

IBM SPSS Custom Tables 26



Uwaga

Przed skorzystaniem z niniejszych informacji oraz produktu, którego one dotyczą, należy zapoznać się z informacjami zamieszczonymi w sekcji “Informacje” na stronie 19.

Informacje o produkcie

Niniejsze wydanie publikacji dotyczy wersji 26, wydania 0, modyfikacji 0 produktu IBM® SPSS Statistics oraz wszystkich następnych wersji i modyfikacji do czasu, aż w kolejnym wydaniu publikacji zostanie zawarta informacja o stosownej zmianie.

Spis treści

Tabele użytkownika	1	Informacje	19
Interfejs modułu Custom Tables	1	Znaki towarowe	21
Interfejs konstruktora tabel	1		
Tworzenie tabel	1	Indeks	23
Tabele użytkownika: Testy	7		
Pliki przykładowe	9		

Tabele użytkownika

Następujące funkcje dotyczące tabel użytkownika są zawarte w SPSS Statistics Standard Edition lub komponencie opcjonalnym Custom Tables.

Interfejs modułu Custom Tables

Interfejs konstruktora tabel

Moduł Custom Tables wykorzystuje prosty interfejs konstruktora tabel, obsługujący funkcję „przeciągnij i upuść”, umożliwiającą podgląd tabeli przy wybieraniu zmiennych i opcji. Udostępnia on również funkcje nie występujące w typowym „oknie dialogowym”, w tym możliwość zmiany rozmiaru okna i jego poszczególnych paneli.

Tworzenie tabel

Zmienne i miary podsumowujące, które będą występować w tabelach, są wybierane w interfejsie modułu Custom Tables.

Analiza > Tabele > Tabele użytkownika

Lista zmiennych. Zmienne zawarte w pliku danych są wyświetlane w lewym panelu okna dialogowego. Moduł Custom Tables stosuje dwa różne poziomy pomiarów zmiennych i postępuje z nimi w sposób zależny od poziomu pomiaru danej zmiennej:

Jakościowe. Dane posiadające ograniczoną liczbę odrębnych wartości lub kategorii (np. płeć czy religia). Zmienne kategoryjne mogą być zmiennymi łańcuchowymi (alfanumerycznymi) lub numerycznymi, wykorzystującymi kody liczbowe reprezentujące kategorie (np. 0 = *mężczyzna* i 1 = *kobieta*). Zmienne kategoryjne nazywane są również danymi jakościowymi. Zmienne kategoryjne mogą być **nominalne** lub **porządkowe**.

- *Nominalny.* Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rangowania; na przykład wydział, na którym są zatrudnieni pracownicy. Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy.* Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorię z wewnętrznym rangowaniem, na przykład poziomy zadowolenia z usługi od bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.

Zmienne kategoryjne definiują kategorie (w wierszach, kolumnach i warstwach) tabeli, a domyślną statystyką podsumowującą jest liczebność (liczba obserwacji w każdej kategorii). Na przykład domyślna tabela zmiennej jakościowej Płeć pokazuje po prostu liczbę kobiet i mężczyzn.

Ilościowe. Dane mierzone na skali interwałowej lub ilorazowej, których wartości określają zarówno ich porządek, jak i odległość między nimi. Na przykład roczna pensja w wysokości 72 195 PLN jest wyżej niż pensja wynosząca 52 398 PLN, a odległość między tymi dwiema wartościami wynosi 19 797 PLN. Zmienne ilościowe są również zwane danymi ilościowymi lub ciągłymi.

Zmienne ilościowe są zwykle podsumowywane w obrębie kategorii zmiennych kategoryjnych, a domyślną statystyką podsumowującą jest średnia. Na przykład domyślna tabela zmiennej Dochód w obrębie kategorii płci pokazuje średni dochód kobiet i mężczyzn.

Podsumowania mogą również obejmować same zmienne ilościowe, bez ich grupowania według zmiennej kategoryjnej. Jest to wykorzystywane głównie do **zestawiania** podsumowań wielu zmiennych ilościowych.

Zestawy wielokrotnych odpowiedzi

Moduł Custom Tables obsługuje również specjalny rodzaj zmiennej o nazwie **zestaw wielokrotnych odpowiedzi**. Zestawy wielokrotnych odpowiedzi nie są „zmiennymi” w normalnym tego słowa znaczeniu. Nie występują w Edytorze danych, a inne procedury nie rozpoznają ich. Zestawy wielokrotnych odpowiedzi wykorzystują wiele zmiennych do rejestracji odpowiedzi na pytania w przypadku, gdy respondent może udzielić więcej niż jednej odpowiedzi. Są one traktowane podobnie jak zmienne kategoryjne i można je, w większości przypadków, poddawać podobnym operacjom.

Ikona obok każdej zmiennej na liście zmiennych określa jej rodzaj.

Kategorie. Po wybraniu z listy zmiennych zmiennej kategoryjnej kategorii zdefiniowanej dla niej są wyświetlane na panelu Informacje o zmiennych. Po zastosowaniu zmiennej w tabeli są one również wyświetlane w panelu obszaru roboczego. Jeśli dla zmiennej nie zdefiniowano żadnych kategorii, na panelu Informacje o zmiennych i w panelu obszaru roboczego wyświetlane są dwie kategorie zastępcze: *Kategoria 1* i *Kategoria 2*.

Zdefiniowane kategorie wyświetlane w konstruktorze tabel odpowiadają **etykiatom wartości**, czyli opisowym etykiatom przypisanym różnym wartościom danych (np. wartości liczbowe 0 i 1, z etykietami *mężczyzna* i *kobieta*). Etykiety wartości można zdefiniować w panelu Informacje o zmiennych w edytorze danych.

Panel obszaru roboczego. Tabelę tworzy się przeciągając zmienne do wierszy i kolumn w panelu obszaru roboczego. Pokazywany jest tam podgląd tabeli, która zostanie utworzona. W panelu obszaru roboczego nie są pokazywane rzeczywiste wartości danych w poszczególnych komórkach, lecz dosyć wiernie przedstawia on układ finalnej tabeli. W przypadku zmiennych kategoryjnych, jeśli plik danych zawiera unikalne wartości, dla których nie zdefiniowano etykiet, rzeczywista tabela może zawierać większą liczbę kategorii niż pokazywanych jest na podglądzie.

Podstawowe zasady i ograniczenia przy tworzeniu tabel

- W przypadku zmiennych kategoryjnych statystyki podsumowujące są obliczane na podstawie najbardziej wewnętrznej zmiennej w wymiarze statystyki źródłowej.
- Domyślny wymiar statystyki źródłowej (wiersz lub kolumna) w przypadku zmiennych kategoryjnych zależy od kolejności przeciągania zmiennych do panelu obszaru roboczego. Jeśli najpierw, na przykład zmienna zostanie przeciągnięta do pola wierszy, wymiar wierszowy będzie domyślnym wymiarem statystyki źródłowej.
- Zmienne ilościowe mogą być podsumowywane tylko w obrębie kategorii najbardziej wewnętrznej zmiennej w wymiarze wierszy lub kolumn. (Zmienną ilościową można umieścić na dowolnym poziomie tabeli, lecz jest ona podsumowywana na najbardziej wewnętrznym poziomie).
- Zmiennych ilościowych nie można podsumowywać w obrębie innych zmiennych ilościowych. Można jednak zestawiać podsumowania wielu zmiennych ilościowych lub podsumowywać je w obrębie kategorii zmiennych kategoryjnych. Nie można zagnieżdżać jednej zmiennej ilościowej w innej ani umieszczać jednej zmiennej ilościowej w wymiarze wierszowym, a innej w wymiarze kolumnowym.
- Jeśli dowolna zmienna w aktywnym zbiorze danych zawiera więcej niż 12 000 zdefiniowanych etykiet wartości, nie można użyć kreatora tabel do tworzenia tabel. Jeśli zmienne przekraczające to ograniczenie nie muszą być uwzględniane w tabelach, można zdefiniować i zastosować zbiory danych, które wyłączają te zmienne. Jeśli konieczne jest uwzględnienie zmiennych zawierających ponad 12 000 zdefiniowanych etykiet wartości, do utworzenia tabel można użyć składni komend **CTABLES**.

