmienne jakościowe, tak że kwantyfikacje odzwierciedlają cechy pierwotnych kategorii. Procedura traktuje kwantyfikowane zmienne jakościowe

tak samo, jak zmienne liczbowe. Użycie transformacji nieliniowych umożliwia analizowanie zmiennych na różnych poziomach w celu znalezienia najlepiej dopasowanego modelu.

Przykład. Regresji jakościowej można użyć np. do opisu zależności zadowolenia z pracy od kategorii stanowiska, regionu geograficznego i ilości podróży służbowych. Może się okazać, że wysoki poziom zadowolenia koresponduje ze stanowiskami menedżerskimi i niewielką ilością podróży. Wynikowe równanie regresji można wykorzystać do przewidywania zadowolenia z pracy na podstawie dowolnej kombinacji trzech zmiennych niezależnych.

Statystyki i wykresy. Częstości, współczynniki regresji, tabela ANOVA, historia iteracji, kwantyfikacje kategorii, korelacje między nieprzekształconymi predyktorami, korelacje między przekształconymi predyktorami, wykresy reszt i wykresy przekształceń.

Wymagania dotyczące danych w regresji dla danych jakościowych

Dane. Procedura CATREG operuje na zmiennych będących wskaźnikami kategorii. Wskaźniki kategorii powinny być dodatnimi liczbami całkowitymi. W oknie dialogowym Dyskretyzacja można przekształcić zmienne o wartościach ułamkowych i łańcuchowych w dodatnie liczby całkowite.

Założenia. Dozwolona jest tylko jedna zmienna odpowiedzi, ale maksymalna liczba predyktorów wynosi 200. Dane muszą zawierać co najmniej trzy ważne obserwacje, a liczba ważnych obserwacji musi być większa od liczby predyktorów powiększonej o jeden.

Procedury pokrewne. Procedura CATREG jest równoważna analizie korelacji kanonicznych dla danych jakościowych ze skalowaniem optymalnym (OVERALS) na dwóch zbiorach, z których jeden zawiera tylko jedną zmienną. Skalowanie wszystkich zmiennych na poziomie ilościowym odpowiada standardowej analizie regresji wielokrotnej.

Uzyskiwanie regresji dla danych jakościowych

1. Wybierz z menu następującą opcję:

Analiza > Regresja > Skalowanie optymalne (CATREG)...

2. Wybierz zmienną zależną i zmienne niezależne.

3. Kliknij **OK**.

Opcjonalnie zmień poziom skalowania dla każdej zmiennej.

Definiowanie skali w regresji dla danych jakościowych

Można ustawić optymalny poziom skalowania dla zmiennych zależnych i niezależnych. Domyślnie są one skalowane jako monotoniczne krzywe sklejące drugiego stopnia (porządkowe) z dwoma węzłami wewnętrznymi. Ponadto można ustawić wagi zmiennych analizy.

Poziom skalowania optymalnego. Możliwe jest również wybranie poziomu skalowania do kwantyfikacji każdej zmiennej.

- **Sklejana w porządkowym.** Kolejność kategorii obserwowanej zmiennej jest zachowywana w optymalnie skalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikiem przekształcenia jest wygładzony, niemalejący wielomian składany wybranego stopnia. Fragmenty są określone na podstawie liczby podanej przez użytkownika i przez procedurę rozmieszczania węzłów wewnętrznych.
- **Sklejana w nominalnym.** Jedyne informacje z obserwowanej zmiennej, które są zachowywane w optymalnie wyskalowanej zmiennej, to grupowanie obiektów w kategoriach. Kolejność kategorii obserwowanej zmiennej nie jest zachowywana. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikiem przekształcenia jest wygładzony, potencjalnie niemalejący wielomian składany wybranego stopnia. Fragmenty są określone na podstawie liczby podanej przez użytkownika i przez procedurę rozmieszczania węzłów wewnętrznych.
- **Porządkowe.** Porządek kategorii obserwowanej zmiennej jest zachowany w optymalnie wyskalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikowe przekształcenie jest lepiej dopasowane niż przekształcenie Sklejana w porządkowym, ale jest mniej gładkie.

- **Nominalne.** Jedyne informacje z obserwowanej zmiennej, które są zachowywane w optymalnie wyskalowanej zmiennej, to grupowanie obiektów w kategoriach. Kolejność kategorii obserwowanej zmiennej nie jest zachowywana. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikowe przekształcenie jest lepiej dopasowane niż przekształcenie Sklejana w nominalnym, ale jest mniej gładkie.
- **Numeryczna.** Kategorie traktowane są jako uporządkowane i równoodległe (poziom przedziałów). Kolejność kategorii i równe odległości między numerami kategorii obserwowanej zmiennej są zachowane w optymalnie wyskalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Kiedy wszystkie zmienne są na poziomie liczbowym, analiza jest analogiczna do standardowej analizy składowych głównych.

Dyskretyzacja regresji dla danych jakościowych

Okno dialogowe Dyskretyzacja pozwala na wybór metody rekodowania zmiennych. Zmienne o wartościach utamkowanych są grupowane w siedem kategorii (lub w tyle kategorii, ile jest odrębnych wartości zmiennych, jeśli liczba ta jest mniejsza od siedmiu) z rozkładem w przybliżeniu normalnym, o ile nie określono inaczej. Zmienne łańcuchowe są zawsze przekształcane w dodatnie liczby całkowite przez przypisanie indyktorów kategorii zgodnie z rosnącym porządkiem alfanumerycznym. Dyskretyzacja zmiennych łańcuchowych odnosi się do tych liczb całkowitych. Inne zmienne domyślnie nie są uwzględniane. Zmienne zdyskretyzowane są następnie używane w analizie.

Metoda. Wybierz grupowanie, rangowanie albo mnożenie.

- **Grupowanie.** Wykonaj rekodowanie na określonej liczbie kategorii lub rekodowanie według przedziałów.
- **Rangowanie.** Zmienna zostanie zdyskredytowana przez rangowanie obserwacji.
- **Mnożenie.** Bieżące wartości zmiennej są standaryzowane, mnożone przez 10, zaokrąglane i dodawana jest do nich stała, tak aby najniższa zdyskretyzowana wartość wynosiła 1.

Grupowanie. Przy dyskretyzacji zmiennych przez grupowanie dostępne są następujące opcje:

- **Liczba kategorii.** Określ liczbę kategorii oraz informację o tym, czy wartości zmiennej powinny być zgodne z rozkładem w przybliżeniu normalnym czy rozkładem jednostajnym w obrębie tych kategorii.
- **Równe przedziały.** Zmienne są rekodowane na kategorie zdefiniowane przez te same przedziały o jednakowej wielkości. Należy określić długość przedziałów.

Braki danych w regresji dla danych jakościowych

Okno dialogowe Braki danych pozwala na wybór strategii postępowania z brakami danych w zmiennych analizowanych i zmiennych pasywnych.

Strategia. Wybierz tę opcję, aby wykluczyć obiekty z brakami danych (usuwanie obserwacjami) albo podstawiać brakujące wartości (aktywne uzupełnianie).

- **Wyklucz obiekty z brakami danych w tej zmiennej.** Obiekty z brakami danych w wybranej zmiennej są wykluczane z analizy. Ta strategia nie jest dostępna w przypadku zmiennych pasywnych.
- **Podstaw braki danych.** Wartości obiektów z brakami danych w wybranej zmiennej są podstawiane. Można wybrać metodę podstawiania. Wybierz opcję **Dominanta**, aby zastąpić braki danych najczęstszymi wartościami kategorii. Jeśli istnieje wiele dominant, używana jest ta z najniższym indykatorem kategorii. Wybierz opcję **Dodatkowa kategoria**, aby zastąpić brakujące wartości tą samą kwantyfikacją kategorii pasywnej. Oznacza to, że obiekty z brakami danych w tej zmiennej uznaje się za należące do tej samej (pasywnej) kategorii.

Opcje regresji dla danych jakościowych

W oknie dialogowym Opcje można wybrać styl wstępnej konfiguracji, określić kryteria iteracji i zbieżności, wybrać obiekty pasywne i skonfigurować etykiety wykresów.

Obiekty pasywne. Dzięki temu można określić obiekty, które mają być traktowane jako pasywne. Wystarczy wpisać numer obiektu pasywnego (lub określić zakres obserwacji), a następnie kliknąć przycisk **Dodaj**. Nie można przypisywać wag do obiektów pasywnych (określone wagi są ignorowane).

Konfiguracja wstępna. Jeśli żadna zmienna nie jest traktowana jako nominalna, wybierz konfigurację **Numeryczna**. Jeśli co najmniej jedna zmienna jest traktowana jako nominalna, wybierz konfigurację **Losowa**.

Jeśli co najmniej jedna zmienna ma poziom skalowania Porządkowy lub Sklejana w porządkowym, to standardowy algorytm dopasowywania modelu może spowodować, że rozwiązanie nie będzie optymalne. Wybranie opcji **Wielokrotne starty systematyczne** ze wszystkimi możliwymi schematami znaków zawsze doprowadzi do znalezienia optymalnego rozwiązania, ale niezbędny czas przetwarzania szybko wydłuży się wraz ze wzrostem liczby zmiennych porządkowych i sklejanych w porządkowym w zbiorze danych. Można zmniejszyć liczbę schematów do przetestowania, określając wartość procentową progu utraty wariancji, przy czym wyższy próg będzie wykluczać więcej schematów znaków. Ta opcja nie gwarantuje optymalnego rozwiązania, ale zmniejsza szansę uzyskania rozwiązania suboptymalnego. Ponadto, jeśli optymalne rozwiązanie nie zostanie znalezione, szansa, że rozwiązanie suboptymalne bardzo różni się od optymalnego, zostaje ograniczona. W przypadku wybrania opcji wielu startów systematycznych znaki współczynników regresji dla każdego startu są zapisywane w zewnętrznym pliku danych IBM SPSS Statistics lub zbiorze danych w bieżącej sesji. Więcej informacji można znaleźć w temacie [“Regresja dla danych jakościowych: Zapisz”](#) na stronie 14.

Wyniki poprzedniego uruchomienia z wieloma systematycznymi startami umożliwiają **użycie stałych znaków dla współczynników regresji**. Znaki (wskazywane przez liczby 1 i -1) muszą znajdować się w wierszu określonego zbioru danych lub pliku. Całkowitoliczbowy numer początkowy jest numerem obserwacji wiersza w tym pliku, który zawiera właściwe znaki.

Kryteria. Można określić maksymalną liczbę iteracji, jaką może przejść procedura regresji w obliczeniach. Można także wybrać wartość kryterium zbieżności. Procedura regresji kończy iteracje, jeśli różnica w łącznym dopasowaniu między dwiema ostatnimi iteracjami jest mniejsza od wartości zbieżności lub jeśli osiągnięta zostanie maksymalna liczba iteracji.

Użyj do opisu wykresów. Umożliwia określenie, czy na wykresach mają być używane zmienne i etykiety wartości, czy nazwy zmiennych i wartości. Można również określić maksymalną długość etykiet.

Regresja dla danych jakościowych: Regularyzacja

Metoda. Metody regularyzacji mogą poprawić błąd predykcji modelu poprzez zmniejszenie zmienności w oszacowaniach współczynnika regresji przez zmniejszenie wartości szacunkowych do 0. Regresja Lasso i Elastic Net zmniejszą pewne szacunki współczynnika do dokładnie 0, zapewniając w ten sposób formę zmiennego wyboru. W przypadku żądania użycia regularyzacji model regularyzowany i współczynniki dla każdej wartości współczynnika kary są zapisywane w zewnętrznym pliku IBM SPSS Statistics lub zbiorze danych w bieżącej sesji. Więcej informacji można znaleźć w temacie [“Regresja dla danych jakościowych: Zapisz”](#) na stronie 14.

- **Regresja grzbietowa.** Regresja grzbietowa zmniejsza współczynniki, wprowadzając składnik kary równy sumie kwadratów współczynników pomnożonej przez **współczynnik kary**. Współczynnik ten może mieć wartość od 0 (bez kary) do 1; procedura poszuka „najlepszej” wartości kary, jeśli użytkownik określi zakres i przyrost.
- **Lasso.** Składnik kary w metodzie Lasso jest oparty na sumie współczynników bezwzględnych, a specyfikacja współczynnika kary jest podobna, jak w regresji grzbietowej, jednak, metoda Lasso jest bardziej intensywna obliczeniowo.
- **Elastyczna sieć.** Metoda Elastyczna sieć łączy kary regresji Lasso i regresji grzbietowej i przeszukuje tabelę określonych wartości w celu znalezienia „najlepszych” współczynników kary regresji Lasso i regresji grzbietowej. Dla danej pary kar regresji Lasso i regresji grzbietowej Elastyczna sieć nie jest dużo bardziej intensywna obliczeniowo niż Lasso.

Wyświetl wykresy regularyzacji. Są to wykresy współczynników regresji w funkcji kary regularyzacji. Podczas wyszukiwania zakresu wartości „najlepszego” współczynnika kary obrazuje to zmiany współczynników regresji w tym zakresie.

Wykresy sieci elastycznej. W przypadku metody Elastyczna sieć wykresy regularyzacji tworzone są na podstawie wartości kary regresji grzbietowej. Opcja **Wszystkie możliwe wykresy** wykorzystuje każdą wartość z zakresu określonego przez podane minimalne i maksymalne wartości kary regresji grzbietowej.

Opcja **Dla niektórych kar regresji grzbietowej** umożliwia określenie podzbioru wartości w zakresie między minimum a maksimum. Wystarczy wpisać numer wartości kary (lub określić zakres wartości), a następnie kliknąć przycisk **Dodaj**.

Wyniki regresji dla danych jakościowych

W oknie wyników można wybrać statystyki, które mają być wyświetlane w wynikach.

Tabele. Generuje tabele do:

- **Wielorakie R.** Włącza do wyników R^2 , skorygowany R^2 i skorygowany R^2 z uwzględnieniem optymalnego skalowania.
- **ANOVA.** Ta opcja powoduje włączenie do wyników regresji i sumy kwadratów reszt, średnich kwadratów i F . Prezentowane są dwie tabele ANOVA: jedna z liczbą stopni swobody regresji równą liczbie predyktorów i jedna z liczbą stopni swobody regresji określoną z uwzględnieniem skalowania optymalnego.
- **Współczynniki.** Ta opcja generuje trzy tabele: tabelę Współczynniki zawierającą wartości beta, błędy standardowe wartości beta, wartości t i istotność; tabelę Współczynnik-skalowanie optymalne zawierającą błąd standardowy wartości beta z uwzględnieniem liczby stopni swobody dla skalowania optymalnego; oraz tabelę z korelacją rzędu zerowego, części i cząstkową, miarę względnej wartości Pratta dla przekształconych predyktorów oraz tolerancję przed i po przekształceniu.
- **Przebieg iteracji.** Dla każdej iteracji, w tym dla wartości początkowych algorytmu, wyświetlana jest wartość wielorakiego R i błędu regresji. Począwszy od pierwszej iteracji podawany jest przyrost wielorakiego R .
- **Korelacje oryginalnych zmiennych.** Prezentowana jest macierz obrazująca korelacje między nieprzekształconymi zmiennymi.
- **Korelacje przekształconych zmiennych.** Prezentowana jest macierz obrazująca korelacje między przekształconymi zmiennymi.
- **Regularyzowane modele i współczynniki.** Przedstawia wartości kar, R -kwadrat i współczynniki regresji dla każdego regularyzowanego modelu. W przypadku określenia metody ponownego próbkowania lub określenia obiektów pasywnych (obserwacji testowych) wyświetlany jest również błąd predykcji lub błąd średniokwadratowy testu.

Ponowne próbkowanie. Metody ponownego próbkowania generują estymację błędu predykcji modelu.

- **Walidacja krzyżowa.** Walidacja krzyżowa dzieli próbę na kilka podprób (złożeń). Następnie generowane są modele regresji dla danych jakościowych, kolejno z wykluczeniem dane z poszczególnych podprób. Pierwszy model jest oparty na wszystkich obserwacjach z wyjątkiem tych w pierwszej krotności próby; drugi model jest oparty na wszystkich obserwacjach z wyjątkiem drugiej krotności próby itd. Dla każdego modelu szacowany jest błąd predykcji poprzez zastosowanie modelu do podpróby wykluczonej podczas generowania modelu.
- **632 Bootstrap.** W przypadku metody bootstrap obserwacje są pobierane z danych losowo ze zwracaniem, a proces ten jest powtarzany pewną liczbę razy w celu uzyskania szeregu prób bootstrapowych. Model jest dopasowany do każdej próby bootstrapowej. Błąd predykcji dla każdego modelu jest szacowany przez zastosowanie dopasowanego modelu do obserwacji, które nie należą do próby bootstrapowej.

Kwantyfikacje kategorii. Prezentowane są tabele zawierające przekształcone wartości wybranych zmiennych.

Statystyki opisowe. Prezentowane są tabele zawierające częstości, braki danych i dominanty wybranych zmiennych.

Regresja dla danych jakościowych: Zapisz

Okno dialogowe Zapisz umożliwia zapisanie wartości przewidywanych, reszt i przekształconych wartości do aktywnego zbioru danych i/lub zapisanie zdyskretyzowanych danych, przekształconych wartości,

regularyzowanych modeli i współczynników oraz znaków współczynników regresji w zewnętrznym pliku danych IBM SPSS Statistics lub w zbiorze danych w bieżącej sesji.

- Zbiory danych są dostępne podczas bieżącej sesji, ale nie są dostępne w kolejnych sesjach, chyba że zostaną wprost zapisane jako pliki danych. Nazwy zbiorów danych muszą być zgodne z regułami nazewnictwa zmiennych.
- Nazwy plików lub nazwy zbiorów danych muszą być różne dla każdego typu zapisywanych danych.

Regularyzowane modele i współczynniki są zapisywane za każdym razem, gdy w oknie dialogowym Regularyzacja wybierana jest metoda regularyzacji. Domyślnie procedura tworzy nowy zestaw danych o unikalnej nazwie, ale można oczywiście określić własną nazwę lub zapis do pliku zewnętrznego.

Znaki współczynników regresji są zapisywane za każdym razem, gdy jako początkowa konfiguracja w oknie dialogowym Opcje używana jest opcja Wielokrotne starty systematyczne. Domyślnie procedura tworzy nowy zestaw danych o unikalnej nazwie, ale można oczywiście określić własną nazwę lub zapis do pliku zewnętrznego.

Regresja dla danych jakościowych: Wykresy przekształceń

Okno dialogowe Wykresy pozwala na określenie zmiennych, na podstawie których będą generowane wykresy przekształceń i reszt.

Wykresy przekształceń. Dla każdej z tych zmiennych, kwantyfikacje kategorii są wykreślane w funkcji pierwotnych wartości kategorii. Puste kategorie są wyświetlane na osi poziomej, ale nie mają wpływu na obliczenia. Kategorie te można poznać po przerwach w linii łączącej kwantyfikacje.

Wykresy reszt. Dla każdej z tych zmiennych reszty (obliczone dla zmiennej zależnej przewidywanej na podstawie wszystkich predyktorów z wyjątkiem wskazanego predyktora) są wykreślane wobec wskaźników kategorii i optymalnych kwantyfikacji kategorii pomnożonych przez współczynnik beta dla indykatów kategorii.

Dodatkowe właściwości komendy CATREG

Można dostosować regresję dla danych jakościowych, wklejając wybrane elementy do okna edytora komend i edytować wynikającą składnię komendy CATREG. Język składni komend umożliwia również:

- Określenie rdzeni przekształconych zmiennych podczas zapisywania ich do aktywnego zbioru danych (za pomocą opcji komendy SAVE).