Tworzenie tabeli

1. Z menu wybierz kolejno następujące pozycje:
Analiza > Tabele > Tabele użytkownika
2. Przeciągnij jedną lub kilka zmiennych do pola wierszy i/lub kolumn w panelu obszaru roboczego.
3. Kliknij przycisk **Utwórz**, aby utworzyć tabelę.

Usuwanie zmiennej z panelu obszaru roboczego

1. Zaznacz (kliknij) zmienną w panelu obszaru roboczego.
2. Kliknij prawym przyciskiem myszy i wybierz opcję **Usuń zmienną** z menu rozwijanego.

Zagnieżdżanie zmiennych

Zagnieżdżanie, podobnie jak umieszczanie w tabelach krzyżowych, umożliwia pokazanie zależności pomiędzy dwiema zmiennymi kategoryjnymi, z tą różnicą, że jedna zmienna jest zagnieżdżona w drugiej w tym samym wymiarze. Można na przykład zagnieżdżyć zmienną *Płeć* w zmiennej *Kategoria wieku* w wymiarze wierszowym, pokazując liczbę kobiet i mężczyzn w każdej kategorii wiekowej.

Można również zagnieżdżyć zmienną ilościową w zmiennej kategoryjnej. Na przykład zagnieżdżenie zmiennej *Dochód* w zmiennej *Płeć* spowoduje wyświetlenie odrębnej średniej (lub mediany czy innej miary podsumowującej) wartości dochodów kobiet i mężczyzn.

Zagnieżdżanie zmiennych

1. Przeciągnij zmienną kategoryjną do pola wierszy lub kolumn w panelu obszaru roboczego.
2. Przeciągnij zmienną kategoryjną lub ilościową na zmienną wiersza lub kolumny kategoryjnej.
3. Wybierz z menu opcję **Zagnieżdżaj nad wszystkimi zmiennymi**, **Zagnieżdżaj po lewej** albo **Zagnieżdżaj po prawej**.

Tabela 1. Zagnieżdżone zmienne kategoryjne

Zmienna 1	Zmienna 2	Statystyka podsumowująca
Kategoria 1	Kategoria 1	12
	Kategoria 2	34
	Kategoria 3	56
Kategoria 2	Kategoria 1	12
	Kategoria 2	34
	Kategoria 3	56

Uwaga: Moduł Custom Tables nie uwzględnia przetwarzania plików powstałych w wyniku podzielenia danych wg warstw. Aby uzyskać taki sam wynik, jak dla plików powstałych w wyniku podzielenia danych wg warstw, umieść zmienne podziału danych w warstwach na skrajnym zewnętrznym poziomie zagnieżdżenia w tabeli.

Edytuj Statystyki

Panel Edytuj statystyki umożliwia:

- Dodawanie i usuwanie statystyk podsumowujących do/z tabeli.

To, które statystyki (i inne opcje) będą dostępne w panelu Edytuj statystyki, zależy od poziomu pomiaru zmiennej źródłowej. Źródło statystyk (zmienna, na podstawie której obliczane są statystyki) jest określone przez:

- **Poziom pomiaru.** Jeśli tabela (lub sekcja tabeli w przypadku tabeli zestawionej) zawiera zmienną ilościową, statystyki obliczane są na podstawie tej zmiennej.
- **Kolejność wybierania zmiennych.** Domyślny wymiar statystyki źródłowej (wiersz lub kolumna) w przypadku zmiennych kategoryjnych zależy od kolejności przeciągania zmiennych do panelu obszaru roboczego. Jeśli najpierw, na przykład zmienna zostanie przeciągnięta do pola wierszy, wymiar wierszowy będzie domyślnym wymiarem statystyki źródłowej.
- **Zagnieżdżanie.** W przypadku zmiennych kategoryjnych statystyki są obliczane na podstawie najbardziej wewnętrznej zmiennej w wymiarze statystyki źródłowej.

Statystyki podsumowujące dla zmiennych kategoryjnych: Do podstawowych statystyk dostępnych dla zmiennych kategoryjnych należą liczebności i procenty. Dla podsumowań ogółem i sum pośrednich można również określić

statystyki użytkownika. Takie statystyki użytkownika obejmują miary tendencji centralnej (np. średnia i mediana) oraz rozproszenia (np. odchylenie standardowe), które mogą być odpowiednie dla niektórych porządkowych zmiennych kategoryalnych.

Liczebności. Liczba obserwacji we wszystkich komórkach tabeli lub liczba odpowiedzi w przypadku zestawów wielokrotnych odpowiedzi. Jeśli stosowane jest ważenie, ta wartość jest ważoną liczebnością.

- Jeśli stosowane jest ważenie, wartość jest ważoną liczebnością.
- Liczebność ważona jest taka sama zarówno przy globalnym ważeniu zbioru danych (**Dane > Ważenie obserwacji...**),

Nieważona liczebność. Nieważona liczba obserwacji we wszystkich komórkach tabeli. Różni się od liczebności tylko, jeśli obowiązuje ważenie.

Skorygowana liczebność. Skorygowana liczebność jest używana przy obliczeniach z ważeniem według podstawy rzeczywistej. Jeśli nie jest używana zmienna będąca rzeczywistą podstawą ważenia, to skorygowana liczebność jest taka sama, jak nieskorygowana.

Procent w wierszu. Procent w każdym wierszu. Suma procentów w każdym wierszu podtabeli (w przypadku procentów prostych) wynosi 100%. Procenty w wierszu są zwykle przydatne tylko wówczas, kiedy w *kolumnie* występuje zmienna jakościowa.

Procent w kolumnie. Procent w każdej kolumnie. Suma procentów w każdej kolumnie podtabeli (w przypadku procentów prostych) wynosi 100%. Procenty w kolumnie są zwykle przydatne tylko wówczas, kiedy w *wierszu* występuje zmienna jakościowa.

Procent w podtabeli. Procenty dla każdej komórki są obliczane na podstawie liczby obserwacji w podtabeli. Wszystkie odsetki komórek w podtabeli obliczane są na podstawie tej samej całkowitej liczby obserwacji, a ich suma w podtabeli wynosi 100%. W tabelach zagnieżdżonych zmienna poprzedzająca najwyższy poziom zagnieżdżenia definiuje podtabelę. Na przykład w tabeli określającej *stan cywilny* w podziale na *pleć* w podziale na *kategorie wieku*, zmienna *Pleć* definiuje podtabelę.

Procent w tabeli. Procenty dla każdej komórki obliczane są na podstawie liczby obserwacji w całej tabeli. Wszystkie odsetki komórek obliczane są na podstawie tej samej całkowitej liczby obserwacji, a ich suma w całej tabeli wynosi 100% (w przypadku procentów prostych).

Przedziały ufności

- Dostępna jest górna i dolna granica ufności liczebności, wartości procentowych, średniej, mediany, centyli i sumy.
- Łącuch "&[Poziom ufności]" w etykiecie tabeli zawiera poziom ufności.
- Dostępny jest błąd standardowy liczebności, wartości procentowych, średniej i sumy.
- Przedziały ufności i błędy standardowe nie są dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Poziom 1

Poziom ufności przedziałów ufności wyrażony jako wartość procentowa. Wartość musi być liczbą większą od 0 i mniejszą niż 100.

Zestawy wielokrotnych odpowiedzi

Zestawy wielokrotnych odpowiedzi mogą zawierać procenty wyliczone na podstawie obserwacji, odpowiedzi lub liczebności. Więcej informacji zawiera temat "Statystyki podsumowujące dla zestawów wielokrotnych odpowiedzi" na stronie 5.

Podstawa obliczania wartości procentowej: Procenty mogą być obliczane na trzy różne sposoby, w zależności od sposobu traktowania braków danych w bazie obliczeniowej:

Procent prosty. Procenty obliczane są na podstawie liczby obserwacji wykorzystanych w tabeli, a ich suma zawsze wynosi 100%. Jeśli kategoria zostanie wykluczona z tabeli, obserwacje do niej należące nie są uwzględniane w bazie do obliczeń. Obserwacje zawierające systemowe braki danych nigdy nie są uwzględniane w bazie do obliczeń. Obserwacje zawierające braki danych zdefiniowane przez użytkownika nie są uwzględniane, jeśli z tabeli zostaną wykluczone kategorie braków danych zdefiniowanych przez użytkownika (ustawienie domyślne), natomiast są uwzględniane, jeśli takie kategorie zostaną włączone do tabeli. Każdy procent niezawierający w nazwie terminu *N ważnych* lub *N Ogółem* jest procentem prostym.

Procent z N ogółem. Obserwacje zawierające systemowe i zdefiniowane przez użytkownika braki danych są uwzględniane przy obliczaniu procentów prostych. Suma procentów może być mniejsza niż 100%.

Procent z N ważnych. Obserwacje zawierające braki danych zdefiniowane przez użytkownika nie są uwzględniane przy obliczaniu procentów prostych nawet wówczas, gdy tabela zawiera kategorie braków danych zdefiniowanych przez użytkownika.

Uwaga: Obserwacje w kategoriach wykluczonych ręcznie, poza kategoriami braków danych zdefiniowanych przez użytkownika, nigdy nie są uwzględniane przy obliczeniach.

Statystyki podsumowujące dla zestawów wielokrotnych odpowiedzi: Dla zestawów wielokrotnych odpowiedzi dostępne są następujące statystyki dodatkowe.

% z odpowiedzi w kolumnie/wierszu/warstwie. Procent na podstawie odpowiedzi.

% z odpowiedzi w kolumnie/wierszu/warstwie (Baza: liczebność). Licznikiem jest liczba odpowiedzi, a mianownikiem liczebność ogółem.

% z liczebności w kolumnie/wierszu/warstwie (Baza: odpowiedzi). Licznikiem jest liczebność, a mianownikiem całkowita liczba odpowiedzi.

Kolumnowy/Wierszowy % z odpowiedzi w warstwie %. Procent w podtabelach. Procent na podstawie odpowiedzi.

Kolumnowy/Wierszowy % z odpowiedzi w warstwie (Baza: liczebność). Procenty w podtabelach. Licznikiem jest liczba odpowiedzi, a mianownikiem liczebność ogółem.

Kolumnowy/Wierszowy % z odpowiedzi w warstwie (Baza: odpowiedzi). Procenty w podtabelach. Licznikiem jest liczebność, a mianownikiem całkowita liczba odpowiedzi.

Odpowiedzi. Liczebność odpowiedzi.

% z odpowiedzi w podtabeli/tabeli. Procent na podstawie odpowiedzi.

% z odpowiedzi w podtabeli/tabeli (Baza: liczebność). Licznikiem jest liczba odpowiedzi, a mianownikiem liczebność ogółem.

% z liczebności w podtabeli/tabeli (Baza: odpowiedzi). Licznikiem jest liczebność, a mianownikiem całkowita liczba odpowiedzi.

Statystyki podsumowujące dla zmiennych ilościowych i podsumowania użytkownika dla zmiennych jakościowych: Oprócz liczebności i procentów dostępnych dla zmiennych kategoryjnych, dla zmiennych ilościowych dostępne są następujące statystyki podsumowujące oraz jako statystyki podsumowania wybierane przez użytkownika i sumy pośrednie dla zmiennych kategoryjnych. Statystyki takie nie są dostępne dla zestawów wielokrotnych odpowiedzi ani zmiennych łańcuchowych (alfanumerycznych).

Średnia. Średnia arytmetyczna; suma podzielona przez liczbę obserwacji.

Mediana. Wartość, powyżej i poniżej której przypada połowa obserwacji; 50. percentyl.

Dominanta. Wartość najczęściej występująca. Jeżeli występuje wiązanie, pokazywana jest wartość najmniejsza.

Minimum. Najmniejsza (najniższa) wartość.

Maksimum. Największa (najwyższa) wartość.

Braki danych. Liczba braków danych (zdefiniowanych i systemowych).

Percentyl. Można uwzględnić 5., 25., 75., 95. i/lub 99. percentyl.

Przedział. Różnica pomiędzy wartością maksymalną a minimalną.

Odchylenie standardowe. Miara rozproszenia wokół średniej. W przypadku rozkładu normalnego, 68% obserwacji znajduje się w obszarze oddalonym o jedno odchylenie standardowe od średniej, zaś 95% - w przedziale oddalonym o dwa odchylenia standardowe. Na przykład jeśli średni wiek wynosi 45 lat, przy odchyleniu standardowym równym 10, w przypadku rozkładu normalnego 95% obserwacji mieściłoby się w przedziale od 25 do 65 lat (pierwiastek kwadratowy z wariancji).

Suma. Suma wartości.

Procent sumy. Procenty na podstawie sum. Statystyka dostępna dla wierszy i kolumn (w podtabelach), całych wierszy i kolumn (dla wszystkich podtabel), warstw, podtabel i całych tabel.

Ogółem N. Liczba wartości bez braków danych, ze zdefiniowanymi brakami danych i systemowymi brakami danych. Nie obejmuje obserwacji w kategoriach wykluczonych ręcznie innych niż kategorie braków danych zdefiniowanych przez użytkownika.

Skorygowane - N Ogółem. Skorygowana suma N używana przy obliczeniach z ważeniem według podstawy rzeczywistej. Jeśli nie jest używana zmienna będąca rzeczywistą podstawą ważenia (określona na karcie Opcje), to skorygowana suma N liczebność jest taka sama, jak nieskorygowana. Ta statystyka jest niedostępna w przypadku zestawów wielokrotnych odpowiedzi.

N ważnych. Liczba wartości bez braków danych. Nie obejmuje obserwacji w kategoriach wykluczonych ręcznie innych niż kategorie braków danych zdefiniowanych przez użytkownika.

Skorygowane - N ważnych. Skorygowana liczba N ważnych jest stosowana w obliczeniach z ważeniem według podstawy rzeczywistej. Jeśli nie jest używana zmienna będąca rzeczywistą podstawą ważenia (określona na karcie Opcje), to skorygowana liczba N ważnych jest taka sama, jak nieskorygowana. Ta statystyka jest niedostępna w przypadku zestawów wielokrotnych odpowiedzi.

Wariancja. Miara rozproszenia wokół średniej, równa sumie podniesionych do kwadratu odchyżeń od średniej, podzielonej przez liczbę obserwacji minus jeden. Wariancja mierzona jest w jednostkach będących kwadratem jednostek samej zmiennej (kwadrat odchylenia standardowego).