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

Analiza głównych składowych dla danych jakościowych (CATPCA)

Ta procedura jednocześnie określa ilościowo zmienne jakościowe przy jednoczesnym ograniczeniu wymiarowości danych. Analiza głównych składowych dla danych jakościowych jest również określana z użyciem akronimu CATPCA (z ang. *categorical principal components analysis*).

Zadaniem analizy głównych składowych jest zredukowanie oryginalnego zestawu zmiennych do najmniejszego zestawu nieskorelowanych składowych, który reprezentuje większość informacji znajdujących się w zmiennych oryginalnych. Technika ta jest najbardziej przydatna, gdy duża liczba zmiennych uniemożliwia skuteczną interpretację relacji pomiędzy obiektami (obiekty i jednostki). Poprzez zredukowanie wymiarowości można interpretować kilka składowych zamiast dużej liczby zmiennych.

Standardowa analiza głównych składowych zakłada liniowe relacje pomiędzy zmiennymi liczbowymi. Z drugiej strony, skalowanie optymalne umożliwia skalowanie zmiennych na różnych poziomach. Zmienne jakościowe są optymalnie kwantyfikowane przy określonej wymiarowości. W wyniku tego możliwe jest modelowanie nieliniowych relacji pomiędzy zmiennymi.

Przykład. Analiza głównych składowych dla danych jakościowych (CATPCA) może zostać użyta do graficznego przedstawienia relacji pomiędzy kategorią zadania, podziałem zadania, regionem, kosztami podróży (wysokie, średnie i niskie) oraz satysfakcją z wykonania zadania. Możliwe, że dwa wymiary uwzględniają dużą liczbę wariacji. Pierwszy wymiar może oddzielać kategorię zadania od regionu,

podczas gdy drugi wymiar może oddzielać podział zadania od kosztów podróży. Może się również okazać, że wysoki poziom satysfakcji z wykonania zadania jest powiązany ze średnimi kosztami podróży.

Statystyki i wykresy. Częstości, braki danych, poziom optymalnego skalowania, tryb, wyjaśniona wariancja dla współrzędnych środka częstości, współrzędne wektora, suma dla zmiennej i dla wymiaru, ładunki składowych dla kwantyfikowanych zmiennych wektorowych, kwantyfikacje i współrzędne kategorii, przebieg iteracji, korelacje przekształconych zmiennych i wartości własnych macierzy korelacji, korelacje oryginalnych zmiennych i wartości własnych macierzy korelacji, oceny obiektów, wykresy kategorii, łączne wykresy kategorii, wykresy przekształceń, wykresy reszt, wykresy rzutowanych środków ciężkości, wykresy obiektów, łączne wykresy rozrzutu (biplot), łączne wykresy rozrzutu (triplot) i wykresy ładunków składowych.

Wymagania dotyczące danych dla analizy głównych składowych dla danych jakościowych

Dane. Wartości zmiennych łańcuchowych zawsze są przekształcane na dodatnie liczby całkowite rosnąco według kolejności alfanumerycznej. Braki danych zdefiniowane przez użytkownika, systemowe braki danych i wartości mniejsze niż 1 są traktowane jako braki danych; istnieje możliwość rekodowania lub dodawania stałych do zmiennych o wartościach mniejszych niż 1, tak aby przestały być brakami danych.

Założenia. Dane muszą zawierać co najmniej trzy ważne obserwacje. Analiza jest wykonywana na podstawie dodatnich danych całkowitych. Opcja dyskretyzacji automatycznie dobiera kategorię zmiennej zawierającej dane ułamkowe poprzez pogrupowanie jej wartości na kategorie z rozkładem zbliżonym do normalnego i automatycznie przekształca wartości zmiennych łańcuchowych na dodatnie liczby całkowite. Można określić więcej schematów dyskretyzacji.

Procedury pokrewne. Skalowanie wszystkich zmiennych na poziomie numerycznym odpowiada standardowej analizie głównych składowych. W przypadku użycia zmiennych przekształconych w standardowej liniowej analizie głównych składowych dostępne są alternatywne funkcje tworzenia wykresów. Jeśli wszystkie zmienne mają wielokrotne nominalne poziomy skalowania, analiza głównych składowych dla danych jakościowych przebiega identycznie jak wielowymiarowa analiza korespondencji. Jeśli przedmiotem zainteresowania są zestawy zmiennych, należy użyć analizy korelacji kanonicznych (nieliniowej) dla danych jakościowych.

Aby uzyskać analizę głównych składowych dla danych jakościowych

1. Wybierz z menu następującą opcję:

Analiza > Redukcja wymiarów > Skalowanie optymalne...

2. Wybierz opcję **Niektóre zmienne nie są nominalne wielokrotne**.

3. Wybierz opcję **Jeden zestaw**.

4. Kliknij przycisk **Definiuj**.

5. Wybierz co najmniej dwie zmienne analizy i określ liczbę wymiarów w rozwiązaniu.

6. Kliknij **OK**.

Opcjonalnie można określić zmienne dodatkowe, które są dopasowane do rozwiązania lub zmienne etykietujące wykresów.

Definiowanie skali i wagi w procedurze CATPCA

Możliwe jest ustawienie poziomu skalowania optymalnego dla zmiennych analizy i zmiennych pasywnych. Domyślnie są one skalowane jako monotoniczne krzywe sklejące drugiego stopnia (porządkowe) z dwoma węzłami wewnętrznymi. Ponadto można ustawić wagi zmiennych analizy.

Waga zmiennej. Dla każdej zmiennej można zdefiniować wagę. Waga musi być dodatnią liczbą całkowitą. Wartość domyślna to 1.

Poziom skalowania optymalnego. Możliwe jest również wybranie poziomu skalowania, który ma być używany do kwantyfikacji każdej zmiennej.

- **Sklejana w porządkowym.** Kolejność kategorii obserwowanej zmiennej jest zachowywana w optymalnie skalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikiem przekształcenia jest wygładzony, niemalejący

wielomian składany wybranego stopnia. Fragmenty są określone na podstawie liczby podanej przez użytkownika i przez procedurę rozmieszczania węzłów wewnętrznych.

- **Sklejana w nominalnym.** Jedyne informacje z obserwowanej zmiennej, które są zachowywane w optymalnie wyskalowanej zmiennej, to grupowanie obiektów w kategoriach. Kolejność kategorii obserwowanej zmiennej nie jest zachowywana. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikiem przekształcenia jest wygładzony, potencjalnie niemalejący wielomian składany wybranego stopnia. Fragmenty są określone na podstawie liczby podanej przez użytkownika i przez procedurę rozmieszczania węzłów wewnętrznych.
- **Nominalny wielokrotnie.** Jedyne informacje z obserwowanej zmiennej, które są zachowywane w optymalnie wyskalowanej zmiennej, to grupowanie obiektów w kategoriach. Kolejność kategorii obserwowanej zmiennej nie jest zachowywana. Kategorie znajdują się w centroidach obiektów w poszczególnych kategoriach. Opcja *Wielokrotnie* wskazuje, że dla każdego wymiaru uzyskano inne zestawy kwantyfikacji.
- **Porządkowe.** Porządek kategorii obserwowanej zmiennej jest zachowany w optymalnie wyskalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikowe przekształcenie jest lepiej dopasowane niż przekształcenie Sklejana w porządkowym, ale jest mniej gładkie.
- **Nominalne.** Jedyne informacje z obserwowanej zmiennej, które są zachowywane w optymalnie wyskalowanej zmiennej, to grupowanie obiektów w kategoriach. Kolejność kategorii obserwowanej zmiennej nie jest zachowywana. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Wynikowe przekształcenie jest lepiej dopasowane niż przekształcenie Sklejana w nominalnym, ale jest mniej gładkie.
- **Numeryczna.** Kategorie traktowane są jako uporządkowane i równoodległe (poziom przedziałów). Kolejność kategorii i równe odległości między numerami kategorii obserwowanej zmiennej są zachowane w optymalnie wyskalowanej zmiennej. Kategorie będą znajdować się na linii prostej (wektorze), przechodzącej przez początek układu. Kiedy wszystkie zmienne są na poziomie liczbowym, analiza jest analogiczna do standardowej analizy składowych głównych.

Analiza głównych składowych dla danych jakościowych: Dyskretyzacja

Okno dialogowe Dyskretyzacja pozwala na wybór metody rekodowania zmiennych. Zmienne o wartościach ułamkowych są grupowane w siedem kategorii (lub w tyle kategorii, ile jest odrębnych wartości zmiennych, jeśli liczba ta jest mniejsza od siedmiu) z rozkładem w przybliżeniu normalnym, o ile nie określono inaczej. Zmienne łańcuchowe są zawsze przekształcane w dodatnie liczby całkowite przez przypisanie indyktorów kategorii zgodnie z rosnącym porządkiem alfanumerycznym. Dyskretyzacja zmiennych łańcuchowych odnosi się do tych liczb całkowitych. Inne zmienne domyślnie nie są uwzględniane. Zmienne zdyskretyzowane są następnie używane w analizie.

Metoda. Wybierz grupowanie, rangowanie albo mnożenie.

- **Grupowanie.** Wykonaj rekodowanie na określoną liczbę kategorii lub rekodowanie według przedziałów.
- **Rangowanie.** Zmienna zostanie zdyskredytowana przez rangowanie obserwacji.
- **Mnożenie.** Bieżące wartości zmiennej są standaryzowane, mnożone przez 10, zaokrąglane i dodawana jest do nich stała, tak aby najniższa zdyskretyzowana wartość wynosiła 1.

Grupowanie. Przy dyskretyzacji zmiennych przez grupowanie dostępne są następujące opcje:

- **Liczba kategorii.** Określ liczbę kategorii oraz informację o tym, czy wartości zmiennej powinny być zgodne z rozkładem w przybliżeniu normalnym czy rozkładem jednostajnym w obrębie tych kategorii.
- **Równe przedziały.** Zmienne są rekodowane na kategorie zdefiniowane przez te same przedziały o jednakowej wielkości. Należy określić długość przedziałów.

Analiza głównych składowych dla danych jakościowych: Braki danych

Użyj okna dialogowego Braki danych, aby wybrać strategię postępowania z brakami danych w zmiennych analizowanych i zmiennych pasywnych.

Strategia. Wybierz opcję wykluczania braków danych (traktowanie pasywne), podstawiania braków danych (traktowanie aktywne) lub wykluczania obiektów z brakami danych (usuwanie obserwacjami).

- **Wyklucz braki danych; dla korelacji podstawianie po kwantyfikacji.** Obiekty z brakami danych w wybranej zmiennej nie są uwzględniane w analizie tej zmiennej. Jeśli wszystkie zmienne traktowane są pasywnie, to obiekty z brakami danych we wszystkich zmiennych są traktowane jako pasywne. Jeśli w oknie dialogowym wyników określono korelacje, to (po analizie) braki danych są podstawiane najczęściej występującą kategorią (dominantą) zmiennej dla korelacji pierwotnych zmiennych. W przypadku korelacji zmiennych optymalnie wyskalowanych można wybrać metodę podstawiania.
 - **Dominanta.** Zastąp braki danych dominantą optymalnie wyskalowanej zmiennej.
 - **Dodatkowa kategoria.** Zastąp braki danych kwantyfikacją pasywnej kategorii. Ustawienie to określa, że obiekty z brakami danych w tej zmiennej uznaje się za należące do tej samej (pasywnej) kategorii.
 - **Losowa kategoria.** Każdy brak danych należy przypisać zmiennej o kwantyfikowanej wartości numeru innej losowej kategorii w oparciu o brzegowe częstotliwości kategorii danej zmiennej.
- **Podstaw braki danych.** Wartości obiektów z brakami danych w wybranej zmiennej są podstawiane. Można wybrać metodę podstawiania.
 - **Dominanta.** Zastąp braki danych najczęstszą kategorią. Jeśli istnieje wiele dominant, używana jest ta z najniższym indykatorem kategorii.
 - **Dodatkowa kategoria.** Zastąp brakujące wartości taką samą kwantyfikacją dodatkowej kategorii. Ustawienie to określa, że obiekty z brakami danych w tej zmiennej uznaje się za należące do tej samej (pasywnej) kategorii.
 - **Losowa kategoria.** Zastąp braki danych zmienną z numerem innej losowej kategorii na podstawie brzegowych częstotliwości kategorii.
- **Wyklucz obiekty z brakami danych w tej zmiennej.** Obiekty z brakami danych w wybranej zmiennej są wykluczane z analizy. Ta strategia nie jest dostępna w przypadku zmiennych pasywnych.

Opcje analizy głównych składowych dla danych jakościowych

W oknie dialogowym Opcje można wybrać początkową konfigurację, określić kryteria iteracji i zbieżności, wybrać metodę normalizacji, wybrać metodę tworzenia etykiet wykresów i określić obiekty dodatkowe.

Obiekty pasywne. Określ numer obserwacji obiektu (lub numer pierwszej i ostatniej obserwacji w przypadku zakresu obiektów), który ma stać się pasywny, a następnie kliknij przycisk **Dodaj**. Wagi obserwacji dla obiektu pasywnego są ignorowane.

Metoda normalizacji. Można określić jedną z pięciu opcji normalizacji ocen obiektu i zmiennych: W danej analizie można używać tylko jednej metody normalizacji.

- **Metoda zmienna.** Ta opcja optymalizuje związki między zmiennymi. Współrzędne zmiennych w przestrzeni obiektów są ładunkami czynnikowymi (korelacjami z głównymi składowymi, takimi jak wymiary i oceny obiektów). Ta metoda jest przydatna, gdy interesuje nas przede wszystkim korelacja między zmiennymi.
- **Metoda obiektowa.** Ta opcja optymalizuje odległości między obiektami. Ta metoda jest przydatna, gdy interesują nas przede wszystkim różnice lub podobieństwa między obiektami.
- **Symetryczna.** Tej opcji normalizacji należy użyć, jeśli analiza ma ujawnić przede wszystkim relację między obiektami i zmiennymi.
- **Niezależna.** Tej opcji normalizacji należy użyć, chcąc osobno zbadać odległości między obiektami i korelacje między zmiennymi.
- **Użytkownika.** Można określić dowolną liczbę rzeczywistą z przedziału zamkniętego $[-1, 1]$. Wartość 1 odpowiada metodzie obiektowej. Wartość 0 odpowiada metodzie symetrycznej. Wartość -1 odpowiada metodzie zmiennej. Podając wartość większą niż -1 i mniejszą niż 1, można rozłożyć wartość własną między obiekty i zmienne. Ta metoda jest użyteczna do tworzenia indywidualnie opracowanych wykresów biplot lub triplot.

Kryteria. Można określić maksymalną liczbę iteracji, jaką może przejść procedura w obliczeniach. Można także wybrać wartość kryterium zbieżności. Algorytm kończy iteracje, jeśli różnica w łącznym dopasowaniu między dwiema ostatnimi iteracjami jest mniejsza od wartości zbieżności lub jeśli osiągnięta zostanie maksymalna liczba iteracji.

Użyj do opisu wykresów. Umożliwia określenie, czy na wykresach mają być używane zmienne i etykiety wartości, czy nazwy zmiennych i wartości. Można również określić maksymalną długość etykiet.

Wykresy wymiarów. Umożliwia sterowanie wymiarami wyświetlanymi w wynikach.

- **Pokaż wszystkie wymiary w rozwiązaniu.** Wszystkie wymiary w rozwiązaniu są widoczne w macierzowym wykresie rozrzutu.
- **Ogranicz liczbę przedstawianych wymiarów.** Widoczne są tylko wykreślone pary. W przypadku ograniczenia widoczności wymiarów należy wybrać najniższy i najwyższy wymiar do wykreślenia. Jako najniższy można wybrać wymiar od 1 do liczby wymiarów w rozwiązaniu minus 1. Ten wymiar jest kreślony wobec wyższych wymiarów. Jako najwyższy można wybrać wymiar od 2 do liczby wymiarów w rozwiązaniu; jest to najwyższy wymiar, który ma być używany przy kreśleniu par wymiarów. Te wybory mają zastosowanie do wszystkich żądanych wykresów wielowymiarowych.

Rotacja. Aby uzyskać wyniki rotowane, można wybrać metodę rotacji.

Uwaga: Te metody rotacji nie są dostępne, jeśli w oknie dialogowym **Metoda bootstrap** wybrano opcję **Realizacja metody bootstrapowej**.

- **Varimax.** Metoda rotacji ortogonalnej, która minimalizuje liczbę zmiennych z wysokimi ładunkami dla poszczególnych składowych. Upraszcza interpretację składowych.
- **Quartimax.** Metoda rotacji, która minimalizuje liczbę składowych potrzebnych do wyjaśnienia każdej zmiennej. Upraszcza interpretację obserwowanych zmiennych.
- **Equamax.** Metoda rotacji jest kombinacją metody Varimax, która upraszcza składowe, i metody Quartimax, która upraszcza zmienne. Minimalizowana jest liczba zmiennych, które mają wysokie ładunki na poszczególnych składowych, oraz liczba składników potrzebnych do wyjaśnienia poszczególnych zmiennych.
- **Oblimin.** Metoda rotacji ukośnej (nieortogonalnej). Gdy delta jest równa 0, składowe są najbardziej ukośne. Im bardziej ujemna staje się delta, tym się mniej ukośne stają się składowe. Wartości dodatnie pozwalają na dodatkową korelację składowych. Wartość **Delta** musi być mniejsza lub równa 0,8.
- **Promax.** Rotacja ukośna (nieortogonalna), która pozwala na skorelowanie składowych. Można ją wyliczyć szybciej niż rotację prostą Oblimin, dlatego jest ona użyteczna w przypadku dużych zbiorów danych. Dopuszczalny stopień korelacji (ukośność) zależy od parametru kappa. Wartość **Kappa** musi być większa lub równa 1 i mniejsza niż 10 000.

Configuration. Można odczytać dane z pliku zawierającego współrzędne konfiguracji. Pierwsza zmienna w pliku zawiera współrzędne dla pierwszego wymiaru. Druga zmienna zawiera współrzędne dla pierwszego wymiaru itd.

- **Początkowa.** Konfiguracja z określonego pliku, która będzie używana jako punkt początkowy analizy.
- **Stała.** Konfiguracja z określonego pliku, która będzie używana do dopasowania zmiennych. Dopasowywane zmienne muszą być wybrane jako zmienne analizowane, ale ponieważ konfiguracja jest ustalona, są traktowane jako pasywne (zatem nie trzeba ich wybierać jako zmiennych pasywnych).

Analiza głównych składowych dla danych jakościowych: Wyniki

Okno dialogowe Wyniki steruje wyświetlaniem wyników.

Oceny obiektów. Wyświetla oceny obiektów i zawiera następujące opcje:

- **Uwzględnij kategorie.** Wyświetla wskaźniki kategorii wybranych zmiennych analizowanych.
- **Etykieta dla ocen obiektów.** Z listy zmiennych określonych jako zmienne etykietujące można wybrać zmienną, która będzie źródłem etykiet obiektów.

Ładunki składowych. Wyświetla ładunki czynnikowe dla wszystkich zmiennych, które nie mają wielu nominalnych poziomów skalowania. Ładunki czynnikowe można sortować według wielkości.

Przebieg iteracji. Dla każdej iteracji prezentowana jest wariancja wyjaśniona, strata i przyrost wariancji wyjaśnionej.

Korelacje oryginalnych zmiennych. Przedstawia macierz korelacji pierwotnych zmiennych i wartości własne tej macierzy.