Przedziały ufności

- Dostępna jest górna i dolna granica ufności liczebności, wartości procentowych, średniej, mediany, centyli i sumy.
- Łańcuch "&[Poziom ufności]" w etykiecie tabeli zawiera poziom ufności.
- Dostępny jest błąd standardowy liczebności, wartości procentowych, średniej i sumy.
- Przedziały ufności i błędy standardowe nie są dostępne w przypadku zestawów wielokrotnych odpowiedzi.

Poziom I

Poziom ufności przedziałów ufności wyrażony jako wartość procentowa. Wartość musi być liczbą większą od 0 i mniejszą niż 100.

Zestawione tabele

Każda sekcja tabeli zdefiniowana przez zmienną zestawiającą jest traktowana jako odrębna tabela i w ten sposób obliczane są statystyki podsumowujące.

Kategorie i podsumowania

Moduł Custom Tables umożliwia:

- Zmianę kolejności kategorii.
- Wstawianie podsumowań.
- W przypadku zmiennej bez zdefiniowanych etykiet wartości można tylko sortować kategorie i wstawiać podsumowania.

Uzyskiwanie dostępu do opcji kategorii i podsumowań

1. Przeciągnij zmienną kategoryjną lub zestaw wielokrotnych odpowiedzi do panelu obszaru roboczego.
2. Kliknij prawym przyciskiem myszy zmienną w panelu obszaru roboczego i z menu kontekstowego wybierz jedną z opcji kategorii lub podsumowań.

Aby posortować kategorie

1. Kliknij prawym przyciskiem myszy zmienną w panelu obszaru roboczego i z menu kontekstowego wybierz opcję **Sortuj kategorie**, a następnie wybierz metodę sortowania:
 - Według wartości
 - Według etykiet
 - Według liczebności
 - Według dolnej granicy

Podsumowania

1. Kliknij prawym przyciskiem myszy zmienną w panelu obszaru roboczego i z menu kontekstowego wybierz opcję **Pokaż sumę**, a następnie wybierz metodę wyświetlania podsumowania:
 - Powyżej kategorii
 - Poniżej kategorii

Jeśli zaznaczona zmienna jest zagnieżdżona w innej zmiennej, podsumowania zostaną wstawione dla każdej podtabeli.

Tabele użytkownika: Testy

Funkcja Testy udostępni testy istotności dotyczące tabel użytkownika.



Testy te nie są dostępne dla tabel, w których etykiety kategorii zostały przeniesione poza swój domyślny wymiar w tabeli lub do przeliczonych kategorii.

Testy średnich kolumnowych i proporcji kolumnowych

Testy średnich kolumnowych są dostępne w przypadku zmiennych ilościowych. Testy proporcji kolumnowych są dostępne w przypadku zmiennych jakościowych.

Porównaj średnie kolumn

Przeprowadzane parami testy równości średnich kolumnowych. Tabela musi zawierać jedną zmienną jakościową w kolumnach, a jako najbardziej wewnętrzny poziom wierszy musi zawierać zmienną ilościową. Tabela musi zawierać średnią jako statystykę podsumowania.

W przypadku zwykłych zmiennych jakościowych wariancję można oszacować na podstawie wszystkich kategorii lub tylko kategorii, które są porównywane. W przypadku zmiennych wielokrotnych odpowiedzi wariancja w teście średnich zawsze oparta jest wyłącznie na kategoriach, które są porównywane.

Porównaj proporcje kolumn

Testy równości proporcji kolumnowych. Tabela musi zawierać co najmniej jedną zmienną jakościową zarówno w kolumnach, jak i wierszach. Tabela musi obejmować liczebności lub wartości procentowe kolumn.

Poziom istotności

Poziom istotności dla testów średnich kolumnowych i proporcji kolumnowych.

- Wartość musi być liczbą większą od 0 i mniejszą od 1.
- W przypadku określenia dwóch poziomów istotności wartości istotności mniejsze lub równe mniejszemu poziomowi są oznaczane wielkimi literami. Wartości istotności mniejsze lub równe większemu poziomowi są oznaczane małymi literami.
- W razie wybrania opcji **Użyj indeksów typu APA** druga wartość jest ignorowana.

Skoryguj wartości p dla wielokrotnych porównań

Korekta metodą **Bonferroniego** koryguje wskaźnik błędu w rodzinie (FWER). Metoda **Benjaminiego-Hochberga** to korekta wskaźnika fałszywych wykryć (FDR). Metoda ta jest mniej konserwatywna niż korekta Bonferroniego.

Identyfikacja istotnych różnic

W testach średnich kolumnowych i proporcji kolumnowych można uwidocznić istotne wyniki w oddzielnej tabeli lub w głównej tabeli.

Przedstaw w oddzielnej tabeli

Wyniki testu istotności są wyświetlane w oddzielnej tabeli. Jeżeli dwie wartości znacząco się różnią, komórka zawierająca większą wartość wyświetla klucz określający kolumnę mniejszej wartości.

Pokaż wartości istotności

Wartości istotności są wyświetlane w nawiasach po każdej wartości klucza w komórce. Ta opcja jest dostępna tylko wtedy, gdy istotne wyniki są wyświetlane w oddzielnej tabeli.

Przedstaw w tabeli głównej

Wyniki testu istotności są wyświetlane w tabeli głównej. Każda kategoria kolumn w tabeli jest oznaczona kluczem alfabetycznym. Dla każdej istotnej pary klucz kategorii z mniejszą średnią lub proporcją kolumn uwidoczniony jest w kategorii z większą średnią lub proporcją kolumn.

- Zatrzymanie wskaźnika myszy nad kluczem w komórce etykiety kolumny wewnątrz tabeli przestawnej powoduje podświetlenie wszystkich komórek tabeli z tym kluczem istotności. W przypadku tabeli z więcej niż jedną zmienną w wymiarze kolumn podświetlane są tylko komórki w tej podtabeli.
- Aby zaznaczyć wszystkie komórki w tabeli (lub podtabeli) mające ten sam klucz istotności, prawym przyciskiem myszy kliknij komórkę z etykieta kolumny i wybierz kolejno opcje **Wybierz > Zaznacz wszystkie komórki z tym kluczem istotności**.

Indeksy dolne w stylu APA

Istnieje możliwość oznaczania istotnych różnic w stylu APA, przy użyciu indeksów dolnych. Jeśli dwie wartości są istotnie różne, to będą miały różne litery w indeksach dolnych. Te indeksy dolne nie są odsyłaczami do przypisów. Gdy działa niniejsza opcja, styl zdefiniowanego przypisu w istniejącej TableLook jest nadpisywany i przypisy są wyświetlane jako liczby z indeksem górnym. Aby zaznaczyć wszystkie komórki znajdujące się w tym samym wierszu i mające ten sam klucz istotności, kliknij prawym przyciskiem myszy komórkę z kluczem istotności i wybierz opcję **Zaznacz komórki o podobnym znaczeniu**

Testy niezależności (chi-kwadrat)

Powoduje wykonanie testu niezależności chi-kwadrat dla tabel, w których zarówno w wierszach, jak i kolumnach występuje przynajmniej jedna zmienna jakościowa.

W miejsce kategorii z sumami pośrednimi użyj sum pośrednich

Wszystkie sumy pośrednie zastępują kategorie testów istotności. W innym wypadku tylko sumy pośrednie, dla których kategorie z sumami pośrednimi zostały ukryte, zastępują kategorie do testowania.

Uwzględnij zmienne wielokrotnych odpowiedzi

Powoduje, że kategorie zestawów wielokrotnych odpowiedzi są uwzględnione w testach istotności. W przeciwnym razie kategorie zestawów wielokrotnych odpowiedzi są uwzględnione w testach istotności.

Pliki przykładowe

Przykładowe pliki zainstalowane wraz z produktem można znaleźć w podkatalogu *Samples* w katalogu instalacyjnym. W podkatalogu *Samples* znajdują się osobne foldery dla każdego z następujących języków: angielskiego, francuskiego, niemieckiego, włoskiego, japońskiego, koreańskiego, polskiego, rosyjskiego, chińskiego uproszczonego, hiszpańskiego i chińskiego tradycyjnego.

Nie wszystkie przykładowe pliki są dostępne we wszystkich językach. Jeżeli przykładowy plik nie jest dostępny w danym języku, ten folder językowy zawiera angielską wersję przykładowego pliku.

Opisy

Poniżej znajdują się krótkie opisy przykładowych plików wykorzystywanych w różnych przykładach zamieszczonych w dokumentacji.

- **accidents.sav.** Jest to plik danych hipotetycznych dotyczący przedsiębiorstwa ubezpieczeniowego, które analizuje związane z wiekiem i płcią czynniki ryzyka wypadków samochodowych w danym regionie. Każda obserwacja odpowiada klasyfikacji krzyżowej kategorii wieku i płci.
- **adl.sav.** Jest to plik danych hipotetycznych dotyczący starań służących określeniu korzyści płynących z proponowanego typu terapii dla pacjentów po udarze. Lekarze losowo przypisali pacjentów po udarze do jednej z dwóch grup. W pierwszej grupie przeprowadzono standardową fizjoterapię, w drugiej zaś także dodatkową terapię psychologiczną. Trzy miesiące po leczeniu zdolności każdego pacjenta do wykonywania codziennych czynności zostały ocenione jako zmienne porządkowe.
- **advert.sav.** Jest to plik danych hipotetycznych dotyczący badań firmy handlowej nad relacją między nakładami finansowymi na reklamę a wynikami sprzedaży. W tym celu zgromadzono wyniki sprzedaży z przeszłości i powiązane z nimi koszty reklamy.
- **aflatoxin.sav.** Jest to plik danych hipotetycznych dotyczący badań nasion kukurydzy pod kątem obecności aflatoksyny, trucizny, której stężenie różni się znacznie między plonami zbóż i w ich obrębie. Zakład przetwórstwa zbóż otrzymał 16 próbek z każdego z 8 plonów i zmierzył poziom aflatoksyny w cząsteczkach na miliard (PPB).
- **anorectic.sav.** Pracując nad ustandaryzowaną symptomatologią zachowań anorektycznych i bulimicznych,¹ przebadali 55 nastolatków ze stwierdzonymi zaburzeniami odżywiania. Każdy pacjent był badany cztery razy w ciągu czterech lat, co dało łącznie 220 obserwacji. W każdej obserwacji u pacjentów oceniano każdy z 16 symptomów. Brakuje ocen symptomów pacjenta 71 przy 2. badaniu, pacjenta 76 przy 2. badaniu i pacjenta 47 przy 3. badaniu, co daje 217 ważnych obserwacji.
- **anticonvulsants.sav.** Badacze pracujący w dziedzinach medycznych mogą używać uogólnionego liniowego modelu mieszanego w celu ustalenia, czy nowy lek przeciwdrgawkowy może zmniejszać częstotliwość ataków padaczkowych u pacjentów. Powtarzalne pomiary dla tego samego pacjenta są zwykle skorelowane pozytywnie, dlatego odpowiedni powinien być model mieszany z pewnymi efektami losowymi. Zmienna przewidywana — liczba ataków — przyjmuje dodatnie wartości całkowite, dlatego odpowiedni może być uogólniony liniowy model mieszany o rozkładzie Poissona i logarytmiczna funkcja łączenia.

1. Van der Ham, T., J. J. Meulman, D. C. Van Strien i H. Van Engeland. 1997. Empirically based subgrouping of eating disorders in adolescents: A longitudinal perspective. *British Journal of Psychiatry*, 170, 363-368.

- **bankloan.sav.** Ten plik danych hipotetycznych dotyczy dążeń banku do zmniejszenia wskaźnika niespłaconych kredytów. Plik zawiera informacje finansowe i demograficzne dotyczące 850 przeszłych i obecnych klientów. Pierwsze 700 obserwacji to klienci, którym przyznano już kredyty. Ostatnie 150 obserwacji to przyszli, potencjalni klienci, których bank musi sklasyfikować jako mających dobrą lub złą sytuację kredytową.
- **bankloan_binning.sav.** Jest to plik danych hipotetycznych, zawierający informacje finansowe i demograficzne o 5000 byłych klientów.
- **bankloan_cs.sav.** Jest to plik danych hipotetycznych dotyczących dążeń banku do identyfikacji cech, które są charakterystyczne dla osób, które nie spłacają pożyczek, a następnie do wykorzystania tych cech w celu identyfikacji dobrego i złego ryzyka kredytowego.
- **bankloan_cs_noweights.sav.** Jest to plik danych hipotetycznych dotyczących dążeń banku do identyfikacji cech, które są charakterystyczne dla osób, które nie spłacają pożyczek, a następnie do wykorzystania tych cech w celu identyfikacji dobrego i złego ryzyka kredytowego. Wagi prób nie są uwzględnione w pliku.
- **behavior.sav.** W klasycznym przykładzie ² 52 studentów poproszono o ocenę kombinacji 15 sytuacji i 15 zachowań za pomocą 10-punktowej skali, od 0 = „wyjątkowo właściwe” do 9 = „wyjątkowo niewłaściwe”. Wartości zostały uśrednione i zarejestrowane jako niepodobieństwa.
- **behavior_ini.sav.** Ten plik danych zawiera konfigurację początkową dwuwymiarowego rozwiązania zagadnienia z pliku *behavior.sav*.
- **brakes.sav.** Jest to plik danych hipotetycznych, który dotyczy kontroli jakości w zakładzie produkującym hamulce tarczowe dla samochodów wysokiej klasy. Plik danych zawiera wyniki pomiarów średnicy 16 tarcz hamulcowych pochodzących z każdej z 8 maszyn produkcyjnych. Docelowa średnica tarcz wynosi 322 milimetry.
- **breakfast.sav.** W klasycznym badaniu ³ 21 studentów MBA z Wharton School i ich małżonków poproszono o uszeregowanie 15 artykułów śniadaniowych według preferencji: od 1 = „najbardziej preferowany” do 15 = „najmniej preferowany”. Preferencje zarejestrowano w ramach sześciu różnych scenariuszy, od „Preferencji ogólnej” do „Przekąska, tylko z napojem”.
- **breakfast-overall.sav.** Ten plik danych zawiera preferencje dotyczące artykułów śniadaniowych tylko dla pierwszego scenariusza, „Preferencja ogólna”.
- **broadband_1.sav.** Jest to plik danych hipotetycznych, zawierający liczbę abonentów (według regionu) krajowej stacji telewizyjnej. Plik danych zawiera miesięczne liczby abonentów z 85 regionów w okresie czterech lat.
- **broadband_2.sav.** Ten plik danych jest taki sam jak plik *broadband_1.sav*, ale zawiera dane za trzy dodatkowe miesiące.
- **cable_survey.sav.** Dostawcy telewizji kablowej, telefonii przewodowej i internetu przewodowego chcą uzyskać dodatkowe informacje na temat potencjalnych klientów. Przeprowadzają ankietę, w której uczestniczy 2000 osób z regionów świadczenia usług — ankietę zawiera następujące pytania: (1) czy nie korzystają z usługi; (2) są odbiorcami usługi u innych dostawców albo (3) czy korzystają z oferty firmy w każdej z tych trzech usług. Ankietę dodatkowo gromadzi pewne dane demograficzne, takie jak płeć, kategoria wiekowa (4 poziomy), kategoria wykształcenia (3 poziomy), kategoria dochodów (3 poziomy), kategoria typu mieszkania (4 poziomy), liczba lat zamieszkania w bieżącej kategorii mieszkania (3 poziomy), liczba osób w gospodarstwie domowym itp.
- **car_insurance_claims.sav.** Zbiór danych przedstawiony i przeanalizowany w innej pracy ⁴ dotyczy roszczeń o odszkodowania motoryzacyjne. Średnią kwotę roszczenia można zamodelować jako zmienną o rozkładzie gamma, używając odwróconej funkcji łączenia w celu powiązania średniej zmiennej zależnej z liniową kombinacją wieku ubezpieczonego, typu pojazdu i jego wieku. Liczba wniesionych roszczeń może być wykorzystana jako waga skalująca.
- **car_sales.sav.** Ten plik danych zawiera hipotetyczne oszacowania sprzedaży, ceny i specyfikacje fizyczne różnych marek i modeli pojazdów. Ceny i specyfikacje fizyczne uzyskano naprzemiennie z witryny *edmunds.com* i z witryn producentów.
- **car_sales_uprepared.sav.** Jest to zmodyfikowana wersja pliku *car_sales.sav*, która nie obejmuje żadnych przetransformowanych wersji tych pól.