Korelacje przekształconych zmiennych. Przedstawia macierz korelacji zmiennych przekształconych (optymalnie wyskalowanych) i wartości własne tej macierzy.

Wariancja wyjaśniona. Wyświetla wielkość wariancji wyjaśnionej według współrzędnych środków ciężkości, współrzędnych wektora i sumy (współrzędnych środków ciężkości i wektora łącznie) dla zmiennej i dla każdego wymiaru.

Kwantyfikacje kategorii. Podaje kwantyfikację kategorii i współrzędne dla każdego z wybranych wymiarów zmiennych.

Statystyki opisowe. Wyświetla częstości, liczby brakujących wartości i dominantę wybranych zmiennych.

Analiza głównych składowych dla danych jakościowych: Zapisz

Elementy sterujące w oknie dialogowym Zapisz zapisują dane po dyskretyzacji, oceny obiektów, przekształcone wartości i inne wyniki do aktywnego zbioru danych, nowego zbioru danych w bieżącej sesji lub do pliku zewnętrznego.

- Zbiory danych są dostępne podczas bieżącej sesji, ale nie są dostępne w kolejnych sesjach, chyba że zostaną wprost zapisane jako pliki danych. Nazwy zbiorów danych muszą być zgodne z regułami nazewnictwa zmiennych.
- Nazwy plików lub nazwy zbiorów danych muszą być różne dla każdego typu zapisywanych danych.
- Zapisując oceny obiektów i przekształcone wartości w aktywnym zbiorze danych, można podać liczbę wymiarów nominalnych.
- Opcje w grupie **Elipsy ufności bootstrap** są dostępne tylko wtedy, gdy w oknie dialogowym Metoda bootstrap wybrano opcję **Realizacja metody bootstrapowej**.

Analiza głównych składowych dla danych jakościowych: Wykresy obiektów

Okno dialogowe Wykresy obiektów i zmiennych umożliwia określenie typów wykresów i zmiennych, dla których będą tworzone wykresy.

Punkty obiektów. Wyświetlany jest wykres punktów obiektów.

Obiekty i zmienne (biplot). Punkty obiektów są wykreślane na podstawie dokonanego przez użytkownika wyboru współrzędnych: ładunków czynnikowych albo środków ciężkości zmiennych.

Obiekty, ładunki i środki ciężkości (triplot). Punkty obiektów są wykreślane na podstawie środków ciężkości w przypadku zmiennych z wieloma nominalnymi poziomami skalowania i na podstawie ładunków czynnikowych w przypadku innych zmiennych.

Zmienne na łącznych wykresach rozrzutu. Można wybrać opcję użycia wszystkich zmiennych dla wykresów biplot i triplot albo wybrać podzbiór.

Etykiety obiektów. Można nakazać, aby etykietami obiektów były kategorie wybranych zmiennych (w oknie dialogowym Opcje można wybrać wartości wskaźników kategorii lub etykiety wartości). W przypadku wybrania opcji **Zmienna** generowany jest jeden wykres na każdą zmienną.

Analiza głównych składowych dla danych jakościowych (CATPCA): Wykresy kategorii

Okno dialogowe Wykresy kategorii umożliwia określenie typów wykresów i zmiennych, dla których będą tworzone wykresy.

Wykresy kategorii. Dla każdej wybranej zmiennej wykreślany jest wykres środków ciężkości i współrzędnych wektora. W przypadku zmiennych z wieloma nominalnymi poziomami skalowania

kategorie znajdują się w środku ciężkości obiektów w poszczególnych kategoriach. W przypadku wszystkich innych poziomów skalowania kategorie znajdują się na wektorze przebiegającym przez początek układu współrzędnych.

Łączne wykresy kategorii. Jest to pojedynczy wykres środków ciężkości i wektorów dla każdej wybranej zmiennej.

Wykresy przekształceń. Wyświetla wykres optymalnych kwantyfikacji kategorii wobec wskaźników kategorii. Można określić liczbę wymiarów dla zmiennych z wieloma nominalnymi poziomami skalowania; dla każdego wymiaru zostanie wygenerowany jeden wykres. Można również wyświetlić wykresy reszt dla każdej wybranej zmiennej.

Rzutuj środki ciężkości. Można wybrać zmienną i rzutować jej środki ciężkości na wybrane zmienne. Nie można rzutować na zmienne z wieloma nominalnymi poziomami skalowania. Gdy użytkownik wywoła ten wykres, wyświetlana jest również tabela ze współrzędnymi rzutowanych środków ciężkości.

Wykresy ładunków w analizie głównych składowych danych jakościowych

Okno dialogowe Wykresy ładunków umożliwia określenie, które zmienne mają być uwzględnione na wykresie, uwidocznienie środków ciężkości na wykresie ładunków oraz wyświetlenie wykresów wariancji wyjaśnionych.

Wariancja wyjaśniona. Dla każdego wymiaru wyświetlany jest wykres wariancji wyjaśnionej.

Pokaż ładunki składowych. W przypadku wybrania tej opcji wyświetlany jest wykres ładunków czynnikowych.

Zmienne na wykresie ładunków. Można wybrać opcję użycia wszystkich zmiennych dla wykresu ładunków czynnikowych lub wybrać podzbiór.

Uwzględnij środki ciężkości. Zmienne o wielu poziomach skalowania nominalnego nie mają ładunków czynnikowych, ale można uwzględnić na wykresie środki ciężkości tych zmiennych. Można wybrać opcję użycia wszystkich zmiennych nominalnych lub wybrać podzbiór.

Analiza głównych składowych dla danych jakościowych: Metoda bootstrap

Okno dialogowe Bootstrap określa parametry do analizy metodą bootstrap.

Realizacja metody bootstrapowej. Wykonuje ponowne próbkowanie metodą bootstrap.

- Jeśli wymagane są oceny ładunków, kategorii lub głównej składowej, wyświetlane są dodatkowe wykresy. Wykresy te wyświetlają punkty dla próby danych oraz ocen metodą bootstrap. Wykresy przekształceń obejmują regiony ufności. Wyświetlany jest również wykres dla wartości własnych.
- Jeśli określone jest rozwiązanie dwuwymiarowe, wyświetlane są wykresy elipsy ufności dla wartości własnych, ładunków składowych, punktów kategorii i punktów obiektów.
- Ponowne próbkowanie metodą bootstrap nie jest dostępne, jeśli w oknie dialogowym Opcje określono metodą rotacji.

Można wybrać metodę bootstrap typu **Zrównoważone** lub **Niezrównoważone**.

Liczba prób. Liczba prób bootstrapowych, które są wykorzystywane do obliczenia ocen metodą bootstrap. Wartość musi być dodatnią liczbą całkowitą.

Poziom ufności. Poziom ufności ocen metodą bootstrap wyrażony jako wartość procentowa. Wartość musi być liczbą dodatnią mniejszą od 100.

Metoda dopasowania. Dostępne są dwie opcje **Prokrusta** i **Odbicie**.

Elipsy ufności. Kontrolują obszar progę dla elips ufności na wykresach. Dla obszarów większych niż określona wartość (operator „>”), liczba elips zmniejsza się wraz ze wzrostem wartości progowej.

- Ustawienia w tej grupie są dostępne tylko wtedy, gdy liczba wymiarów, które są określone w głównym oknie dialogowym wynosi 2.

- Elipsy ufności dla wykresów ładunków są dostępne tylko wtedy, gdy w oknie dialogowym Wykresy ładunków wybrano opcję **Pokaż ładunki składowych**.
- Elipsy ufności dla wykresów obiektów są dostępne tylko wtedy, gdy w oknie dialogowym Wykresy obiektów i zmiennych wybrano opcję **Punkty obiektów**.
- Elipsy ufności dla wykresów kategorii są dostępne tylko wtedy, gdy w oknie dialogowym Wykresy kategorii wybrano opcję **Wykresy kategorii**.

Liczba punktów konturów elips. Liczba punktów wykresu użytych do narysowania każdej elipsy ufności. Im większa wartość, tym gładziej elipsa. Wartość musi być dodatnią liczbą całkowitą mniejszą lub równą 100.

Dodatkowe właściwości komendy CATPCA

Istnieje możliwość zmodyfikowania analizy głównej składowych dla danych jakościowych poprzez wklejenie zaznaczenia do okna komend i edycję uzyskanej komendy CATPCA. Język składni komend umożliwi również:

- Określenie rdzeni przekształconych zmiennych, ocen obiektów i przybliżeń podczas zapisywania ich do aktywnego zbioru danych (za pomocą opcji komendy SAVE).
- Określenie maksymalnej długości etykiety oddzielnie dla każdego wykresu (za pomocą opcji komendy PLOT).
- Określenie osobnej listy zmiennych dla wykresu reszt (za pomocą opcji komendy PLOT).

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

Nieliniowa analiza korelacji kanonicznych (OVERALS)

Nieliniowa analiza korelacji kanonicznych odpowiada jakościowej analizie korelacji kanonicznych z optymalnym skalowaniem. Celem tej procedury jest określenie, na ile podobne są do siebie zestawy zmiennych jakościowych. Nieliniowa analiza korelacji kanonicznych jest również znana pod akronimem OVERALS.

Standardowa analiza korelacji kanonicznych jest rozszerzeniem regresji wielokrotnej, w której drugi zestaw nie zawiera pojedynczej zmiennej odpowiedzi, lecz wiele zmiennych odpowiedzi. Celem jest wyjaśnienie możliwie jak największej części wariacji w relacjach między dwoma zestawami zmiennych liczbowych w przestrzeni niskowymiarowej. Początkowo zmienne w każdym zestawie są liniowo łączone w taki sposób, by uzyskane kombinacje liniowe miały maksymalną korelację. Biorąc pod uwagę te kombinacje, określa się kolejne kombinacje liniowe, które nie są skorelowane z poprzednimi kombinacjami i charakteryzują się największą możliwą korelacją.

Optymalne skalowanie rozszerza standardową analizę na trzy sposoby. Po pierwsze, metoda OVERALS dopuszcza więcej niż dwa zestawy zmiennych. Po drugie, zmienne mogą być skalowane jako nominalne, porządkowe lub liczbowe. W rezultacie możliwe jest analizowanie nieliniowych relacji pomiędzy zmiennymi. Wreszcie, zamiast maksymalizować korelacje między zestawami zmiennych, procedura porównuje zbiory z nieznanym zbiorem kompromisowym, który jest zdefiniowany przez oceny obiektów.

Przykład. Analizy korelacji kanonicznych z optymalnym skalowaniem można użyć do graficznego przedstawienia relacji między jednym zestawem zmiennych zawierającym kategorie Praca i Lata edukacji oraz innym zestawem zmiennych zawierających Region zamieszkania i Płeć. Może się okazać, że lata edukacji i region zamieszkania zapewniają lepsze rozróżnienie niż pozostałe zmienne. Można się również okazać, że lata edukacji zapewniają najlepsze rozróżnienie w pierwszym wymiarze.

Statystyki i wykresy. Częstości, środki ciężkości, przebieg iteracji, oceny obiektów, kwantyfikacja kategorii, wagi, ładunki składników, pojedyncze i wielokrotne dopasowanie, wykresy ocen obiektów, wykresy współrzędnych kategorii, wykresy ładunków składowych, wykresy środków ciężkości kategorii, wykresy przekształceń.

Wymagania dotyczące danych w nieliniowej analizie korelacji kanonicznych

Dane. Użyj liczb całkowitych do zakodowania zmiennych kategorialnych (skalowanie na poziomie nominalnym lub porządkowym). Aby zminimalizować wyniki, należy zakodować zmienne przy użyciu

kolejnych liczb całkowitych, zaczynając od 1. Zmienne skalowane na poziomie liczbowym nie powinny być rekodowane na kolejne liczby całkowite. Aby zminimalizować wyniki, od każdej wartości każdej zmiennej skalowanej na poziomie liczbowym należy odjąć najmniejszą zaobserwowaną wartość i dodać 1. Wartości ułamkowe są obcinane po przecinku.

Założenia. Zmienne można klasyfikować do dwóch lub większej liczby zestawów. Zmienne w analizie są skalowane jako wielokrotne nominalne, pojedyncze nominalne, porządkowe lub liczbowe. Maksymalna liczba wymiarów używanych w procedurze zależy od optymalnego poziomu skalowania zmiennych. Jeśli wszystkie zmienne są określone jako porządkowe, pojedyncze nominalne lub liczbowe, maksymalna liczba wymiarów jest mniejszą z następujących dwóch wartości: liczba obserwacji minus 1 albo łączna liczba zmiennych. Jeśli jednak zdefiniowano tylko dwa zestawy zmiennych, to maksymalna liczba wymiarów jest liczbą zmiennych w mniejszym zestawie. Jeśli niektóre zmienne są wielokrotnymi nominalnymi, to maksymalna liczba wymiarów jest równa łącznej liczbie kategorii wielokrotnych nominalnych powiększonej o liczbę niewielokrotnych zmiennych nominalnych pomniejszonej o liczbę wielokrotnych zmiennych nominalnych. Jeśli na przykład analiza obejmuje pięć zmiennych, z których jedna jest wielokrotną nominalną z czterema kategoriami, maksymalna liczba wymiarów to $(4 + 4 - 1)$, czyli 7. Jeśli zostanie podana liczba większa od maksymalnej, zostanie użyta wartość maksymalna.

Procedury pokrewne. Jeśli każdy zestaw zawiera jedną zmienną, nieliniowa analiza korelacji kanonicznych jest odpowiednikiem analizy głównych składowych z optymalnym skalowaniem. Jeśli każda z tych zmiennych jest nominalną wielokrotną, analiza jest odpowiednikiem wielowymiarowej analizy korespondencji. Jeśli w analizie uczestniczą dwa zestawy zmiennych, a jeden z zestawów zawiera tylko jedną zmienną, analiza jest identyczna z regresją dla danych jakościowych z optymalnym skalowaniem.

Uzyskiwanie nieliniowej analizy korelacji kanonicznych

1. Z menu wybierz:

Analiza > Redukcja wymiarów > Skalowanie optymalne...

2. Wybierz opcję **Wszystkie zmienne są nominalne wielokrotne** albo **Niektóre zmienne nie są nominalne wielokrotne**.

3. Wybierz **Wiele zestawów**.

4. Kliknij przycisk **Definiuj**.

5. Zdefiniuj co najmniej dwa zestawy zmiennych. Wybierz zmienne, które mają należeć do pierwszego zestawu. Aby przejść do następnego zestawu, kliknij przycisk **Dalej**, a następnie wybierz zmienne, które mają zostać uwzględnione w drugim zestawie. Istnieje możliwość dodania dodatkowych zestawów. Kliknij przycisk **Wstecz**, aby powrócić do poprzednio zdefiniowanego zestawu zmiennych.

6. Zdefiniuj zakres wartości i skalę pomiaru (optymalny poziom skalowania) dla każdej wybranej zmiennej.

7. Kliknij **OK**.

8. Opcjonalnie:

- Wybierz jedną lub więcej zmiennych, aby na ich podstawie tworzone były etykiety punktów na wykresach ocen obiektów. Z każdej zmiennej tworzony jest oddzielny wykres z punktami oznaczonymi wartościami tej zmiennej. Dla każdej z tych zmiennych etykiety wykresu należy zdefiniować zakres. Gdy używane jest okno dialogowe, jedna zmienna nie może być używana zarówno w analizie, jak i jako zmienna etykietująca. Jeśli chcesz, aby na wykresie ocen obiektów etykiety pochodziły ze zmiennej, która jest używana w analizie, użyj narzędzia Compute (dostępnego z menu Przekształcenia), aby utworzyć kopię tej zmiennej. Użyj nowej zmiennej do oznaczenia wykresu etykietami. Ewentualnie można użyć komend.
- Określ liczbę wymiarów, które mają zostać zawarte w rozwiązaniu. Z reguły należy wybrać tylko tyle wymiarów, ile potrzeba do objaśnienia większości zmienności. Jeśli analiza obejmuje więcej niż dwa wymiary, generowane są wykresy trójwymiarowe z pierwszych trzech wymiarów. Inne wymiary można wyświetlić, edytując wykres.

Definiuj zakres i typ skali

Należy zdefiniować zakres dla każdej zmiennej. Określona wartość maksymalna musi być liczbą całkowitą. Ułamkowe wartości danych są podczas analizy obcinane. Wartość kategorii poza określonym zakresem zostanie zignorowana w analizie. Aby zminimalizować wyniki, należy użyć funkcji Automatyczne rekodowanie (dostępnej z menu Przekształcenie), aby utworzyć kolejne kategorie zaczynające się od 1 dla zmiennych, które są traktowane jako zmienne nominalne lub porządkowe. Rekodowanie do kolejnych liczb całkowitych nie jest zalecane w przypadku zmiennych skalowanych na poziomie liczbowym. Aby zminimalizować wyniki dla zmiennych, które są traktowane jako zmienne porządkowe, od wartości dla każdej zmiennej należy odjąć wartość minimalną, a następnie dodać 1.

Należy również wybrać poziom skalowania, który ma być używany do kwantyfikacji każdej zmiennej.

- **Porządkowe.** Porządek kategorii obserwowanej zmiennej jest zachowany w optymalnie kwantyfikowanej zmiennej.
- **Nominalna pojedyncza.** W zmiennej kwantyfikowanej obiekty w tej samej kategorii otrzymują taki sam wynik.
- **Nominalna wielokrotna.** Kwantyfikacje mogą być inne w każdym wymiarze.
- **Ilościowa dyskretna.** Kategorie traktowane są jako uporządkowane i równoodległe. Różnice między numerami kategorii i kolejnością kategorii obserwowanej zmiennej są zachowane w zmiennej kwantyfikowanej.

Definiuj zakres

Należy zdefiniować zakres dla każdej zmiennej. Określona wartość maksymalna musi być liczbą całkowitą. Ułamkowe wartości danych są podczas analizy obcinane. Wartość kategorii poza określonym zakresem zostanie zignorowana w analizie. Aby zminimalizować wyniki, należy użyć funkcji Automatyczne rekodowanie (dostępnej z menu Przekształcenie), aby utworzyć kolejne kategorie zaczynające się od 1.

Należy również zdefiniować zakres dla każdej zmiennej, która jest używana do oznaczania wykresów ocen obiektów. Na wykresach pojawiają się jednak etykiety dla kategorii z wartościami danych, które znajdują się poza zdefiniowanym zakresem dla danej zmiennej.

Nieliniowa analiza korelacji kanonicznych: Opcje

Okno dialogowe Opcje umożliwia wybranie opcjonalnych statystyk i wykresów, zapisywanie ocen obiektów jako nowych zmiennych w aktywnym zbiorze danych, określanie iteracji i kryterium zbieżności oraz określanie początkowej konfiguracji analizy.

Pokaż. Do dostępnych statystyk należą: częstości (liczebności) brzegowe, środki ciężkości, historia iteracji, wagi i ładunki składowych, kwantyfikacje kategorii, oceny obiektów oraz statystyki pojedynczego dopasowania i wielokrotnych dopasowań.

- *Środki ciężkości.* Kwantyfikacje kategorii oraz rzutowane i rzeczywiste średnie z ocen obiektów (obserwacji) włączonych do każdego zestawu, należących do tej samej kategorii zmiennej.
- *Wagi i ładunki składowych.* Współczynniki regresji w każdym wymiarze dla każdej zmiennej kwantyfikowanej w zestawie, w którym wyniki są ustępują na zmiennych kwantyfikowanych oraz projekcji zmiennej kwantyfikowanej w obszarze obiektu. Obie wielkości wskazują na wielkość udziału każdej zmiennej w tworzeniu wymiaru w obrębie każdego zbioru.
- *Dopasowanie.* Umożliwia mierzenie dobroci dopasowania współrzędnych kategorii pojedynczych i wielokrotnych/kwantyfikacji kategorii względem obiektów.
- *Kwantyfikacje kategorii.* Wartości skali optymalnej przypisane do kategorii zmiennej.
- *Oceny obiektów.* Optymalna ocena przypisywana obiektowi (obserwacji) w konkretnym wymiarze.

Wykres. Możliwe jest generowanie wykresów współrzędnych kategorii, ocen obiektów, ładunków składowych, środków ciężkości kategorii i przekształceń.

Zapisz oceny obiektów. Oceny obiektów można zapisać jako nowe zmienne w aktywnym zbiorze danych. Oceny obiektów zapisywane są z uwzględnieniem liczby wymiarów określonych w głównym oknie dialogowym.

Użyj losowej konfiguracji początkowej. Losowej konfiguracji początkowej należy używać, jeśli niektóre lub wszystkie zmienne są pojedynczymi zmiennymi nominalnymi. Gdy ta opcja nie jest wybrana, używana jest zagnieżdżona konfiguracja początkowa.

Kryteria. Można określić maksymalną liczbę iteracji obliczeniowych, jaką może przejść procedura nieliniowej analizy korelacji kanonicznej. Można także wybrać wartość kryterium zbieżności. Analiza kończy iteracje, jeśli różnica w łącznym dopasowaniu między dwiema ostatnimi iteracjami jest mniejsza od wartości zbieżności lub jeśli osiągnięta zostanie maksymalna liczba iteracji.

Dodatkowe właściwości komendy OVERALS

Istnieje możliwość zmodyfikowania nieliniowej analizy korelacji kanonicznych poprzez wklejenie zaznaczenia do okna komend i edycję uzyskanej komendy OVERALS. Język składni komend umożliwia również:

- Określenie pary wymiarów do wykreślenia, by nie były kreślone wszystkie wyodrębnione wymiary (za pomocą słowa kluczowego NDIM opcji komendy PLOT).
- Określenie liczby znaków etykiety wartości używanej do oznaczania punktów na wykresach (za pomocą opcji komendy PLOT).
- Wyznaczenie więcej niż pięciu zmiennych jako zmiennych etykietujących dla wykresów ocen obiektów (za pomocą opcji komendy PLOT).
- Wybranie zmiennych używanych w analizie jako zmienne etykietujące dla wykresów ocen obiektów (za pomocą opcji komendy PLOT).
- Wybranie zmiennych określających etykiety punktów na wykresie oceny kwantyfikacji (za pomocą opcji komendy PLOT).
- Określenie liczby obserwacji, które mają być uwzględnione w analizie, aby nie były używane wszystkie obserwacje z aktywnego zbioru danych (za pomocą opcji komendy NOBSEVATIONS).
- Określenie trzonów nazwy zmiennych tworzonych podczas zapisywania ocen obiektów (za pomocą opcji komendy SAVE).
- Określenie liczby wymiarów do zapisania zamiast zapisywania wszystkich wyodrębnionych wymiarów (za pomocą opcji komendy SAVE).
- Zapisanie kwantyfikacji kategorii do pliku macierzy (za pomocą opcji komendy MATRIX).
- Wygenerowanie wykresów o niższej rozdzielczości, które mogą być łatwiejsze do odczytania niż zwykłe wykresy o wysokiej rozdzielczości (za pomocą komendy SET).
- Wygenerowanie wykresów środka ciężkości i wykresów przekształceń tylko dla wybranych zmiennych (za pomocą opcji komendy PLOT).

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

Analiza korespondencji

Jednym z celów analizy korespondencji jest opisanie relacji między dwiema zmiennymi nominalnymi w tabeli korespondencji w przestrzeni niskowymiarowej, przy jednoczesnym opisanu relacji między kategoriami dla każdej zmiennej. Dla każdej zmiennej odległość między punktami kategorii na wykresie odzwierciedla relację między kategoriami, przy czym podobne kategorie są wykreślone blisko siebie. Rzutowanie punktów jednej zmiennej na wektor od początku układu współrzędnych do punktu kategorii drugiej zmiennej opisuje relację między tymi zmiennymi.

Analiza tabel kontyngencji często obejmuje badanie profili wierszy i kolumn oraz testowanie niezależności przy użyciu statystyki chi-kwadrat. Jednak liczba profili może być duża, a test chi-kwadrat nie ujawnia struktury zależności. Procedura Tabele krzyżowe oferuje kilka miar związków i testów związków, ale nie pozwala na graficzne przedstawienie relacji między zmiennymi.

Analiza czynnikowa jest standardową techniką opisu relacji między zmiennymi w przestrzeni niskowymiarowej. Jednak analiza czynnikowa wymaga danych interwałowych, a liczba obserwacji powinna być pięciokrotnie większa od liczby zmiennych. Natomiast w analizie korespondencji przyjmuje się zmienne za nominalne i analiza ta pozwala na opisanie relacji między kategoriami każdej zmiennej oraz relacji między zmiennymi. Ponadto analizę korespondencji można wykorzystać do analizowania dowolnej tabeli dodatnich miar korespondencji.

Przykład. Analiza korespondencji pozwala np. na graficzne przedstawienie relacji między kategorią stanowiska a paleniem tytoniu. Może się okazać, że menedżerowie niższego szczebla pod tym względem różnią się od sekretarek, ale sekretarki nie różnią się od menedżerów wyższego szczebla. Może też okazać się, że intensywne palenie tytoniu jest powiązane z menedżerami niższego szczebla, a okazjonalne palenie z sekretarkami.

Statystyki i wykresy. Miary korespondencji, profile wierszy i kolumn, wartości osobliwe, oceny wierszy i kolumn, bezwładność, masa, statystyki ufności oceny wierszy i kolumn, oceny ufności wartości osobliwych, wykresy przekształceń, wykresy punktowe wierszy, wykresy punktowe kolumn i biploty.

Wymagania dotyczące danych do analizy korespondencji

Dane. Zmienne jakościowe przeznaczone do analizy są skalowane nominalnie. W przypadku danych zagregowanych lub dla miary korespondencji innej niż częstości należy używać zmiennej ważącej o dodatnich wartościach podobieństwa. Alternatywą, w przypadku tabeli danych, jest odczyt za pomocą komend.

Założenia. Maksymalna liczba wymiarów używanych w procedurze zależy od liczby aktywnych kategorii wierszy i kolumn oraz od liczby ograniczeń równości. Jeśli nie są stosowane ograniczenia równości i wszystkie kategorie są aktywne, maksymalna liczba wymiarów jest o jeden mniejsza od liczby kategorii zmiennej z najmniejszą liczbą kategorii. Na przykład, jeśli jedna zmienna ma pięć kategorii, a druga cztery, to maksymalna liczba wymiarów wynosi trzy. Kategorie uzupełniające nie są aktywne. Na przykład, jeśli jedna zmienna ma pięć kategorii, z czego dwie są uzupełniające, a druga zmienna ma cztery kategorie, maksymalna liczba wymiarów wynosi dwa. Wszystkie zbiory kategorii podlegające ograniczeniu równości należy traktować jak jedną kategorię. Na przykład, jeśli zmienna ma pięć kategorii, z czego trzy podlegają ograniczeniu równości, przy wyznaczaniu maksymalnej liczby wymiarów zmienną tę należy traktować jako mającą trzy kategorie. Dwie kategorie są nieograniczone, a trzecia odpowiada trzem ograniczonym kategoriom. W przypadku określenia liczby wymiarów większej od maksymalnej przyjmowana jest wartość maksymalna.

Procedury pokrewne. Jeśli w analizie uczestniczą więcej niż dwie zmienne, należy użyć wielowymiarowej analizy korespondencji. Jeśli zmienne powinny być skalowane porządkowo, należy użyć analizy głównych składowych dla danych jakościowych.

Wykonywanie analizy korespondencji

1. Z menu wybierz:

Analiza > Redukcja wymiarów > Analiza korespondencji...

2. Wybierz zmienną wierszową.

3. Wybierz zmienną kolumnową.

4. Zdefiniuj zakresy dla zmiennych.

5. Kliknij **OK**.

Definiowanie zakresu zmiennej wierszowej w analizie korespondencji

Dla zmiennej wierszowej należy zdefiniować zakres. Podane wartości minimum i maksimum muszą być liczbami całkowitymi. Ułamkowe wartości danych są podczas analizy obcinane. Wartość kategorii poza określonym zakresem zostanie zignorowana w analizie.

Wszystkie kategorie są początkowo nieograniczone i aktywne. Można ograniczyć kategorie wierszy do równych innym kategoriom wierszy albo zdefiniować kategorię wiersza jako uzupełniającą.

• **Kategorie muszą być równe.** Kategorie muszą mieć równe oceny. Należy użyć ograniczeń równości, jeśli uzyskany porządek kategorii jest niepożądanym lub nieintuicyjnym. Maksymalna liczba kategorii

wierszy, które mogą być ograniczone, jest równa łącznej liczbie kategorii aktywnych wierszy minus 1. Aby narzucić różne ograniczenia równości dla zestawów kategorii, należy użyć składni. Na przykład za pomocą komend można ograniczyć kategorie 1 i 2 tak, by były sobie równe, i kategorie 3 oraz 4 tak, by były sobie równe.

- **Kategoria jest pasywna.** Kategorie uzupełniające nie wpływają na analizę, ale są odzwierciedlone w przestrzeni zdefiniowanej przez aktywne kategorie. Kategorie uzupełniające nie odgrywają roli przy definiowaniu wymiarów. Maksymalna liczba uzupełniających kategorii wierszy jest równa łącznej liczbie kategorii wierszy pomniejszonej o 2.

Definiowanie zakresu kolumn w analizie korespondencji

Dla zmiennej kolumnowej należy zdefiniować zakres. Podane wartości minimum i maksimum muszą być liczbami całkowitymi. Ułamkowe wartości danych są podczas analizy obcinane. Wartość kategorii poza określonym zakresem zostanie zignorowana w analizie.

Wszystkie kategorie są początkowo nieograniczone i aktywne. Można ograniczyć kategorie kolumn do równych innym kategoriom kolumn albo zdefiniować kategorię kolumny jako uzupełniającą.

- **Kategorie muszą być równe.** Kategorie muszą mieć równe oceny. Należy użyć ograniczeń równości, jeśli uzyskany porządek kategorii jest niepożądanym lub nieintuicyjnym. Maksymalna liczba kategorii kolumn, które mogą być ograniczone, jest równa łącznej liczbie kategorii aktywnych kolumn minus 1. Aby narzucić różne ograniczenia równości dla zestawów kategorii, należy użyć składni. Na przykład za pomocą komend można ograniczyć kategorie 1 i 2 tak, by były sobie równe, i kategorie 3 oraz 4 tak, by były sobie równe.
- **Kategoria jest pasywna.** Kategorie uzupełniające nie wpływają na analizę, ale są odzwierciedlone w przestrzeni zdefiniowanej przez aktywne kategorie. Kategorie uzupełniające nie odgrywają roli przy definiowaniu wymiarów. Maksymalna liczba uzupełniających kategorii kolumn jest równa łącznej liczbie kategorii kolumn pomniejszonej o 2.

Analiza korespondencji: Model

Okno dialogowe Model umożliwia określenie liczby wymiarów, miary odległości, metody standaryzacji i metody normalizacji.

Wymiary w rozwiązaniu. Określ liczbę wymiarów. Z reguły należy wybrać tylko tyle wymiarów, ile potrzeba do objaśnienia większości zmienności. Maksymalna liczba wymiarów zależy od liczby aktywnych kategorii używanych w analizie oraz od ograniczeń równości. Maksymalna liczba wymiarów jest mniejsza niż:

- Liczba aktywnych kategorii wierszy minus liczba kategorii wierszy ograniczonych do równości plus liczba ograniczonych zestawów kategorii wierszy.
- Liczba aktywnych kategorii kolumn minus liczba kategorii kolumn ograniczonych do równości plus liczba ograniczonych zestawów kategorii kolumn.

Miara odległości. Można wybrać miarę odległości między wierszami i kolumnami tabeli korespondencji. Wybierz jedną z następujących możliwości:

- **Chi-kwadrat.** Użyj odległości profilu ważonego, gdzie waga jest masą wierszy lub kolumn. Ta miara jest wymagana do standardowej analizy korespondencji.
- **Euklidesowa.** Użyj pierwiastka kwadratowego sumy kwadratów różnic między parami wierszy i parami kolumn.

Metoda standaryzacji. Wybierz jedną z następujących możliwości:

- **Średnie wierszy i kolumn są usuwane.** Zarówno wiersze, jak i kolumny są wycentrowane. Ta metoda jest wymagana do standardowej analizy korespondencji.
- **Średnie wierszy są usuwane.** Tylko wiersze są wycentrowane.
- **Średnie kolumn są usuwane.** Tylko kolumny są wycentrowane.

- **Sumy wierszy są wyrównywane i średnie są usuwane.** Przed wycentrowaniem wierszy ich marginesy są zrównywane.
- **Sumy kolumn są wyrównywane i średnie są usuwane.** Przed wycentrowaniem kolumn ich marginesy są zrównywane.

Metoda normalizacji. Wybierz jedną z następujących możliwości:

- **Symetryczna.** Dla każdego wymiaru oceny wierszy są średnimi ważonymi ocen kolumn podzielonymi przez dopasowaną wartość osobliwą, a oceny kolumn są ważonymi średnimi ocen wierszy podzielonymi przez dopasowaną wartość osobliwą. Metoda ta umożliwia zbadanie różnic lub podobieństw między kategoriami dwóch zmiennych.
- **Kolumnowo-wierszowa.** Odległości między punktami wierszowymi i kolumnowymi są przybliżeniami odległości w tabeli korespondencji zgodnie z wybraną miarą odległości. Tej metody należy użyć, aby zbadać różnice między kategoriami jednej lub obu zmiennych, zamiast różnice między tymi dwiema zmiennymi.
- **Wierszowa.** Odległości między punktami wierszowymi są przybliżeniami odległości w tabeli korespondencji zgodnie z wybraną miarą odległości. Oceny wierszy są średnimi ważonymi ocen kolumn. Tej metody należy użyć, aby zbadać różnice lub podobieństwa między kategoriami zmiennej wierszowej.
- **Kolumnowa.** Odległości między punktami kolumnowymi są przybliżeniami odległości w tabeli korespondencji zgodnie z wybraną miarą odległości. Oceny kolumn są średnimi ważonymi ocen wierszy. Metoda ta jest odpowiednia wtedy, gdy chce się analizować różnice lub podobieństwa pomiędzy kategoriami zmiennej kolumnowej.
- **Użytkownika.** Należy określić wartość z zakresu od -1 do 1. Wartość -1 odpowiada nazwie użytkownika kolumny. Wartość 1 odpowiada wierszowej składowej głównej. Wartość 0 odpowiada rozkładowi symetrycznemu. Pozostałe wartości rozkładają bezwładność w różnym stopniu między oceny wierszy i kolumn. Ta metoda jest użyteczna do tworzenia indywidualnie opracowanych wykresów biplot.

Analiza korespondencji: Statystyki

Okno dialogowe Statystyka umożliwia określenie generowanych wyników liczbowych.

Tabela korespondencji. Tabela krzyżowa zmiennych wejściowych z sumami brzegowymi wierszy i kolumn.

Przegląd punktów wierszowych. Dla każdej kategorii wierszy: oceny, masa, bezwładność, udział wymiaru w bezwładności i udział wymiaru w bezwładności punktu.

Przegląd punktów kolumnowych. Dla każdej kategorii kolumn: oceny, masa, bezwładność, udział wymiaru w bezwładności i udział wymiaru w bezwładności punktu.

Profile wierszy. Dla każdej kategorii wierszy rozkład po kategoriach zmiennej kolumnowej.

Profile kolumn. Dla każdej kategorii kolumn rozkład po kategoriach zmiennej kolumnowej.

Permutacje tabeli korespondencji. Tabela korespondencji zreorganizowana w taki sposób, że wiersze i kolumny znajdują się w porządku rosnącym zgodnie z ocenami w pierwszym wymiarze. Opcjonalnie można określić maksymalną liczbę wymiarów, dla których wygenerowane zostaną tabele permutowane. Generowana jest tabela permutowana dla każdego wymiaru od 1 do podanej liczby wymiarów.

Statystyki ufności dla punktów wierszowych. Uwzględnia odchylenie standardowe i korelacje dla wszystkich niezapełniających punktów wierszy.

Statystyki ufności dla punktów kolumnowych. Uwzględnia odchylenie standardowe i korelacje dla wszystkich niezapełniających punktów kolumn.

Analiza korespondencji: Wykresy

Okno dialogowe Wykresy umożliwia określenie, które wykresy są generowane.

Wykres rozrzutu. Generuje macierz wykresów wymiarów parami. Dostępne wykresy rozrzutu:

- **Biplot (Łączny wykres rozrzutu).** Generuje macierz połączonych wykresów punktów wierszy i kolumn. Jeśli wybrana jest normalizacja kolumnowo-wierszowa, biplot jest niedostępny.
- **Punkty wierszowe.** Generuje macierz wykresów punktów wierszy.
- **Punkty kolumnowe.** Generuje macierz wykresów punktów kolumn.

Opcjonalnie można określić liczbę znaków w etykietach punktów. Wartość ta musi być nieujemną liczbą całkowitą mniejszą lub równą 20.

Wykresy liniowe. Generuje wykres dla każdego wymiaru wybranej zmiennej. Dostępne wykresy liniowe:

- **Przekształcone kategorie zmiennej wierszowej.** Generuje wykres pierwotnych wartości kategorii wierszy wobec odpowiednich ocen wierszy.
- **Przekształcone kategorie zmiennej kolumnowej.** Generuje wykres pierwotnych wartości kategorii kolumn wobec odpowiednich ocen kolumn.

Opcjonalnie można określić liczbę znaków w etykietach kategorii. Wartość ta musi być nieujemną liczbą całkowitą mniejszą lub równą 20.

Wykresy wymiarów. Umożliwia wpływanie na to, które wymiary będą widoczne w wynikach.

- **Pokaż wszystkie wymiary w rozwiązaniu.** Wszystkie wymiary w rozwiązaniu są widoczne w macierzowym wykresie rozrzutu.
- **Ogranicz liczbę przedstawianych wymiarów.** Widoczne są tylko wykreślone pary. W przypadku ograniczenia widoczności wymiarów należy wybrać najniższy i najwyższy wymiar do wykreślenia. Jako najniższy można wybrać wymiar od 1 do liczby wymiarów w rozwiązaniu minus 1. Ten wymiar jest kreślony wobec wyższych wymiarów. Jako najwyższy można wybrać wymiar od 2 do liczby wymiarów w rozwiązaniu; jest to najwyższy wymiar, który ma być używany przy kreśleniu par wymiarów. Te wybory mają zastosowanie do wszystkich żądanych wykresów wielowymiarowych.

Dodatkowe właściwości komendy CORRESPONDENCE

Istnieje możliwość dostosowania analizy korespondencji poprzez wklejenie wybranych elementów do okna komend i edycję uzyskanej komendy CORRESPONDENCE. Język składni komend umożliwia również:

- Określenie danych tabeli jako danych wejściowych zamiast używania danych obserwacji (użycie opcji TABLE = ALL).
- Określenie liczby znaków etykiety wartości używanej do opisu etykietami punktów każdego typu w macierzowym wykresie rozrzutu lub w macierzowym łącznym wykresie rozrzutu (biplot) (za pomocą opcji PLOT).
- Określenie liczby znaków etykiety wartości używanej do opisu etykietami punktów każdego typu wykresu liniowego (za pomocą opcji PLOT).
- Zapisywanie macierzy ocen wierszy i kolumn w macierzowym pliku danych (za pomocą opcji OUTFILE).
- Zapisywanie macierzy statystyk ufności (wariancje i kowariancje) dla pojedynczych wartości i ocen w macierzowym pliku danych (za pomocą opcji OUTFILE).
- Określenie wielu zestawów kategorii jako równe (za pomocą opcji EQUAL).