2. Price, R. H., i D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, 579-586.

3. Green, P. E., i V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.

4. McCullagh, P., i J. A. Nelder. 1989. *Generalized Linear Models*, wydanie drugie. London: Chapman & Hall.

- **carpet.sav.** W popularnym przykładzie⁵ spółka zainteresowana wprowadzeniem na rynek nowego środka do czyszczenia dywanów chce zbadać, jaki wpływ na preferencje klientów ma pięć czynników: wzornictwo opakowania, marka, cena, znak *Good Housekeeping* i gwarancja zwrotu pieniędzy. Istnieją trzy poziomych czynników dotyczące konstrukcji opakowań (różniących się położeniem aplikatora), trzy marki (*K2R*, *Glory* i *Bissell*), trzy poziomy cen i dwa poziomy (tak lub nie) dla każdego z dwóch ostatnich czynników. Dziesięciu klientów ocenia 22 profile zdefiniowane przez te czynniki. Zmienna *Preferencja* zawiera rangę średnich ocen każdego profilu. Niskie rangi odpowiadają wysokiej preferencji. Zmienna ta odzwierciedla ogólną miarę preferencji każdego z profili.
- **carpet_prefs.sav.** Ten plik danych bazuje na przykładzie omówionym w opisie pliku *carpet.sav*, ale zawiera rzeczywiste rangowanie uzyskane od każdego z 10 klientów. Klientów poproszono o uszeregowanie 22 profili produktów, od najbardziej do najmniej preferowanego. Zmienne od *PREF1* do *PREF22* zawierają identyfikatory powiązanych profili, zgodnie z definicją w pliku *carpet_plan.sav*.
- **catalog.sav.** Ten plik danych zawiera hipotetyczne miesięczne wielkości sprzedaży trzech produktów oferowanych przez firmę wydającą katalogi. Uwzględniono także dane pięciu możliwych predyktorów.
- **catalog_seasfac.sav.** Ten plik danych jest taki sam jak plik *catalog.sav*, ale dodano w nim zestaw czynników sezonowych obliczonych za pomocą procedury Dekompozycja sezonowa oraz towarzyszące im zmienne daty.
- **cellular.sav.** Jest to plik danych hipotetycznych dotyczący dążeń operatora telefonii komórkowej do zmniejszenia poziomu odejścia klientów. Oceny tendencji do odejścia są zastosowane do klientów porangowanych od 0 do 100. Klienci z oceną równą 50 lub wyższą mogą rozważać zmianę dostawcy usług.
- **ceramics.sav.** Jest to plik danych hipotetycznych dotyczący dążeń producenta do określenia, czy nowy stop wysokiej klasy ma większą odporność termiczną niż stop standardowy. Każda obserwacja odpowiada odrębnemu testowi jednego ze stopów i zawiera wartość temperatury, przy której łożysko uległo awarii.
- **cereal.sav.** Jest to plik danych hipotetycznych, który dotyczy badania 880 osób w zakresie ich preferencji śniadaniowych, rejestrującego także ich wiek, płeć, stan cywilny i informacje o aktywnym stylu życia (w zależności od tego, czy co najmniej dwa razy w tygodniu uprawiają sport). Każda obserwacja reprezentuje jednego respondenta.
- **clothing_defects.sav.** Jest to plik danych hipotetycznych, który dotyczy procesu kontroli jakości w fabryce odzieży. Z każdej partii produkowanej w zakładzie inspektorzy pobierają próbę odzieży i liczą, ile sztuk odzieży jest nie do przyjęcia.
- **coffee.sav.** Ten plik danych dotyczy postrzegania wizerunków sześciu marek kawy mrożonej⁶. Dla każdego z 23 atrybutów kawy mrożonej respondenci wybierali wszystkie marki, które były opisywane przez dany atrybut. Sześć marek oznaczono symbolami AA, BB, CC, DD, EE i FF, aby zachować poufność.
- **contacts.sav.** Jest to plik danych hipotetycznych, który dotyczy list kontaktowych dla grupy przedstawicieli handlowych sprzedających komputery dla firm. Każdy kontakt jest przydzielony do kategorii według działu firmy, w którym pracuje dany przedstawiciel, oraz jego pozycji w firmie. Zarejestrowano także kwotę ostatniej zawartej transakcji sprzedaży, czas, jaki upłynął od ostatniej transakcji, oraz wielkość firmy osoby kontaktowej.
- **credit_card.sav.** Hipotetyczne badanie korzystania z kart kredytowych śledzi miesięczne wydatki każdego podmiotu z podstawowej karty przez dwa lata i z podziałem na typy transakcji (artykuły spożywcze, handel detaliczny, rozrywka, podróże i inne). Każdy rekord w zbiorze danych odpowiada danemu miesiącowi wydatków i typowi transakcji, dlatego dane gromadzone dla każdego obiektu muszą obejmować 2 lata \times 12 miesięcy na rok \times 5 typów transakcji = 120 rekordów
- **creditpromo.sav.** Jest to plik danych hipotetycznych, który dotyczy dążeń domu towarowego do oszacowania skuteczności ostatniej promocji kart kredytowych. W tym celu wybrano losowo 500 posiadaczy kart. Połowa z nich otrzymała reklamę promującą obniżoną stopę procentową na zakupy dokonane w ciągu trzech następnych miesięcy. Druga połowa otrzymała standardową sezonową reklamę.
- **cross_sell.sav.** Sklep internetowy prowadzi kluby dla wielbicieli książek i dla wielbicieli płyt CD. Co miesiąc firma przygotowuje oferty specjalne, które są dostępne dla członków klubów. Firma zamierza przygotować model specjalnej oferty comiesięcznej dostępnej w zależności od łącznych zakupów książek i płyt CD, a także typ oferty dostępnej dla członków klubów. W tej sytuacji odpowiednia jest procedura podwójna MNK, ponieważ pieniądze

5. Green, P. E., i Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.

6. Kennedy, R., C. Riquier i B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement, and Analysis for Marketing*, 5, 56-70.

wydane na oferty specjalne nie są wydawane na książki ani płyty CD; z tego względu nie istnieje pętla sprzężenia zwrotnego między odpowiedzią a tymi dwoma predyktorami.