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

Wielowymiarowa analiza korespondencji

Wielowymiarowa analiza korespondencji kwantyfikuje dane nominalne (jakościowe), przypisując wartości liczbowe do obserwacji (obiektów) i kategorii, tak aby obiekty w tej samej kategorii znajdowały się blisko siebie, a obiekty w różnych kategoriach znajdowały się daleko od siebie. Każdy obiekt znajduje się najbliżej jak to możliwe względem punktów kategorii dla kategorii, która dotyczy obiektu. W ten sposób kategorie dzielą obiekty na jednorodne podgrupy. Zmienne są uznawane za jednorodne, jeśli klasyfikują obiekty w tych samych kategoriach do tych samych podgrup.

Przykład. Wielowymiarowa analiza korespondencji pozwala np. na graficzne przedstawienie relacji między kategorią stanowiska, przynależnością do mniejszości i płcią. Może okazać się, że przynależność

do mniejszości i płeć są cechami rozróżniającymi, ale kategoria stanowiska — nie jest. Może się także okazać, że kategoria latynoska i afroamerykańska są do siebie podobne.

Statystyki i wykresy. Oceny obiektów, miary rozróżniające, historia iteracji, korelacje zmiennych pierwotnych i przekształconych, kwantyfikacje kategorii, statystyki opisowe, wykresy punktów obiektów, wykresy biplot, wykresy kategorii, połączone wykresy kategorii, wykresy przekształceń i wykresy miary rozróżniającej.

Wymagania dotyczące danych do wielowymiarowej analizy korespondencji

Dane. Wartości zmiennych łańcuchowych zawsze są przekształcane na dodatnie liczby całkowite rosnąco według kolejności alfanumerycznej. Braki danych zdefiniowane przez użytkownika, systemowe braki danych i wartości mniejsze niż 1 są traktowane jako braki danych; istnieje możliwość rekodowania lub dodawania stałych do zmiennych o wartościach mniejszych niż 1, tak aby przestały być brakami danych.

Założenia. Wszystkie zmienne mają poziom skalowania wielokrotny nominalny. Dane muszą zawierać co najmniej trzy ważne obserwacje. Analiza jest wykonywana na podstawie dodatnich danych całkowitych. Opcja dyskretyzacji automatycznie dobiera kategorię zmiennej zawierającej dane ułamkowe poprzez pogrupowanie jej wartości na kategorie z rozkładem zbliżonym do normalnego i automatycznie przekształca wartości zmiennych łańcuchowych na dodatnie liczby całkowite. Można określić więcej schematów dyskretyzacji.

Procedury pokrewne. Dla dwóch zmiennych wielowymiarowa analiza korespondencji jest analogiczna do analizy korespondencji. Jeśli zachodzi podejrzenie, że zmienne mają cechy porządkowe lub liczbowe, należy użyć Analizy głównych składowych dla danych jakościowych. Jeśli analiza ma dotyczyć zbiorów zmiennych, należy użyć nieliniowej analizy korelacji kanonicznych.

Wykonywanie wielowymiarowej analizy korespondencji

1. Z menu wybierz:

Analiza > Redukcja wymiarów > Skalowanie optymalne...

2. Wybierz opcję **Wszystkie zmienne są nominalne wielokrotne.**

3. Wybierz opcję **Jeden zestaw.**

4. Kliknij przycisk **Definiuj.**

5. Wybierz co najmniej dwie zmienne analizy i określ liczbę wymiarów w rozwiązaniu.

6. Kliknij **OK.**

Opcjonalnie można określić zmienne dodatkowe, które są dopasowane do rozwiązania lub zmienne etykietujące wykresów.

Definiowanie wagi zmiennej w wielowymiarowej analizie korespondencji

Można ustawić wagi zmiennych analizy.

Waga zmiennej. Dla każdej zmiennej można zdefiniować wagę. Waga musi być dodatnią liczbą całkowitą. Wartość domyślna to 1.

Dyskretyzacja wielowymiarowej analizy korespondencji

Okno dialogowe Dyskretyzacja pozwala na wybór metody rekodowania zmiennych. Zmienne o wartościach ułamkowych są grupowane w siedem kategorii (lub w tyle kategorii, ile jest odrębnych wartości zmiennych, jeśli liczba ta jest mniejsza od siedmiu) z rozkładem w przybliżeniu normalnym, o ile nie określono inaczej. Zmienne łańcuchowe są zawsze przekształcane w dodatnie liczby całkowite przez przypisanie indyktorów kategorii zgodnie z rosnącym porządkiem alfanumerycznym. Dyskretyzacja zmiennych łańcuchowych odnosi się do tych liczb całkowitych. Inne zmienne domyślnie nie są uwzględniane. Zmienne zdyskretyzowane są następnie używane w analizie.

Metoda. Wybierz grupowanie, rangowanie albo mnożenie.

- **Grupowanie.** Wykonaj rekodowanie na określoną liczbę kategorii lub rekodowanie według przedziałów.
- **Rangowanie.** Zmienna zostanie zdyskredytowana przez rangowanie obserwacji.

- **Mnożenie.** Bieżące wartości zmiennej są standaryzowane, mnożone przez 10, zaokrąglane i dodawana jest do nich stała, tak aby najniższa zdyskretyzowana wartość wynosiła 1.

Grupowanie. Przy dyskretyzacji zmiennych przez grupowanie dostępne są następujące opcje:

- **Liczba kategorii.** Określ liczbę kategorii oraz informację o tym, czy wartości zmiennej powinny być zgodne z rozkładem w przybliżeniu normalnym czy rozkładem jednostajnym w obrębie tych kategorii.
- **Równe przedziały.** Zmienne są rekodowane na kategorie zdefiniowane przez te same przedziały o jednakowej wielkości. Należy określić długość przedziałów.

Wielowymiarowa analiza korespondencji (MCA) – Braki danych

Okno dialogowe Braki danych pozwala na wybór strategii postępowania z brakami danych w zmiennych analizowanych i zmiennych pasywnych.

Traktowanie braków danych. Wybierz opcję wykluczania braków danych (traktowanie pasywne), podstawiania braków danych (traktowanie aktywne) lub wykluczania obiektów z brakami danych (usuwanie obserwacji).

- **Wyklucz braki danych; dla korelacji podstawianie po kwantyfikacji.** Obiekty z brakami danych w wybranej zmiennej nie są uwzględniane w analizie tej zmiennej. Jeśli wszystkie zmienne traktowane są pasywnie, to obiekty z brakami danych we wszystkich zmiennych są traktowane jako pasywne. Jeśli w oknie dialogowym wyników określono korelacje, to (po analizie) braki danych są podstawiane najczęściej występującą kategorią (dominantą) zmiennej dla korelacji pierwotnych zmiennych. W przypadku korelacji zmiennych optymalnie wyskalowanych można wybrać metodę podstawiania. Wybierz opcję **Dominanta**, aby zastąpić braki danych dominantą optymalnie wyskalowanej zmiennej. Wybierz opcję **Dodatkowa kategoria**, aby zastąpić brakujące wartości kwantyfikacją kategorii dodatkowej. Oznacza to, że obiekty z brakami danych w tej zmiennej uznaje się za należące do tej samej (pasywnej) kategorii.
- **Podstaw braków danych.** Wartości obiektów z brakami danych w wybranej zmiennej są podstawiane. Można wybrać metodę podstawiania. Wybierz opcję **Dominanta**, aby zastąpić braki danych najczęstszymi wartościami kategorii. Jeśli istnieje wiele dominant, używana jest ta z najniższym indykatorem kategorii. Wybierz opcję **Dodatkowa kategoria**, aby zastąpić brakujące wartości tą samą kwantyfikacją kategorii pasywnej. Oznacza to, że obiekty z brakami danych w tej zmiennej uznaje się za należące do tej samej (pasywnej) kategorii.
- **Wyklucz obiekty z brakami danych w tej zmiennej.** Obiekty z brakami danych w wybranej zmiennej są wykluczane z analizy. Ta strategia nie jest dostępna w przypadku zmiennych pasywnych.

Wielowymiarowa analiza korespondencji (MCA) – Opcje

W oknie dialogowym Opcje można wybrać początkową konfigurację, określić kryteria iteracji i zbieżności, wybrać metodę normalizacji, wybrać metodę tworzenia etykiet wykresów i określić obiekty dodatkowe.

Obiekty pasywne. Określ numer obserwacji obiektu (lub numer pierwszej i ostatniej obserwacji w przypadku zakresu obiektów), który ma stać się pasywny, a następnie kliknij przycisk **Dodaj**. Kontynuuj, dopóki nie zostaną wskazane wszystkie obiekty pasywne. Wagi obserwacji dla obiektu pasywnego są ignorowane.

Metoda normalizacji. Można określić jedną z pięciu opcji normalizacji ocen obiektu i zmiennych: W danej analizie można używać tylko jednej metody normalizacji.

- **Metoda zmienna.** Ta opcja optymalizuje związki między zmiennymi. Współrzędne zmiennych w przestrzeni obiektów są ładunkami czynnikowymi (korelacjami z głównymi składowymi, takimi jak wymiary i oceny obiektów). Jest to przydatne, gdy interesuje nas przede wszystkim korelacja między zmiennymi.
- **Metoda obiektowa.** Ta opcja optymalizuje odległości między obiektami. Jest to przydatne, gdy interesują nas przede wszystkim różnice lub podobieństwa między obiektami.
- **Symetryczna.** Tej opcji normalizacji należy użyć, jeśli analiza ma ujawnić przede wszystkim relację między obiektami i zmiennymi.

- **Niezależna.** Tej opcji normalizacji należy użyć, chcąc osobno zbadać odległości między obiektami i korelacje między zmiennymi.
- **Użytkownika.** Można określić dowolną liczbę rzeczywistą z przedziału zamkniętego $[-1, 1]$. Wartość 1 jest równoważna metodzie obiektowej, wartość 0 jest równoważna metodzie symetrycznej, a wartość -1 jest równoważna metodzie zmiennej. Podając wartość większą niż -1 i mniejszą niż 1, można rozłożyć wartość własną między obiekty i zmienne. Ta metoda jest użyteczna do tworzenia indywidualnie opracowanych wykresów biplot lub triplot.

Kryteria. Można określić maksymalną liczbę iteracji, jaką może przejść procedura w obliczeniach. Można także wybrać wartość kryterium zbieżności. Algorytm kończy iteracje, jeśli różnica w łącznym dopasowaniu między dwiema ostatnimi iteracjami jest mniejsza od wartości zbieżności lub jeśli osiągnięta zostanie maksymalna liczba iteracji.

Użyj do opisu wykresów. Umożliwia określenie, czy na wykresach mają być używane zmienne i etykiety wartości, czy nazwy zmiennych i wartości. Można również określić maksymalną długość etykiet.

Wykresy wymiarów. Umożliwia wpływ na to, które wymiary będą widoczne w wynikach.

- **Pokaż wszystkie wymiary w rozwiązaniu.** Wszystkie wymiary w rozwiązaniu są widoczne w macierzowym wykresie rozrzutu.
- **Ogranicz liczbę przedstawianych wymiarów.** Widoczne są tylko wykreślone pary. W przypadku ograniczenia widoczności wymiarów należy wybrać najniższy i najwyższy wymiar do wykreślenia. Jako najniższy można wybrać wymiar od 1 do liczby wymiarów w rozwiązaniu minus 1. Ten wymiar jest kreślony wobec wyższych wymiarów. Jako najwyższy można wybrać wymiar od 2 do liczby wymiarów w rozwiązaniu; jest to najwyższy wymiar, który ma być używany przy kreśleniu par wymiarów. Te wybory mają zastosowanie do wszystkich żądanych wykresów wielowymiarowych.

Configuration. Można odczytać dane z pliku zawierającego współrzędne konfiguracji. Pierwsza zmienna w pliku powinna zawierać współrzędne pierwszego wymiaru, druga zmienna powinna zawierać współrzędne drugiego wymiaru i tak dalej.

- **Początkowa.** Konfiguracja z określonego pliku będzie używana jako punkt początkowy analizy.
- **Stać.** Konfiguracja z określonego pliku będzie używana do dopasowania zmiennych. Dopasowywane zmienne muszą być wybrane jako zmienne analizowane, ale ponieważ konfiguracja jest ustalona, są traktowane jako pasywne (zatem nie trzeba ich wybierać jako zmiennych pasywnych).

Wielowymiarowa analiza korespondencji (MCA) – Wynik

Okno dialogowe wyników umożliwia generowanie tabel ocen obiektów, miar rozróżniających, historii iteracji, korelacji zmiennych pierwotnych i przekształconych, kwantyfikacji kategorii dla wybranych zmiennych i statystyk opisowych dla wybranych zmiennych.

Oceny obiektów. Wyświetla oceny obiektów, w tym masę, bezwładność i udziały, a także umożliwia korzystanie z następujących opcji:

- **Uwzględnij kategorie.** Wyświetla wskaźniki kategorii wybranych zmiennych analizowanych.
- **Etykieta dla ocen obiektów.** Z listy zmiennych określonych jako zmienne etykietujące można wybrać zmienną, która będzie źródłem etykiet obiektów.

Miary dyskryminacyjne. Wyświetla miary rozróżniające na każdą zmienną i każdy wymiar.

Przebieg iteracji. Dla każdej iteracji prezentowana jest wariancja wyjaśniona, strata i przyrost wariancji wyjaśnionej.

Korelacje oryginalnych zmiennych. Przedstawia macierz korelacji pierwotnych zmiennych i wartości własne tej macierzy.

Korelacje przekształconych zmiennych. Przedstawia macierz korelacji zmiennych przekształconych (optymalnie wyskalowanych) i wartości własne tej macierzy.

Kwantyfikacje kategorii i udziały. Podaje kwantyfikacje kategorii (współrzędne), w tym masę i bezwładność oraz udziały dla każdego wymiaru wybranych zmiennych.

Uwaga: współrzędne i udziały (w tym masa i bezwładność) są wyświetlane w oddzielnych warstwach wynikowej tabeli przestawnej, przy czym współrzędne domyślnie są wyświetlane. Aby wyświetlić udziały, należy aktywować (kliknąć dwukrotnie) tabelę i wybrać opcję Udziały z listy rozwijanej Warstwa.

Statystyki opisowe. Wyświetla częstości, liczby brakujących wartości i dominantę wybranych zmiennych.

Wielowymiarowa analiza korespondencji (MCA) – Zapisz

Okno dialogowe Zapisz umożliwia zapisanie dyskretyzowanych danych, ocen obiektów i przekształconych wartości w zewnętrznym pliku danych programu IBM SPSS Statistics lub w zbiorze danych w bieżącej sesji. Można także zapisać przekształcone wartości i oceny obiektów w aktywnym zbiorze danych.

- Zbiory danych są dostępne podczas bieżącej sesji, ale nie są dostępne w kolejnych sesjach, chyba że zostaną wprost zapisane jako pliki danych. Nazwy zbiorów danych muszą być zgodne z regułami nazewnictwa zmiennych.
- Nazwy plików lub nazwy zbiorów danych muszą być różne dla każdego typu zapisywanych danych.
- Zapisując oceny obiektów i przekształcone wartości w aktywnym zbiorze danych, można podać liczbę wymiarów nominalnych.

Wielowymiarowa analiza korespondencji (MCA) – Wykresy obiektów

Okno dialogowe Wykresy obiektów umożliwia określenie typów wykresów i zmiennych, dla których będą tworzone wykresy.

Punkty obiektów. Wyświetlany jest wykres punktów obiektów.

Obiekty i środki ciężkości (biplot). Punkty obiektów są wykreślane razem ze środkami ciężkości zmiennych.

Zmienne na łącznych wykresach rozrzutu. Można wybrać opcję użycia wszystkich zmiennych dla wykresu biplot albo wybrać podzbiór.

Etykiety obiektów. Można nakazać, aby etykietami obiektów były kategorie wybranych zmiennych (w oknie dialogowym Opcje można wybrać wartości wskaźników kategorii lub etykiety wartości). W przypadku wybrania opcji **Zmienna** generowany jest jeden wykres na każdą zmienną.

Wielowymiarowa analiza korespondencji (MCA) – Wykresy zmiennych

Okno dialogowe Wykresy zmiennych umożliwia określenie typów wykresów i zmiennych, dla których będą tworzone wykresy.

Wykresy kategorii. Dla każdej wybranej zmiennej wykreślany jest wykres współrzędnych środków ciężkości. Kategorie znajdują się w środkach ciężkości obiektów w poszczególnych kategoriach.

Łączne wykresy kategorii. Jest to pojedynczy wykres środków ciężkości dla każdej wybranej zmiennej.

Wykresy przekształceń. Wyświetla wykres optymalnych kwantyfikacji kategorii wobec wskaźników kategorii. Można określić liczbę wymiarów; dla każdego wymiaru zostanie wygenerowany jeden wykres. Można również wyświetlić wykresy reszt dla każdej wybranej zmiennej.

Miary dyskryminacyjne. Generuje jeden wykres miar różnicujących dla wybranych zmiennych.

Dodatkowe właściwości komendy MULTIPLE CORRESPONDENCE

Wielowymiarową analizę korespondencji można dostosować, wklejając wybrane elementy do okna edytora komend i edytować wynikającą składnię komendy MULTIPLE CORRESPONDENCE. Język składni komend umożliwia również:

- Określenie rdzeni przekształconych zmiennych, ocen obiektów i przybliżeń podczas zapisywania ich do aktywnego zbioru danych (za pomocą opcji komendy SAVE).
- Określenie maksymalnej długości etykiety oddzielnie dla każdego wykresu (za pomocą opcji komendy PLOT).

- Określenie osobnej listy zmiennych dla wykresu reszt (za pomocą opcji komendy PLOT).

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

Skalowanie wielowymiarowe (PROXSCAL)

Procedura skalowania wielowymiarowego służy do znajdowania struktury w zbiorze miar bliskości pomiędzy obiektami. Jest to możliwe dzięki przypisywaniu obserwacji do poszczególnych miejsc w niskowymiarowej przestrzeni pojęciowej w taki sposób, że odległości między punktami w przestrzeni możliwie blisko odpowiadają danym miarom (nie)podobieństwa. Wynikiem jest reprezentacja najmniejszych kwadratów obiektów w przestrzeni niskowymiarowej, co (w wielu przypadkach) pomaga lepiej zrozumieć dane.

Przykład. Skalowanie wielowymiarowe może okazać się bardzo pomocne przy określaniu relacji postrzegalnych. Przykładowo, w wypadku wizerunku produktu można przeprowadzić badanie, które dostarczy zbiór danych opisujących dostrzegane podobieństwo (lub bliskość) między wybranym produktem a towarami konkurencyjnymi. Używając takich miar bliskości i niezależnych zmiennych (np. ceny), można spróbować określić, które zmienne decydują o tym, jak konsumenci postrzegają dany produkt, a następnie dostosować jego wizerunek do wyników takiej analizy.

Statystyki i wykresy. Przebieg iteracji, miary STRESS, dekompozycja STRESS, współrzędne przestrzeni wspólnej, odległości obiektów w ostatecznej konfiguracji, wagi indywidualnych przestrzeni, przestrzenie indywidualne, przekształcone odległości, przekształcone zmienne niezależne, wykresy STRESS, wykresy rozrzutu przestrzeni wspólnej, wykresy rozrzutu wag indywidualnych przestrzeni, wykresy rozrzutu indywidualnych przestrzeni, wykresy przekształceń, wykresy reszt Sheparda oraz wykresy przekształceń zmiennych niezależnych.

Wymagania dotyczące danych przy skalowaniu wielowymiarowym

Dane. Dane mogą być dostarczane w formie macierzy odległości lub zmiennych, które są konwertowane w takie macierze. Macierze mogą być formatowane w kolumnach lub między kolumnami. Odległości można skalować na poziomie ilorazowym, przedziałowym, porządkowym lub sklejanym.

Założenia. Należy określić co najmniej trzy zmienne. Liczba wymiarów nie może być większa niż liczba obiektów pomniejszona o jeden. Redukcja wymiarowości jest pomijana, jeśli łączy się z wieloma losowymi startami. Jeśli określone jest tylko jedno źródło, wszystkie modele stanowią ekwiwalent modelu typu IDENTITY; z tego powodu analiza domyślnie odwołuje się do modelu typu IDENTITY.

Procedury pokrewne. Skalowanie wszystkich zmiennych na poziomie ilościowym odpowiada standardowej analizie skalowania wielowymiarowego.

Wykonywanie skalowania wielowymiarowego

1. Z menu wybierz:

Analiza > Skalowanie > Skalowanie wielowymiarowe (PROXSCAL)...

Spowoduje to otwarcie okna dialogowego Format danych.

2. Określ format swoich danych:

Format danych. Określ, czy Twoje dane zawierają miary bliskości czy też chcesz wygenerować odległości z danych.

Liczba źródeł. Jeśli Twoje dane są odległościami, określ, czy masz jedno czy wiele źródeł miar bliskości.

Jedno źródło. Jeśli odległości pochodzą z jednego źródła, określ, czy Twój zbiór danych sformatowano za pomocą odległości w macierzy rozłożonej na kolumny czy w pojedynczej kolumnie z dwiema oddzielnymi zmiennymi. Pozwoli to określić wiersz i kolumnę każdej odległości.

- *Odległości w macierzy rozłożonej na kolumny.* Macierz bliskości jest rozłożona na liczbę kolumn równą liczbie obiektów. Prowadzi to do okna dialogowego Odległości w kolumnach w macierzach.

- *Odległości w pojedynczej kolumnie.* Macierz bliskości jest zwinięta do pojedynczej kolumny lub zmiennej. Niezbędne są dwie dodatkowe zmienne, identyfikujące wiersz, kolumnę komórek. Prowadzi to do okna dialogowego Odległości w jednej kolumnie.

Wiele źródeł. Jeśli odległości pochodzą z wielu źródeł, określ, czy Twój zbiór danych sformatowano za pomocą odległości w zestawionych macierzach rozłożonych na kolumny, wielu kolumnach z jednym źródłem przypisanym do każdej czy w pojedynczej kolumnie.

- *Odległości w zestawionych macierzach rozłożonych na kolumny.* Macierze zbliżeniowe są rozmieszczone w wielu kolumnach, które są równe liczbie obiektów i są ułożone nad jedną inną, przez liczbę wierszy równą liczbie obiektów, w których liczba źródeł jest równa liczbie obiektów. Prowadzi to do okna dialogowego Odległości w kolumnach w macierzach.
- *Odległości w kolumnach, jedno źródło na jedną kolumnę.* Macierze zbliżeniowe są zwijane do wielu kolumn lub zmiennych. Niezbędne są dwie dodatkowe zmienne, identyfikujące wiersz, kolumnę komórek. Prowadzi to do okna dialogowego Odległości w kolumnach.
- *Odległości zestawione w pojedynczej kolumnie.* Macierze zbliżeniowe są zwijane w jedną kolumnę lub zmienną. Niezbędne są trzy dodatkowe zmienne, identyfikujące wiersz, kolumnę i źródło komórek. Prowadzi to do okna dialogowego Odległości w jednej kolumnie.

3. Kliknij przycisk **Definiuj**.

Odległości w macierzach rozłożonych na kolumny

Po wybraniu odległości w modelu danych macierzy dla jednego lub wielu źródeł w oknie dialogowym Format danych wykonaj następujące kroki:

1. Wybierz co najmniej trzy zmienne odległości. (Upewnij się, że kolejność zmiennych na liście jest zgodna z kolejnością kolumn odległości).
2. Opcjonalnie możesz wybrać liczbę zmiennych definiujących wagi równą liczbie zmiennych odległości. (Upewnij się, że kolejność wag jest zgodna z kolejnością odległości, które opisują).
3. Oprócz tego, jeśli istnieje wiele źródeł, możesz wybrać zmienną je definiującą. (Liczba obserwacji w każdej zmiennej odległości powinna być równa iloczynowi liczby zmiennych odległości i liczby źródeł).

Oprócz tego możliwe jest określenie modelu skalowania wielowymiarowego, wyznaczenie ograniczeń przestrzeni wspólnej, wskazanie kryteriów zbieżności, określenie początkowej konfiguracji, a także wybranie wykresów i danych wyjściowych.

Bliskości w kolumnach

Jeśli w oknie dialogowym Format danych wybrano model wielu kolumn dla wielu źródeł, wykonaj następujące czynności:

1. Wybierz co najmniej dwie zmienne bliskości. Zakłada się, że każda zmienna jest macierzą odległości z oddzielnego źródła.
2. Wybierz zmienną wierszową, aby określić lokalizację wierszy bliskości w każdej zmiennej bliskości.
3. Wybierz zmienną kolumnową, aby określić lokalizację kolumn bliskości w zmiennej odległości. (Komórki macierzy odległości bez wyznaczonego wiersza/kolumny są uznawane za brakujące).
4. Opcjonalnie możesz wybrać liczbę zmiennych definiujących wagi równą liczbie zmiennych odległości.

Oprócz tego możliwe jest określenie modelu skalowania wielowymiarowego, wyznaczenie ograniczeń przestrzeni wspólnej, wskazanie kryteriów zbieżności, określenie początkowej konfiguracji, a także wybranie wykresów i danych wyjściowych.

Odległości w jednej kolumnie

Po wybraniu modelu jednej kolumny dla jednego lub wielu źródeł w oknie dialogowym Format danych wykonaj następujące kroki:

1. Wybierz zmienną odległości (Przyjmuje się, że jest to jedna lub więcej macierzy odległości).

2. Wybierz zmienną wierszową, aby określić lokalizacje wierszy dla odległości w zmiennej odległości.
3. Wybierz zmienną kolumnową, aby określić lokalizacje kolumn dla odległości w zmiennej odległości.
4. Jeśli istnieje wiele źródeł, wybierz zmienną je definiującą. (Komórki macierzy odległości bez wyznaczonego wiersza/kolumny są uznawane za brakujące w wypadku każdego źródła).
5. Opcjonalnie można też wybrać zmienną definiującą wagi.

Oprócz tego możliwe jest określenie modelu skalowania wielowymiarowego, wyznaczenie ograniczeń przestrzeni wspólnej, wskazanie kryteriów zbieżności, określenie początkowej konfiguracji, a także wybranie wykresów i danych wyjściowych.

Tworzenie bliskości na podstawie danych

Jeśli chcesz utworzyć bliskości na podstawie danych w oknie dialogowym Format danych, wykonaj następujące czynności:

1. Jeśli tworzysz odległości między zmiennymi (patrz okno dialogowe Utwórz miary na podstawie danych), wybierz co najmniej trzy zmienne. Zmienne te zostaną użyte do utworzenia macierzy odległości (jednej lub wielu, jeśli istnieje wiele źródeł). W przypadku tworzenia odległości między obserwacjami wymagana jest tylko jedna zmienna.
2. Jeśli istnieje wiele źródeł, wybierz zmienną je definiującą.
3. Opcjonalnie wybierz miarę do tworzenia bliskości.

Oprócz tego możliwe jest określenie modelu skalowania wielowymiarowego, wyznaczenie ograniczeń przestrzeni wspólnej, wskazanie kryteriów zbieżności, określenie początkowej konfiguracji, a także wybranie wykresów i danych wyjściowych.

Tworzenie miary na podstawie danych

W skalowaniu wielowymiarowym, wynik skalowania tworzony jest na podstawie niepodobieństw. W przypadku korzystania z danych o wielu zmiennych (wartości mierzonych zmiennych), w celu uzyskania wyniku skalowania wielowymiarowego konieczne jest utworzenie danych niepodobieństwa. Można określić szczegóły tworzenia miar niepodobieństw na podstawie danych.

Miara. Pozwala na określanie miary niepodobieństwa dla potrzeb analizy. W obszarze Miara dla danych wybierz jedną z alternatyw, odpowiadającą typowi analizowanych danych, a następnie z odpowiadającej jej listy rozwijanej wybierz stosowną miarę. Dostępne alternatywy to:

- **Interwałowe.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Odległość Czebyszewa, Odległość miejska, Odległość Minkowskiego lub Odległość użytkownika.
- **Liczebności.** Odległość chi-kwadrat lub Odległość phi-kwadrat.
- **Binarne.** Odległość euklidesowa, Kwadrat odległości euklidesowej, Różnica wielkości, Różnica wzoru, Wariancja lub Miara Lance'a i Williamsa.

Utwórz macierz odległości. Pozwala na wybranie jednostki analizy. Dostępne alternatywy to: Pomiędzy zmiennymi i Między obserwacjami.

Przekształcanie wartości. W niektórych przypadkach, na przykład, kiedy zmienne mierzone są przy użyciu różnych skali, wartości można wystandaryzować przed rozpoczęciem wyliczania odległości pomiędzy nimi (nie dotyczy danych binarnych). Z listy rozwijanej Standaryzacja wybierz metodę standaryzacji (jeśli standaryzacja nie jest wymagana, wybierz opcję **Brak**).

Definiowanie modelu skalowania wielowymiarowego

Okno dialogowe Model umożliwia określenie modelu skalowania, jego minimalnej i maksymalnej liczby wymiarów, struktury macierzy odległości, przekształcenia, które ma być stosowane wobec bliskości, oraz informacji o tym, czy bliskości mają być przekształcane w każdym źródle osobno, czy niezależnie od źródła.

Model skalowania. Wybierz jedną z następujących opcji:

- **Tożsamość.** Wszystkie źródła mają tę samą konfigurację.
- **Ważony euklidesowy.** Jest to model indywidualnych różnic. Każde źródło ma osobną przestrzeń, w której każdy wymiar przestrzeni wspólnej jest ważony w inny sposób.
- **Uogólniony euklidesowy.** Jest to model indywidualnych różnic. Każde źródło ma indywidualną przestrzeń uzyskaną w wyniku obrotu wspólnej przestrzeni i następującego po nim ważenia różnic wymiarów.
- **Zredukowana ranga.** Jest to uogólniony model euklidesowy, dla którego można określić rangi poszczególnych przestrzeni. Należy określić rangę, która jest większa lub równa 1 i mniejsza niż maksymalna liczba wymiarów.

Kształt. Określ, czy bliskości powinny być pobierane z dolnej, czy z górnej trójkątnej części macierzy odległości. Można też nakazać stosowanie całej macierzy — wówczas analizowana będzie suma ważona wartości z górnej i dolnej części trójkątnej. W każdym przypadku należy określić całą macierz, w tym przekątną, chociaż analizowane będą tylko wybrane jej części.

Odległości. Określ, czy macierz bliskości zawiera miary podobieństwa czy niepodobieństwa.

Przekształcenia odległości. Wybierz jedną z następujących opcji:

- **Proporcja jednostek.** Przekształcone bliskości są proporcjonalne do pierwotnych bliskości. Jest to dozwolone tylko w przypadku bliskości o wartościach dodatnich.
- **Interwałowe.** Przekształcone bliskości są proporcjonalne do pierwotnych bliskości powiększonych o wyraz wolny. Dodanie wyrazu wolnego sprawia, że wszystkie przekształcone bliskości na pewno będą dodatnie.
- **Porządkowe.** Przekształcone bliskości zachowują kolejność pierwotnych bliskości. Użytkownik określa, czy powiązane bliskości powinny pozostać powiązane, czy też można je rozdzielić.
- **Krzywa sklejana.** Przekształcenie bliskości jest gładkim, niemającym przekształceniem kawałkami wielomianowym bliskości pierwotnych. Użytkownik określa stopień wielomianu oraz liczbę węzłów wewnętrznych.

Zastosuj przekształcenia. Określ, czy porównywane mają być bliskości tylko w obrębie każdego źródła, czy też porównania nie powinny zależeć od źródła.

Wymiary. Domyślnie rozwiązanie jest obliczane w dwóch wymiarach (Minimum = 2, Maksimum = 2). Należy wybrać minimalną i maksymalną liczbę całkowitą z zakresu od 1 do liczby obiektów minus 1 (przy czym minimum nie może być większe od maksimum). Procedura oblicza rozwiązanie w maksymalnej liczbie wymiarów, a następnie krokowo zmniejsza liczbę wymiarów aż do osiągnięcia najniższej wymiarowości.

Ograniczenia skalowania wielowymiarowego

Okno dialogowe Ograniczenia umożliwia nakładanie ograniczeń na przestrzeń wspólną.

Ograniczenia przestrzeni wspólnej. Określ typ żądanego ograniczenia.

- **Bez ograniczeń.** Na przestrzeń wspólną nie są nakładane żadne ograniczenia.
- **Współrzędne stałe.** Pierwsza wybrana zmienna zawiera współrzędne obiektów w pierwszym wymiarze, druga zmienna odpowiada współrzędnym w drugim wymiarze itd. Brak wartości oznacza, że współrzędna w wymiarze jest swobodna. Liczba wybranych zmiennych musi być równa maksymalnej liczbie żądanych wymiarów.
- **Liniowa kombinacja zmiennych niezależnych.** Wspólna przestrzeń jest ograniczona do liniowej kombinacji wybranych zmiennych.

Zmienne ograniczające. Wybierz zmienne, które definiują ograniczenia wspólnej przestrzeni. Jeśli określono kombinację liniową, należy określić przekształcenie interwałowe, porządkowe lub sklepane zmiennych ograniczających. W każdym z tych przypadków liczba obserwacji w każdej zmiennej musi być równa liczbie obiektów.

Skalowanie wielowymiarowe: Opcje

Okno dialogowe Opcje umożliwia wybranie początkowego stylu konfiguracji, określenie kryteriów iteracji i zbieżności oraz wybranie algorytmów standardowych lub najszybszych.

Konfiguracja wstępna. Wybierz jedną z następujących możliwości:

- **Simplex.** Obiekty są umieszczane w tej samej odległości od siebie w maksymalnej liczbie wymiarów. Wykonywana jest jedna iteracja w celu ulepszenia tej konfiguracji wysokowymiarowej, po czym następuje redukcja liczby wymiarów w celu uzyskania początkowej konfiguracji, która będzie miała maksymalną liczbę wymiarów określoną w oknie dialogowym Model.
- **Torgerson.** Jako początkowej konfiguracji używa się klasycznego rozwiązania skalowania.
- **Pojedynczy losowy start.** Konfiguracja jest wybierana losowo.
- **Wielokrotne starty losowe.** Losowo wybieranych jest kilka konfiguracji, a konfiguracja z najniższym znormalizowanym obciążeniem surowym jest używana jako konfiguracja początkowa.
- **Użytkownika.** Użytkownik może wybrać zmienne, które zawierają współrzędne jego własnej konfiguracji początkowej. Liczba wybranych zmiennych powinna być równa określonej maksymalnej liczbie wymiarów, przy czym pierwsza zmienna odpowiada współrzędnym w wymiarze 1, druga zmienna odpowiada współrzędnym w wymiarze 2 itd. Liczba obserwacji w każdej zmiennej powinna być równa liczbie obiektów.

Kryteria iteracji. Określ wartości kryteriów iteracji.

- **Zbieżność STRESS.** Algorytm zatrzyma iteracje, gdy różnica w kolejnych wartościach znormalizowanego obciążenia surowego będzie mniejsza niż liczba określona w tym polu, która musi należeć do przedziału od 0,0 do 1,0.
- **Minimalna wartość STRESS.** Algorytm zatrzyma się, gdy znormalizowane obciążenie surowe spadnie poniżej liczby określonej w tym miejscu, która musi należeć do zakresu od 0,0 do 1,0.
- **Maksymalna liczba iteracji.** Algorytm wykona określoną liczbę iteracji, chyba że wcześniej zostanie spełnione jedno z powyższych kryteriów.
- **Użyj najszybszych algorytmów.** Wybranie tej opcji powoduje przyspieszenie działania algorytmów; nie można jej używać z modelami innymi niż model tożsamościowy lub z ograniczeniami.

Wykresy skalowania wielowymiarowego, wersja 1

Okno dialogowe Wykresy umożliwia wybranie wykresów, które będą tworzone. W tej sekcji przedstawiono opis okna dialogowego Wykresy, które jest używane, jeśli dane mają format Bliskości w kolumnach. W przypadku wykresów **Wagi indywidualnych przestrzeni, Oryginalne vs przekształcone odległości** oraz **Przekształcone bliskości wobec odległości** określa się źródła, dla których wykresy powinny być tworzone. Listą dostępnych źródeł jest lista zmiennych bliskości w głównym oknie dialogowym.

Wartość STRESS. Tworzony jest wykres znormalizowanego obciążenia surowego wobec wymiarów. Wykres ten jest tworzony tylko wtedy, gdy maksymalna liczba wymiarów jest większa niż minimalna liczba wymiarów.

Przestrzeń wspólna. Prezentowana jest macierz rozrzutu współrzędnych przestrzeni wspólnej.

Indywidualne przestrzenie. Dla każdego źródła przedstawione zostaną macierze rozrzutu współrzędnych indywidualnych przestrzeni. Jest to możliwe tylko wtedy, gdy w oknie dialogowym Model określono jeden z modeli różnic indywidualnych.

Wagi indywidualnych przestrzeni. Tworzony jest wykres rozrzutu poszczególnych wag przestrzeni. Jest to możliwe tylko wtedy, gdy w oknie dialogowym Model określono jeden z modeli różnic indywidualnych. W przypadku ważonego modelu euklidesowego na wykresie prezentowane są wagi z jednym wymiarem na każdej osi. W przypadku uogólnionego modelu euklidesowego dla każdego wymiaru tworzony jest jeden wykres, który prezentuje zarówno obrót, jak i wagę tego wymiaru. Model zredukowanej rangi tworzy ten sam wykres, co uogólniony model euklidesowy, ale z mniejszą liczbą wymiarów dla poszczególnych przestrzeni.

Oryginalne vs przekształcone odległości. Tworzone są wykresy bliskości pierwotnych wobec bliskości przekształconych.

Przekształcone bliskości wobec odległości. Wykreślane są bliskości przekształcone wobec odległości.

Przekształcone zmienne niezależne. Tworzone są wykresy przekształceń zmiennych niezależnych.

Korelacje zmiennych i wymiarów. Zostanie utworzony wykres korelacji między zmiennymi niezależnymi a wymiarami przestrzeni wspólnej.

Wykresy skalowania wielowymiarowego, wersja 2

Okno dialogowe Wykresy umożliwia wybranie wykresów, które będą tworzone. W tym temacie opisano okno dialogowe Wykresy, które jest używane w sytuacji, gdy format danych jest inny niż Odległości w kolumnach. W przypadku wykresów **Wagi indywidualnych przestrzeni**, **Oryginalne vs przekształcone odległości** oraz **Przekształcone bliskości wobec odległości** określa się źródła, dla których wykresy powinny być tworzone. Wprowadzone numery źródeł muszą być wartościami zmiennej źródłowej określonej w głównym oknie dialogowym i należeć do przedziału od 1 do liczby źródeł.

Skalowanie wielowymiarowe: Opcje

Okno dialogowe Wyniki umożliwia wybór rodzajów wyświetlanych wyników i zapisywanie niektórych z nich w oddzielnych plikach.

Pokaż. Wybierz jeden lub kilka z następujących elementów do wyświetlenia:

- **Współrzędne przestrzeni wspólnej.** Przedstawia współrzędne przestrzeni wspólnej.
- **Współrzędne indywidualnych przestrzeni.** Współrzędne indywidualnych przestrzeni są wyświetlane tylko wtedy, gdy model nie jest modelem tożsamościowym.
- **Wagi indywidualnych przestrzeni.** Przedstawia wagi indywidualnych przestrzeni tylko wtedy, gdy określono jeden z modeli indywidualnych różnic. W zależności od modelu wagi przestrzeni są dekomponowane na wagi obrotu i wagi wymiarów, które są także wyświetlane.
- **Odległości.** Przedstawia odległości między obiektami w konfiguracji.
- **Przekształcone odległości.** Przedstawia przekształcone bliskości między obiektami w konfiguracji.
- **Dane wejściowe.** Uwzględnia pierwotne bliskości oraz, o ile są obecne, wagi danych, początkową konfigurację i zakotwiczone (stałe) współrzędne zmiennych niezależnych.
- **STRESS dla każdego losowego startu.** Przedstawia wartość startową generatora liczb losowych i wartość znormalizowanego obciążenia surowego dla każdego losowego startu.
- **Przebieg iteracji.** Wyświetla przebieg iteracji algorytmu głównego.
- **Wiele miar STRESS.** Przedstawia różne wartości obciążenia. Tabela zawiera wartości znormalizowanego obciążenia surowego, Stress-I, Stress-II, S-Stress, wyjaśnione rozproszenie (DAF) oraz współczynnik Tuckera kongruencji.
- **Dekompozycja STRESS.** Przedstawia obiekty i źródła dekompozycji ostatecznego znormalizowanego obciążenia surowego, łącznie ze średnią na obiekt i średnią na źródło.
- **Przekształcone zmienne niezależne.** Jeśli zostało wybrane ograniczenie w postaci kombinacji liniowej, przedstawiane są przekształcone zmienne niezależne i odpowiadające im wagi regresji.
- **Korelacje zmiennych i wymiarów.** Jeśli została wybrana opcja ograniczenia w postaci kombinacji liniowej, wyświetlane są korelacje między zmiennymi niezależnymi i wymiary przestrzeni wspólnej.

Zapisz w nowym pliku. Istnieje możliwość zapisania współrzędnych przestrzeni wspólnej, wag poszczególnych przestrzeni, odległości i przekształconych zmiennych niezależnych w odrębnych plikach danych programu IBM SPSS Statistics.

Dodatkowe właściwości komendy PROXSCAL

Istnieje możliwość zmodyfikowania skalowania wielowymiarowego analizy odległości poprzez wklejenie zaznaczenia do okna komend i edycję uzyskanej komendy PROXSCAL. Język składni komend umożliwia również:

- Określenie odrębnych list zmiennych dla wykresów przekształceń i reszt (za pomocą opcji komendy PLOT).
- Określenie odrębnych list źródłowych dla poszczególnych wykresów wag przestrzeni, przekształceń i reszt (za pomocą opcji komendy PLOT).
- Określenie podzbioru wykresów przekształceń zmiennych niezależnych do wyświetlenia (za pomocą opcji komendy PLOT).

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

Wielowymiarowe skalowanie preferencji (PREFSCAL)

Procedura wielowymiarowego skalowania preferencji próbuje znaleźć wspólną skalę ilościową, która pozwala na wizualne zbadanie relacji między dwoma zestawami obiektów.

Przykłady. Załóżmy, że poprosiliśmy 21 osób o sklasyfikowanie 15 składników śniadania w kolejności preferencji, od 1 do 15. Za pomocą wielowymiarowego skalowania preferencji można stwierdzić, że badane osoby rozróżniają składniki śniadań według dwóch kryteriów: między pieczywem miękkim i twardym oraz między składnikami tuczącymi i nietuczącymi.

Możemy także poprosić grupę kierowców o ocenę 26 modeli samochodów na podstawie 10 atrybutów w 6-punktowej skali od 1=„całkowicie nieprawda” do 6=„zdecydowanie prawda”. Wartości uśrednione w zbiorze osób są traktowane jako podobieństwa. Korzystając z wielowymiarowego skalowania preferencji, można znaleźć skupienia podobnych modeli i atrybutów, z którymi są one najbardziej powiązane.

Statystyki i wykresy. Procedura wielowymiarowego skalowania preferencji może generować przebieg iteracji, miary obciążenia, dekompozycję obciążenia, współrzędne przestrzeni wspólnej, odległości między obiektami w końcowej konfiguracji, wagi indywidualnych przestrzeni, indywidualne przestrzenie, bliskości przekształcone, wykresy obciążenia, wykresy rozrzutu przestrzeni wspólnej, wykresy rozrzutu wag indywidualnych przestrzeni, wykresy rozrzutu indywidualnych przestrzeni, wykresy przekształceń i wykresy reszt Sheparda.

Wymagania dotyczące danych przy wielowymiarowym skalowaniu preferencji

Dane. Dane dostarczane są w postaci prostokątnych macierzy odległości. Każda kolumna jest traktowana jako oddzielny obiekt kolumnowy. Każdy wiersz macierzy odległości jest traktowany jako osobny obiekt wierszowy. Jeśli istnieje wiele źródeł bliskości, macierze są zestawione.

Założenia. Należy określić co najmniej dwie zmienne. Liczba wymiarów w rozwiązaniu nie może przekraczać liczby obiektów pomniejszonej o jeden. Jeśli określone jest tylko jedno źródło, wszystkie modele stanowią ekwiwalent modelu typu IDENTITY; z tego powodu analiza domyślnie odwołuje się do modelu typu IDENTITY.

Uzyskiwanie wielowymiarowego skalowania preferencji

1. Z menu wybierz:

Analiza > Skalowanie > Wielowymiarowe skalowanie preferencji (PREFSCAL)...

2. Wybierz co najmniej dwie zmienne, które identyfikują kolumny w prostokątnej macierzy odległości. Każda zmienna reprezentuje osobny obiekt kolumnowy.
3. Opcjonalnie można wybrać liczbę zmiennych definiujących wagi równą liczbie zmiennych obiektów kolumnowych. Kolejność zmiennych definiujących wagi powinna być zgodna z kolejnością obiektów kolumnowych, których one dotyczą.
4. Opcjonalnie można też wybrać zmienną wierszową. Wartości (lub etykiety wartości) tej zmiennej są używane jako etykiety obiektów wierszowych w wynikach.

5. Jeśli istnieje wiele źródeł, opcjonalnie wybierz zmienną źródłową. Liczba obserwacji w każdym pliku danych powinna być równa iloczynowi liczby obiektów wierszowych i liczby źródeł.

Oprócz tego możliwe jest określenie modelu wielowymiarowego skalowania preferencji, nałożenie ograniczeń na przestrzeń wspólną, wskazanie kryteriów zbieżności, określenie początkowej konfiguracji, a także wybranie wykresów i wyników.

Definiowanie modelu wielowymiarowego skalowania preferencji

Okno dialogowe Model umożliwia określenie modelu skalowania, jego minimalnej i maksymalnej liczby wymiarów, struktury macierzy odległości, przekształcenia, które ma być stosowane wobec bliskości, oraz informacji o tym, czy bliskości mają być przekształcane warunkowo według wiersza, zależnie od źródła, czy niezależnie od źródła.

Model skalowania. Wybierz jedną z następujących opcji:

- **Tożsamość.** Wszystkie źródła mają tę samą konfigurację.
- **Ważony euklidesowy.** Jest to model indywidualnych różnic. Każde źródło ma osobną przestrzeń, w której każdy wymiar przestrzeni wspólnej jest ważony w inny sposób.
- **Uogólniony euklidesowy.** Jest to model indywidualnych różnic. Każde źródło ma indywidualną przestrzeń uzyskaną w wyniku obrotu wspólnej przestrzeni i następującego po nim ważenia różnic wymiarów.

Odległości. Określ, czy macierz bliskości zawiera miary podobieństwa czy niepodobieństwa.

Wymiary. Domyślnie rozwiązanie jest obliczane w dwóch wymiarach (Minimum = 2, Maksimum = 2). Można wybrać minimalną i maksymalną liczbę całkowitą z zakresu od 1 do liczby obiektów minus 1, o ile tylko minimum nie będzie większe od maksimum. Procedura oblicza rozwiązanie w maksymalnej liczbie wymiarów, a następnie krokowo zmniejsza liczbę wymiarów aż do osiągnięcia najniższej wymiarowości.

Przekształcenia odległości. Wybierz jedną z następujących opcji:

- **Brak.** Bliskości nie są przekształcane. Opcjonalnie można wybrać opcję **Uwzględnij wyraz wolny** – wówczas możliwe będzie przesunięcie bliskości o stałą.
- **Liniowy.** Przekształcone bliskości są proporcjonalne do pierwotnych bliskości, to znaczy, że funkcja transformacji szacuje nachylenie, a wyraz wolny jest ustalony na wartość 0. Jest to również nazywane transformacją ilorazową. Opcjonalnie można wybrać opcję **Uwzględnij wyraz wolny** – wówczas możliwe będzie także przesunięcie bliskości o stałą. Jest to tak zwane przekształcenie przedziałowe.
- **Krzywa sklejana.** Przekształcenie bliskości jest gładkim, niemalejącym przekształceniem kawałkami wielomianowym bliskości pierwotnych. Można określić stopień wielomianu oraz liczbę węzłów wewnętrznych. Opcjonalnie można wybrać opcję **Uwzględnij wyraz wolny** – wówczas możliwe będzie także przesunięcie bliskości o stałą.
- **Wygładź.** Przekształcone bliskości zachowują pierwotną kolejność, z uwzględnieniem ograniczenia co do różnic między kolejnymi wartościami. Wynikiem jest „gładkie przekształcenie porządkowe”. Można określić, czy powiązane bliskości powinny pozostać powiązane, czy też można je rozdzielić.
- **Porządkowe.** Przekształcone bliskości zachowują kolejność pierwotnych bliskości. Można określić, czy powiązane bliskości powinny pozostać powiązane, czy też można je rozdzielić.

Zastosuj przekształcenia. Określ, czy porównywane mają być tylko bliskości w obrębie każdego wiersza, czy tylko bliskości w obrębie każdego źródła, czy też porównania mają być niezależne od wiersza lub źródła – to znaczy, czy przekształcenia mają być wykonywane dla każdego wiersza, dla każdego źródła, czy dla wszystkich bliskości naraz.

Ograniczenia wielowymiarowego skalowania preferencji

Okno dialogowe Ograniczenia umożliwia nakładanie ograniczeń na przestrzeń wspólną.

Ograniczenia przestrzeni wspólnej. Istnieje możliwość zakotwiczenia (ustalenia) współrzędnych obiektów wierszowych i/lub kolumnowych w przestrzeni wspólnej.

Zmienne ograniczające wiersza/kolumny. Wybierz plik zawierający ograniczenia i wybierz zmienne, które definiują ograniczenia w przestrzeni wspólnej. Pierwsza wybrana zmienna zawiera współrzędne obiektów w pierwszym wymiarze, druga zmienna odpowiada współrzędnym w drugim wymiarze itd. Brak wartości oznacza, że współrzędna w wymiarze jest swobodna. Liczba wybranych zmiennych musi być równa maksymalnej liczbie żądanych wymiarów. Liczba obserwacji dla każdej zmiennej musi być równa liczbie obiektów.

Wielowymiarowe skalowanie preferencji: Opcje

Okno dialogowe Opcje umożliwia wybranie początkowego stylu konfiguracji, określenie kryteriów iteracji i zbieżności oraz określenie składnika kary za obciążenie.

Konfiguracja wstępna. Wybierz jedną z następujących możliwości:

- **Klasyczna.** Prostokątna macierz odległości jest używana jako uzupełnienie wartości wewnątrzblokowych (wartości między wierszami i kolumnami) kompletnej symetrycznej macierzy MDS. Po utworzeniu kompletnej macierzy jako konfiguracja początkowa przyjmowana jest klasyczna forma skalowania. Wartości wewnątrzblokowe można zapisać poprzez podstawienie metodą nierówności trójkątów lub odległości Spearmana.
- **Ross-Cliff.** W metodzie Ross-Cliff początkowe wartości obiektów wierszowych i kolumnowych są wynikami dekompozycji podwójnie wyśrodkowanej i podniesionej do kwadratu macierzy odległości na pojedyncze wartości będące początkowymi wartościami obiektów wierszy i kolumn.
- **Korespondencja.** W metodzie korespondencji początkowymi wartościami są wyniki analizy korespondencji danych odwróconych (podobieństwa zamiast różnic) z symetryczną normalizacją ocen wierszy i kolumn.
- **Środki ciężkości.** Procedura najpierw ustawia obiekty wierszowe konfiguracji na podstawie dekompozycji wartości własnych. Następnie obiekty kolumnowe są umieszczane w środkach ciężkości określonych wyborów. Jako liczbę wyborów określ liczbę dodatnią z przedziału od 1 do liczby zmiennych bliskości.
- **Wielokrotne starty losowe.** Rozwiązania są obliczane dla kilku konfiguracji początkowych wybranych losowo, a jako najlepsze rozwiązanie wskazywana jest konfiguracja z najmniejszą karą za obciążenie.
- **Użytkownika.** Użytkownik może wybrać zmienne, które zawierają współrzędne jego własnej konfiguracji początkowej. Liczba wybranych zmiennych powinna być równa określonej maksymalnej liczbie wymiarów, przy czym pierwsza zmienna odpowiada współrzędnym w wymiarze 1, druga zmienna odpowiada współrzędnym w wymiarze 2 itd. Liczba obserwacji w każdej zmiennej powinna być równa łącznej liczbie obiektów wierszowych i kolumnowych. Współrzędne wierszowe i kolumnowe powinny być uporządkowane w stos, przy czym współrzędne kolumn powinny następować po współrzędnych wierszy.

Kryteria iteracji. Określ wartości kryteriów iteracji.

- **Zbieżność STRESS.** Algorytm przerwie iteracje, gdy względna różnica kolejnych wartości ukaranego obciążenia jest mniejsza niż określona tutaj liczba, która musi być nieujemna.
- **Minimalna wartość STRESS.** Algorytm zostanie zatrzymany, gdy ukarane obciążenie spadnie poniżej liczby określonej w tym polu. Liczba ta musi być nieujemna.
- **Maksymalna liczba iteracji.** Algorytm wykona liczbę iteracji określoną w tym polu, chyba że wcześniej zostanie spełnione jedno z powyższych kryteriów.

Składnik kary. Algorytm usiłuje zminimalizować ukarane obciążenie, miarę dobroci dopasowania równą iloczynowi Stress-I Kruskala oraz składnika kary na podstawie współczynnika zmienności przekształconych bliskości. Te elementy sterujące umożliwiają ustawienie siły i zakresu składnika kary.

- **Siła.** Im mniejsza jest wartość parametru siły, tym większa jest kara. Należy określić wartość z zakresu od 0,0 do 1,0.
- **Przedział.** Ten parametr określa moment, w którym kara staje się aktywna. Jeśli ma wartość 0,0, kara jest nieaktywna. Zwiększenie wartości powoduje, że algorytm szuka rozwiązania o większej zmienności wśród przekształconych bliskości. Podaj liczbę nieujemną.

Wielowymiarowe skalowanie preferencji: Wykresy

Okno dialogowe Wykresy umożliwia wybranie wykresów, które będą tworzone.

Wykresy. Dostępne są następujące rodzaje wykresów:

- **Wielokrotne starty.** Przedstawia zestawiony histogram ukaranego obciążenia, z widocznym zarówno obciążeniem, jak i karą.
- **Wstępna przestrzeń wspólna.** Wyświetla macierz rozrzutu współrzędnych wstępnej przestrzeni wspólnej.
- **Wartość STRESS dla wymiarów.** Generuje wykres liniowy ukaranego obciążenia wobec liczby wymiarów. Wykres ten jest tworzony tylko wtedy, gdy maksymalna liczba wymiarów jest większa niż minimalna liczba wymiarów.
- **Ostateczna przestrzeń wspólna.** Prezentowana jest macierz rozrzutu współrzędnych przestrzeni wspólnej.
- **Wagi przestrzeni.** Tworzony jest wykres rozrzutu poszczególnych wag przestrzeni. Jest to możliwe tylko wtedy, gdy w oknie dialogowym Model określono jeden z modeli różnic indywidualnych. W przypadku ważonego modelu euklidesowego na wykresie prezentowane są wagi wszystkich źródeł — z jednym wymiarem na każdej osi. W przypadku uogólnionego modelu euklidesowego dla każdego wymiaru tworzony jest jeden wykres, który prezentuje zarówno obrót, jak i wagę tego wymiaru dla każdego źródła.
- **Indywidualne przestrzenie.** Wyświetlana jest macierz rozrzutu współrzędnych dla poszczególnych przestrzeni każdego źródła. Jest to możliwe tylko wtedy, gdy w oknie dialogowym Model określono jeden z modeli różnic indywidualnych.
- **Wykresy przekształceń.** Generowany jest wykres rozrzutu pierwotnych bliskości wobec bliskości przekształconych. W zależności od sposobu przekształcania do każdego wiersza lub źródła przypisywany jest osobny kolor. Przekształcenie niezależne ma jeden wspólny kolor.
- **Wykresy Sheparda.** Pierwotne bliskości względem zarówno przekształconych bliskości, jak i odległości. Odległości są prezentowane jako punkty, a przekształcone bliskości są prezentowane jako linia. W zależności od sposobu przekształcania dla każdego wiersza lub źródła generowana jest osobna linia. Przekształcenie niezależne generuje jedną linię
- **Wykres rozrzutu dopasowania.** Zostanie wyświetlony wykres rozrzutu przekształconych bliskości wobec odległości. W przypadku określenia wielu źródeł do każdego z nich przypisywany jest osobny kolor.
- **Wykresy reszt.** Wyświetlany jest wykres rozrzutu przekształconych bliskości wobec reszt (przekształconych bliskości minus odległości). W przypadku określenia wielu źródeł do każdego z nich przypisywany jest osobny kolor.

Style obiektów wierszowych. Zapewniają dodatkową kontrolę nad wyświetlaniem obiektów wierszowych na wykresach. Wartości opcjonalnej zmiennej kolorów są cyklicznie wybierane jako kolory. Wartości opcjonalnej zmiennej znaczników są cyklicznie wybierane jako znaczniki.

Wykresy źródeł. W przypadku opcji **Przestrzenie indywidualne**, **Rozrzut dopasowania** i **Wykresy reszt** — oraz, jeśli przekształcenia odbywają się według źródła, dla opcji **Wykresy przekształceń** i **Wykresy Sheparda** — można określić źródła, dla których wykresy powinny być tworzone. Wprowadzone numery źródeł muszą być wartościami zmiennej źródłowej określonej w głównym oknie dialogowym i należeć do przedziału od 1 do liczby źródeł.

Wykresy wierszy. Jeśli przekształcenia odbywają się według wierszy, w przypadku wykresów **Wykresy przekształceń** i **Wykresy Sheparda** można określić wiersz, dla którego wykresy powinny być tworzone. Wprowadzane numery wierszy muszą mieścić się w przedziale od 1 do liczby wierszy.

Wielowymiarowe skalowanie preferencji: Wynik

Okno dialogowe Wyniki umożliwia wybór rodzajów wyświetlanych wyników i zapisywanie niektórych z nich w oddzielnych plikach.

Pokaż. Wybierz jedną lub więcej z następujących opcji do wyświetlenia:

- **Dane wejściowe.** Uwzględnia pierwotne bliskości oraz, o ile są obecne, wagi danych, początkową konfigurację i zakotwiczone (stałe) współrzędne.
- **Wielokrotne starty.** Wyświetla wartość startową generatora liczb losowych i wartość ukaranego obciążenia dla każdego losowego startu.
- **Wstępne dane.** Wyświetla współrzędne początkowej przestrzeni wspólnej.
- **Przebieg iteracji.** Wyświetla przebieg iteracji algorytmu głównego.
- **Miary dopasowania.** Wyświetla różne miary. Tabela zawiera kilka miar dobroci dopasowania, złego dopasowania, korelacji, zmienności i braku degeneracji.
- **Dekompozycja STRESS.** Wyświetla dekompozycję obiektów, wierszy i źródeł ukaranego obciążenia, w tym średnie oraz odchylenia standardowe wierszy, kolumn i źródeł.
- **Przekształcone odległości.** Wyświetla przekształcone odległości.
- **Ostateczna przestrzeń wspólna.** Przedstawia współrzędne przestrzeni wspólnej.
- **Wagi przestrzeni.** Wyświetla wagi poszczególnych przestrzeni. Ta opcja jest dostępna tylko wtedy, gdy wybrano jeden z modeli indywidualnych różnic. W zależności od modelu wagi przestrzeni są dekomponowane na wagi obrotu i wagi wymiarów, które są także wyświetlane.
- **Indywidualne przestrzenie.** Wyświetlane są współrzędne poszczególnych przestrzeni. Ta opcja jest dostępna tylko wtedy, gdy wybrano jeden z modeli indywidualnych różnic.
- **Dopasowane odległości.** Przedstawia odległości między obiektami w konfiguracji.

Zapisz w nowym pliku. Istnieje możliwość zapisania współrzędnych przestrzeni wspólnej, wag poszczególnych przestrzeni, odległości i przekształconych bliskości w odrębnych plikach danych programu IBM SPSS Statistics.

Dodatkowe właściwości komendy PREFSCAL

Istnieje możliwość zmodyfikowania wielowymiarowego skalowania preferencji poprzez wklejenie zaznaczenia do okna komend i edycję uzyskanej komendy PREFSCAL. Język składni komend umożliwia również:

- Określenie wielu list źródłowych dla indywidualnych przestrzeni, wykresów rozrzutu dopasowania i wykresów reszt oraz, w przypadku warunkowych transformacji macierzy, dla wykresów przekształceń i wykresów Sheparda, gdy dostępne jest więcej niż jedno źródło (za pomocą opcji komendy PLOT).
- Określenie wielu list wierszy dla wykresów przekształceń i wykresów Sheparda w przypadku przekształceń warunkowych wiersza (za pomocą opcji komendy PLOT).
- Określenie liczby wierszy zamiast zmiennej identyfikującej wiersze (za pomocą opcji komendy INPUT).
- Określenie liczby źródeł zamiast zmiennej identyfikującej źródła (za pomocą opcji komendy INPUT).

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

Uwagi

Niniejsza publikacja została przygotowana z myślą o produktach i usługach oferowanych w Stanach Zjednoczonych. IBM może udostępniać ten materiał w innych językach. Jednakże w celu uzyskania dostępu do takiego materiału istnieje konieczność posiadania egzemplarza produktu w takim języku.

Produktów, usług lub opcji opisywanych w tym dokumencie IBM nie musi oferować we wszystkich krajach. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Odwołanie do produktu, programu lub usługi IBM nie oznacza, że można użyć wyłącznie tego produktu, programu lub usługi IBM. Zamiast nich można zastosować ich odpowiednik funkcjonalny pod warunkiem że nie narusza to praw własności intelektualnej IBM. Jednakże cała odpowiedzialność za ocenę przydatności i sprawdzenie działania produktu, programu lub usługi pochodzących od producenta innego niż IBM spoczywa na użytkowniku.

IBM może posiadać patenty lub złożone wnioski patentowe na produkty, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przesyłać na adres:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Zapytania dotyczące zestawów znaków dwubajtowych (DBCS) należy kierować do lokalnych działów własności intelektualnej IBM (IBM Intellectual Property Department) lub wysłać je na piśmie na adres:

Intellectual Property Licensing

*Legal and Intellectual Property Law
IBM Japan Ltd.
19-21, Nihonbashi-Hakozakicho, Chuo-ku
Tokio 103-8510, Japonia*

INTERNATIONAL BUSINESS MACHINES CORPORATION DOSTARCZA TĘ PUBLIKACJĘ W STANIE, W JAKIM SIĘ ZNAJDUJE ("AS IS") BEZ UDZIELANIA JAKICHKOLWIEK GWARANCJI (W TYM TAKŻE RĘKOJMI), WYRAŻNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA NIE NARUSZA PRAW STRON TRZECICH. Ustawodawstwa niektórych krajów nie dopuszczają zastrzeżeń dotyczących gwarancji wyraźnych lub domniemanych w odniesieniu do pewnych transakcji; w takiej sytuacji powyższe zdanie nie ma zastosowania.

Informacje zawarte w tej publikacji mogą zawierać nieścisłości techniczne lub błędy drukarskie. Informacje te są okresowo aktualizowane, a zmiany te zostaną uwzględnione w kolejnych wydaniach tej publikacji. IBM zastrzega sobie prawo do wprowadzania ulepszeń i/lub zmian w produktach i/lub programach opisanych w tej publikacji w dowolnym czasie, bez wcześniejszego powiadomienia.

Wszelkie wzmianki w tej publikacji na temat stron internetowych firm innych niż IBM zostały wprowadzone wyłącznie dla wygody użytkowników i w żadnym razie nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

IBM ma prawo do używania i rozpowszechniania informacji przystanych przez użytkownika w dowolny sposób, jaki uzna za właściwy, bez żadnych zobowiązań wobec ich autora.

Licencjodawcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) umożliwienia wymiany informacji między niezależnie utworzonymi programami i innymi programami

(łącznie z opisywanym) oraz (ii) wykorzystywania wymienianych informacji, powinni skontaktować się z:

IBM Director of Licensing

IBM Corporation

*North Castle Drive, MD-NC119
Armonk, NY 10504-1785 U.S.A.*

Informacje takie mogą być udostępnione, o ile spełnione zostaną odpowiednie warunki, w tym, w niektórych przypadkach, zostanie uiszczona stosowna opłata.

Licencjonowany program opisany w niniejszej publikacji oraz wszystkie inne licencjonowane materiały dostępne dla tego programu są dostarczane przez IBM na warunkach określonych w Umowie IBM z Klientem, Międzynarodowej Umowie Licencyjnej IBM na Program lub w innych podobnych umowach zawartych między IBM i użytkownikami.

Dane dotyczące wydajności i cytowane przykłady zostały przedstawione jedynie w celu zobrazowania sytuacji. Faktyczne wyniki dotyczące wydajności mogą się różnić w zależności do konkretnych warunków konfiguracyjnych i operacyjnych.

Informacje dotyczące produktów innych podmiotów niż IBM zostały uzyskane od dostawców tych produktów, z ich publicznych ogłoszeń lub innych dostępnych publicznie źródeł. IBM nie testował tych produktów i nie może potwierdzić dokładności pomiarów wydajności, kompatybilności ani żadnych innych danych związanych z produktami firm innych niż IBM. Pytania dotyczące możliwości produktów firm innych niż IBM należy kierować do dostawców tych produktów.

Wszelkie stwierdzenia dotyczące przyszłych kierunków rozwoju i zamierzeń IBM mogą zostać zmienione lub wycofane bez powiadomienia.

Publikacja ta zawiera przykładowe dane i raporty używane w codziennych operacjach działalności gospodarczej. W celu kompleksowego zilustrowania tej działalności podane przykłady zawierają nazwy osób, firm i ich produktów. Wszystkie te nazwy/nazwiska są fikcyjne i jakiegokolwiek podobieństwo do istniejących nazw/nazwisk jest całkowicie przypadkowe.

LICENCJA W ZAKRESIE PRAW AUTORSKICH:

Niniejsza publikacja zawiera przykładowe aplikacje w kodzie źródłowym ilustrujące techniki programowania w różnych systemach operacyjnych. Użytkownik może kopiować, modyfikować i rozpowszechniać te programy przykładowe w dowolnej formie bez uiszczania opłat na rzecz IBM, w celu rozbudowy, użytkowania, handlowego lub w celu rozpowszechniania aplikacji zgodnych z aplikacyjnym interfejsem programowym dla tego systemu operacyjnego, dla którego napisane były programy przykładowe. Programy przykładowe nie zostały gruntownie przetestowane. IBM nie może zatem gwarantować ani sugerować niezawodności, użyteczności i funkcjonalności tych programów. Programy przykładowe są dostarczane w stanie, w jakim się znajdują ("AS IS"), bez jakichkolwiek gwarancji (rękojmię również wyłącza się). IBM nie ponosi odpowiedzialności za jakiegokolwiek szkody wynikające z używania programów przykładowych.

Każda kopia programu przykładowego lub jakiegokolwiek jego fragment, jak też jakiegokolwiek prace pochodne muszą zawierać następujące uwagi dotyczące praw autorskich:

© Copyright IBM Corp. 2021. Fragmenty tego kodu pochodzą z przykładowych programów produktu IBM Corp. Programy przykładowe.

© Copyright IBM Corp. 1989-2021. Wszelkie prawa zastrzeżone.

Znaki towarowe

IBM, logo IBM i ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp., zarejestrowanymi w wielu systemach prawnych na całym świecie. Pozostałe nazwy produktów i usług mogą być znakami towarowymi IBM lub innych przedsiębiorstw. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW, w sekcji "Copyright and trademark

information" (Informacje o prawach autorskich i znakach towarowych), pod adresem www.ibm.com/legal/copytrade.shtml.

Adobe, logo Adobe, PostScript oraz logo PostScript są znakami towarowymi lub zastrzeżonymi znakami towarowymi Adobe Systems Incorporated w Stanach Zjednoczonych i/lub w innych krajach.

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium i Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych w Stanach Zjednoczonych i w innych krajach.

Linux jest zastrzeżonym znakiem towarowym Linusa Torvaldsa w Stanach Zjednoczonych i/lub w innych krajach.

Microsoft, Windows, Windows NT oraz logo Windows są znakami towarowymi Microsoft Corporation w Stanach Zjednoczonych i/lub w innych krajach.

UNIX jest zastrzeżonym znakiem towarowym Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące Java są znakami towarowymi firmy i jej firm zależnych.

Indeks

A

- algorytm bootstrap
 - Analiza głównych składowych dla danych jakościowych [21](#)
- Analiza głównych składowych dla danych jakościowych
 - algorytm bootstrap [21](#)
 - dotądowe właściwości komendy [22](#)
 - zapisywanie zmiennych [20](#)
- Analiza głównych składowych dla danych jakościowych (CATPCA)
 - poziom skalowania optymalnego [16](#)
- Analiza korespondencji
 - dotądowe właściwości komendy [29](#)
 - statystyki [25](#)
 - wykresy [25](#)
- ANOVA
 - w regresji dla danych jakościowych [14](#)

B

- bezwładność
 - w analizie korespondencji [28](#)
- biplot (łączny wykres rozrzutu)
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)
 - w analizie korespondencji [28](#)
 - w wielowymiarowej analizie korespondencji [33](#)
- braki danych
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [17](#)
 - w wielowymiarowej analizie korespondencji [31](#)
- brakujące wartości
 - w regresji dla danych jakościowych [12](#)

D

- dopasowanie
 - w nieliniowej analizie korelacji kanonicznych [24](#)
- dyskretyzacja
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [17](#)
 - w regresji dla danych jakościowych [12](#)
 - w wielowymiarowej analizie korespondencji [30](#)

E

- elastyczna sieć
 - w regresji dla danych jakościowych [13](#)

K

- konfiguracja wstępna
 - w nieliniowej analizie korelacji kanonicznych [24](#)
 - w regresji dla danych jakościowych [12](#)
 - w skalowaniu wielowymiarowym [38](#)

- konfiguracja wstępna (*kontynuacja*)
 - w wielowymiarowym skalowaniu preferencji [42](#)
- korelacje
 - w skalowaniu wielowymiarowym [39](#)
- kryteria iteracji
 - w skalowaniu wielowymiarowym [38](#)
 - w wielowymiarowym skalowaniu preferencji [42](#)
- kwantyfikacje kategorii
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
 - w nieliniowej analizie korelacji kanonicznych [24](#)
 - w regresji dla danych jakościowych [14](#)
 - w wielowymiarowej analizie korespondencji [32](#)

L

- lasso
 - w regresji dla danych jakościowych [13](#)

Ł

- ładunki składowych
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
 - w nieliniowej analizie korelacji kanonicznych [24](#)
- łączne wykresy kategorii
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)
 - w wielowymiarowej analizie korespondencji [33](#)

M

- macierz korelacji
 - w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
 - w wielowymiarowej analizie korespondencji [32](#)
- miary dyskryminacji
 - w wielowymiarowej analizie korespondencji [32](#)
- miary obciążenia
 - w skalowaniu wielowymiarowym [39](#)
 - w wielowymiarowym skalowaniu preferencji [43](#)
- miary odległości
 - w analizie korespondencji [27](#)
- model skalowania
 - w wielowymiarowym skalowaniu preferencji [41](#)
- model tożsamości
 - w wielowymiarowym skalowaniu preferencji [41](#)

N

- najszybsze algorytmy
 - w skalowaniu wielowymiarowym [38](#)
- Nieliniowa analiza korelacji kanonicznych
 - dotądowe właściwości komendy [25](#)
 - statystyki [22](#)
 - wykresy [22](#)

normalizacja
w analizie korespondencji [27](#)

O

obiekty pasywne
w regresji dla danych jakościowych [12](#)
oceny obiektów
w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
w nieliniowej analizie korelacji kanonicznych [24](#)
w wielowymiarowej analizie korespondencji [32](#)
odległości
w skalowaniu wielowymiarowym [39](#)
w wielowymiarowym skalowaniu preferencji [43](#)
ograniczenia
w skalowaniu wielowymiarowym [37](#)
ograniczenia przestrzeni wspólnej
w wielowymiarowym skalowaniu preferencji [41](#)

P

poziom skalowania optymalnego
w analizie głównych składowych dla danych jakościowych (CATPCA) [16](#)
w wielowymiarowej analizie korespondencji [30](#)
PREFSCAL [40](#)
przebieg iteracji
w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
w skalowaniu wielowymiarowym [39](#)
w wielowymiarowej analizie korespondencji [32](#)
w wielowymiarowym skalowaniu preferencji [43](#)
przekształcenia bliskości
w wielowymiarowym skalowaniu preferencji [41](#)
przekształcone bliskości
w skalowaniu wielowymiarowym [39](#)
w wielowymiarowym skalowaniu preferencji [43](#)
przekształcone zmienne niezależne
w skalowaniu wielowymiarowym [39](#)

R

Regresja dla danych jakościowych
dodatkowe właściwości komendy [15](#)
poziom skalowania optymalnego [11](#)
regularyzacja [13](#)
statystyki [10](#)
wykresy [10](#)
zapisywanie [14](#)
regresja grzbietowa
w regresji dla danych jakościowych [13](#)

S

Skalowanie wielowymiarowe
dane wyjściowe [39](#)
dodatkowe właściwości komendy [40](#)
model [36](#)
ograniczenia [37](#)
opcje [38](#)
statystyki [34](#)
wykresy [34](#), [38](#), [39](#)

składnik kary
w wielowymiarowym skalowaniu preferencji [42](#)
standaryzacja
w analizie korespondencji [27](#)
statystyki opisowe
w regresji dla danych jakościowych [14](#)
statystyki ufności
w analizie korespondencji [28](#)

Ś

środki ciężkości
w nieliniowej analizie korelacji kanonicznych [24](#)

T

triplot
w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)

U

uogólniony model euklidesowy
w wielowymiarowym skalowaniu preferencji [41](#)

W

waga zmiennej
w analizie głównych składowych dla danych jakościowych (CATPCA) [16](#)
w wielowymiarowej analizie korespondencji [30](#)
wagi
w nieliniowej analizie korelacji kanonicznych [24](#)
wagi indywidualnych przestrzeni
w skalowaniu wielowymiarowym [39](#)
w wielowymiarowym skalowaniu preferencji [43](#)
wariancja wyjaśniona
w analizie głównych składowych dla danych jakościowych (CATPCA) [19](#)
ważony model euklidesowy
w wielowymiarowym skalowaniu preferencji [41](#)
wielokrotne R
w regresji dla danych jakościowych [14](#)
Wielowymiarowa analiza korespondencji
dodatkowe właściwości komendy [33](#)
poziom skalowania optymalnego [30](#)
zapisywanie zmiennych [33](#)
Wielowymiarowe skalowanie preferencji (PREFSCAL)
dodatkowe właściwości komendy [44](#)
model [41](#)
ograniczenia przestrzeni wspólnej [41](#)
opcje [42](#)
statystyki [40](#)
wykresy [40](#), [43](#)
wyniki [43](#)
współczynniki regresji
w regresji dla danych jakościowych [14](#)
współrzędne indywidualnych przestrzeni
w wielowymiarowym skalowaniu preferencji [43](#)
współrzędne przestrzeni wspólnej
w skalowaniu wielowymiarowym [39](#)
w wielowymiarowym skalowaniu preferencji [43](#)
wykres rozrzutu dopasowania

wykres rozrzutu dopasowania (*kontynuacja*)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy
 w analizie korespondencji [28](#)
 w nieliniowej analizie korelacji kanonicznych [24](#)
 w regresji dla danych jakościowych [15](#)
 w skalowaniu wielowymiarowym [38, 39](#)

wykresy indywidualnych przestrzeni
 w skalowaniu wielowymiarowym [38](#)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy kategorii
 w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)
 w wielowymiarowej analizie korespondencji [33](#)

wykresy korelacji
 w skalowaniu wielowymiarowym [38](#)

wykresy ładunków czynnikowych
 w analizie głównych składowych dla danych jakościowych (CATPCA) [21](#)

wykresy miar różnicujących
 w wielowymiarowej analizie korespondencji [33](#)

wykresy obciążenia
 w skalowaniu wielowymiarowym [38](#)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy ostatecznej przestrzeni wspólnej
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy przekształceń
 w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)
 w skalowaniu wielowymiarowym [38](#)
 w wielowymiarowej analizie korespondencji [33](#)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy przestrzeni wspólnej
 w skalowaniu wielowymiarowym [38](#)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy punktów obiektów
 w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)
 w wielowymiarowej analizie korespondencji [33](#)

wykresy reszt
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy rzutowanych środków ciężkości
 w analizie głównych składowych dla danych jakościowych (CATPCA) [20](#)

Wykresy Sheparda
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy wag indywidualnych przestrzeni
 w skalowaniu wielowymiarowym [38](#)
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy wag przestrzeni
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy wielokrotnych startów
 w wielowymiarowym skalowaniu preferencji [43](#)

wykresy wstępnej przestrzeni wspólnej
 w wielowymiarowym skalowaniu preferencji [43](#)

wymiary
 w analizie korespondencji [27](#)