- **customer_dbase.sav.** Jest to plik danych hipotetycznych, dotyczący dążeń firmy do wykorzystania informacji znajdujących się w jej magazynie danych w celu złożenia specjalnych ofert klientom, w których przypadku prawdopodobieństwo odpowiedzi jest największe. Losowo wybrano podzbiór klientów z bazy klientów firmy, skierowano do nich oferty specjalne i zarejestrowano ich reakcje.
- **customer_information.sav.** Plik z hipotetycznymi danymi, zawierający informacje mailingowe klientów, takie jak nazwy i adresy.
- **customer_subset.sav.** Podzbiór 80 obserwacji z *customer_dbase.sav.*
- **debate.sav.** Jest to plik danych hipotetycznych, który dotyczy par odpowiedzi na ankietę przeprowadzaną wśród uczestników debaty politycznej przed debatą i po niej. Każda obserwacja odpowiada jednemu respondentowi.
- **debate_aggregate.sav.** Jest to plik danych hipotetycznych, w którym zagregowano odpowiedzi z pliku *debate.sav.* Każda obserwacja odpowiada klasyfikacji krzyżowej preferencji przed debatą i po niej.
- **demo.sav.** Jest to plik danych hipotetycznych, który dotyczy bazy danych klientów zakupionej do celów wysyłania comiesięcznych ofert. Zarejestrowano fakt, czy dany klient zareagował na ofertę, oraz różne informacje demograficzne.
- **demo_cs_1.sav.** Jest to plik danych hipotetycznych, który dotyczy pierwszego etapu dążeń pewnej firmy do skompilowania bazy danych informacji uzyskanych w ankiecie. Każda obserwacja odpowiada innemu miastu. Rejestrowane są region, prowincja, okręg i identyfikator miasta.
- **demo_cs_2.sav.** Jest to plik danych hipotetycznych, który dotyczy drugiego etapu dążeń pewnej firmy do skompilowania bazy danych informacji uzyskanych w ankiecie. Każda obserwacja odpowiada różnym jednostkom gospodarstw domowych z miast wybranych w pierwszym etapie; rejestrowane są region, prowincja, okręg, miasto, oddział i identyfikator jednostki. Uwzględniono także informacje na temat doboru próby z pierwszych dwóch etapów konstruowania bazy.
- **demo_cs.sav.** Jest to plik danych hipotetycznych, który zawiera wyniki ankiety zgromadzone przy użyciu złożonego modelu planu losowania. Każda obserwacja odpowiada innej jednostce gospodarstwa domowego; rejestrowane są różne informacje demograficzne i dotyczące doboru próby.
- **diabetes_costs.sav.** Ten plik danych hipotetycznych zawiera informacje towarzystwa ubezpieczeniowego na temat osób ubezpieczonych chorych na cukrzycę. Każda obserwacja odpowiada jednemu ubezpieczonemu.
- **dietstudy.sav.** Ten plik danych hipotetycznych zawiera wyniki badania nad tzw. dietą Stillmana ⁷. Każda obserwacja odpowiada odrębnemu badanemu i zawiera jego wagę w funtach przed dietą i po niej oraz poziom trójglicerydów we krwi w mg/100 ml.
- **dmdata.sav.** Jest to plik danych hipotetycznych, który zawiera informacje demograficzne i informacje dotyczące zakupów firmy zajmującej się marketingiem bezpośrednim. *dmdata2.sav* zawiera informacje o podzbiornie kontaktów, które otrzymały wiadomości testowe, a *dmdata3.sav* zawiera informacje o pozostałych kontaktach, które nie otrzymały wiadomości testowych.
- **dvdplayer.sav.** Jest to plik danych hipotetycznych dotyczący prac nad nowym odtwarzaczem płyt DVD. Używając prototypu urządzenia, dział marketingu zgromadził dane z grupy fokusowej. Każda obserwacja odpowiada jednemu badanemu użytkownikowi i zawiera pewne dane demograficzne o badanych oraz ich odpowiedzi na pytania dotyczące prototypu.
- **Employee data.sav.** Jest to plik danych hipotetycznych, który zawiera informacje dotyczące pracowników (wyszkolenie, kategoria zatrudnienia, bieżące wynagrodzenie, poprzednie doświadczenie itp.).
- **german_credit.sav.** Ten plik danych pochodzi ze zbioru danych „German credit”, znajdującego się w Repository of Machine Learning Databases ⁸ na Uniwersytecie Kalifornijskim w Irvine.
- **grocery_1month.sav.** Ten plik danych hipotetyczny to plik danych *grocery_coupons.sav.*, w którym tygodniowe wielkości zakupów zostały zagregowane tak, aby każda obserwacja odpowiadała jednemu klientowi. W rezultacie pewne zmienne, które zmieniały się w skali tygodnia, zniknęły, a zarejestrowana kwota wydatków jest teraz sumą wydatków w ciągu czterech tygodni badania.

7. Rickman, R., N. Mitchell, J. Dingman i J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228:, 54-58.

8. Blake, C. L. i C. J. Merz. 1998. "UCI Repository of machine learning databases." Pod adresem <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

- **grocery_coupons.sav.** Jest to plik danych hipotetycznych, który zawiera dane zebrane w ankiecie prowadzonej przez sieć sklepów spożywczych, zainteresowaną nawykami zakupowymi swoich klientów. Każdy klient jest badany przez cztery tygodnie, a każda obserwacja odpowiada odrębnemu tygodniowi tego klienta i zawiera obserwacje o tym, gdzie i kiedy dana osoba robi zakupy oraz ile wydała na artykuły spożywcze w danym tygodniu.
- **guttman.sav.** Bell ⁹ przedstawił tabelę ilustrującą możliwe grupy społeczne. Guttman ¹⁰ wykorzystał część tej tabeli, w której pięć zmiennych opisujących takie zagadnienia, jak interakcje społeczne, poczucie przynależności do grupy, fizyczna bliskość członków grupy i formalność relacji zostało skrzyżowanych z siedmioma teoretycznymi grupami społecznymi, obejmującymi tłum (np. widzów meczu piłki nożnej), publiczność (np. widzów w teatrze lub słuchaczy wykładu), publikę (np. grupę osób oglądających telewizję lub czytających gazetę), mob (podobny do tłumy, ale ze znacznie intensywniejszymi interakcjami), grupę pierwotną (intymną), grupę wtórną (dobrowolne) i nowoczesną społeczność (luźną grupę sprzymierzeńców, ukształtowaną przez bliskość fizyczną i zapotrzebowanie na specjalistyczne usługi).
- **health_funding.sav.** Jest to plik danych hipotetycznych, który zawiera dane na temat finansowania opieki zdrowotnej (kwoty na 100 członków populacji), wskaźników zachorowań (wskaźnik na 10 000 członków populacji) i wizyt w zakładach opieki zdrowotnej (wskaźnik na 10 000 członków populacji). Każda obserwacja reprezentuje inne miasto.
- **hivassay.sav.** Jest to plik danych hipotetycznych, który dotyczy dążeń laboratorium farmaceutycznego do opracowania szybkiej metody oznaczania próbek w celu wykrywania zakażenia wirusem HIV. Wyniki oznaczania to osiem ciemniejących odcieni czerwieni, gdzie ciemniejsze odcienie oznaczają większe prawdopodobieństwo zachorowania. Próba laboratoryjna została przeprowadzona na 2000 próbek krwi, z których połowa była zainfekowana wirusem HIV, druga połowa zaś wolna od wirusa.
- **hourlywagedata.sav.** Jest to plik danych hipotetycznych, który dotyczy wynagrodzeń godzinowych pielęgniarek zatrudnionych w przychodniach i szpitalach i mających różne poziomy doświadczenia.
- **insurance_claims.sav.** Jest to plik danych hipotetycznych, który dotyczy firmy ubezpieczeniowej, która chce zbudować model do flagowania podejrzanych, potencjalnie oszukańczych roszczeń. Każda obserwacja reprezentuje jedno roszczenie.
- **insure.sav.** Jest to plik danych hipotetycznych, który dotyczy firmy ubezpieczeniowej badającej czynniki ryzyka, wskazujących, czy klient będzie musiał wnieść roszczenie w ramach 10-letniej umowy ubezpieczenia na życie. Każda obserwacja w pliku danych reprezentuje parę umów, z których w jednej nastąpiło roszczenie, dopasowanych według wieku i płci.
- **judges.sav.** Jest to plik danych hipotetycznych, który dotyczy ocen przyznawanych przez wykwalifikowanych sędziów (i jednego ochotnika) 300 występom gimnastycznym. Każdy wiersz reprezentuje jeden występ; sędziowie oglądali te same występy.
- **kinship_dat.sav.** Rosenberg i Kim ¹¹ postanowili przeanalizować 15 terminów związanych z pokrewieństwem (ciotka, brat, kuzyn, córka, ojciec, wnuczka, dziadek, babcia, wnuk, matka, bratanek, bratanica, siostra, syn, wuj). Poprosili cztery grupy studentów (dwie męskie i dwie żeńskie) o posortowanie tych terminów na podstawie podobieństw. Dwie grupy (jedną męską i jedną żeńską) poproszono o posortowanie terminów dwa razy, przy czym drugie sortowanie miało być oparte na innym kryterium niż pierwsze. W ten sposób uzyskano łącznie sześć „źródeł”. Każde źródło odpowiada macierzy odległości 15 x 15, której komórki odpowiadają liczbie osób w źródle pomniejszonej o liczbę określającą, ile razy zmienne w tym źródle były łączone w podzbiory.
- **kinship_ini.sav.** Ten plik danych zawiera konfigurację początkową trójwymiarowego rozwiązania zagadnienia z pliku *kinship_dat.sav*.
- **kinship_var.sav.** Ten plik danych zawiera zmienne niezależne *gender* [płeć], *gener(ation)* [pokolenie] i *degree* (of separation) [stopień oddzielenia], które można wykorzystać do interpretacji wymiarów rozwiązania zagadnienia z pliku *kinship_dat.sav*. W szczególności można je wykorzystać do ograniczenia przestrzeni rozwiązania do liniowej kombinacji tych zmiennych.

9. Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.

10. Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, 469-506.

11. Rosenberg, S., i M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, 489-502.

- **marketvalues.sav.** Ten plik danych dotyczy sprzedaży domów na nowym osiedlu w Algonquin (USA) w latach 1999–2000. Dane o tej sprzedaży są ogólnodostępne.
- **nhis2000_subset.sav.** National Health Interview Survey (NHIS) to duże, oparte na całej populacji badanie stanu zdrowia ludności cywilnej w USA. Wywiady są przeprowadzane osobiście w reprezentatywnej w skali kraju próbie gospodarstw domowych. Dla każdego członka rodziny gromadzi się informacje i spostrzeżenia o zachowaniach dotyczących zdrowia i stanie zdrowia. Ten plik danych zawiera podzbiór informacji z badania przeprowadzonego w 2000 r. National Center for Health Statistics. National Health Interview Survey, 2000. Plik danych i dokumentacja do użytku publicznego. ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/. Dostęp uzyskano w 2003 r.
- **ozone.sav.** Dane uwzględniają 330 obserwacji sześciu zmiennych meteorologicznych, służących do przewidywania koncentracji ozonu na podstawie pozostałych zmiennych. Poprzedni badacze, m.in. ¹², ¹³, znaleźli wśród tych zmiennych nieliniowości, które utrudniają zastosowanie standardowych podejść opartych na regresji.
- **pain_medication.sav.** Ten plik danych hipotetycznych zawiera wyniki badań klinicznych leku przeciwzapalnego służącego do leczenia chronicznego bólu stawów. Przedmiotem szczególnego zainteresowania jest czas, jaki potrzebny jest do działania leku, oraz jego porównanie z istniejącymi lekami.
- **patient_los.sav.** Ten plik danych hipotetycznych zawiera dane leczenia pacjentów przyjętych do szpitala z podejrzeniem zawału mięśnia sercowego („atak serca”). Każda obserwacja odpowiada jednemu pacjentowi i zawiera wiele zmiennych związanych z jego pobytem w szpitalu.
- **patlos_sample.sav.** Ten plik danych hipotetycznych zawiera dane leczenia próby pacjentów, u których podczas leczenia zawału mięśnia sercowego („atak serca”) podawano leki trombolityczne (rozpuszczające skrzepy krwi). Każda obserwacja odpowiada jednemu pacjentowi i zawiera wiele zmiennych związanych z jego pobytem w szpitalu.
- **poll_cs.sav.** Jest to plik danych hipotetycznych, który dotyczy dążeń ankietowanych państwowych do określenia poziomu wsparcia publicznego dla ustawy przed jej przyjęciem przez zgromadzenie ustawodawcze. Obserwacje odpowiadają zarejestrowanym osobom głosującym. Każda obserwacja zawiera dane o okręgu, okręgu miejskim i osiedlu, w których dana osoba głosująca mieszka.
- **poll_cs_sample.sav.** Ten plik danych hipotetycznych zawiera próbę osób głosujących w wymienionych w pliku *poll_cs.sav*. Próbę dobrano zgodnie z wzorcem określonym w pliku *poll_csplan*, a w pliku danych zarejestrowano prawdopodobieństwo uwzględnienia i przykładowe wagi. Należy jednak pamiętać, że ponieważ plan dobierania próby wykorzystuje metodę prawdopodobieństwa proporcjonalnego do rozmiaru (PPS), istnieje także plik zawierający prawdopodobieństwa dla łącznego wyboru (*poll_jointprob.sav*). Dodatkowe zmienne odpowiadające danym demograficznym osób głosujących i ich opiniom o proponowanej ustawie zostały zebrane i dodane do pliku danych po doborze próby.
- **property_assess.sav.** Jest to plik danych hipotetycznych, które dotyczą dążeń asesora okręgu do utrzymania aktualności wycen wartości nieruchomości przy ograniczonych zasobach. Obserwacje odpowiadają nieruchomościom sprzedanym w obrębie hrabstwa w ostatnim roku. W każdej obserwacji w pliku danych zarejestrowano okręg miejski, w którym znajduje się nieruchomość, nazwisko asesora, który ostatnio odwiedzał nieruchomość, czas, jaki upłynął od tej wyceny, dokonaną wówczas wycenę wartości oraz wartość sprzedaży nieruchomości.
- **property_assess_cs.sav.** Jest to plik danych hipotetycznych, które dotyczą dążeń urzędnika stanu do utrzymania aktualności wycen wartości nieruchomości przy ograniczonych zasobach. Obserwacja odpowiada nieruchomościom w obrębie stanu. W każdej obserwacji w pliku danych zarejestrowano okręg, okręg miejski i dzielnicę, w której znajduje się nieruchomość, czas, jaki upłynął od ostatniej wyceny, oraz dokonaną wówczas wycenę wartości.
- **property_assess_cs_sample.sav.** Ten plik danych hipotetycznych zawiera próbę nieruchomości wymienionych w pliku *property_assess_cs.sav*. Próbę dobrano zgodnie z wzorcem określonym w pliku *property_assess_csplan*, a w pliku danych zarejestrowano prawdopodobieństwo uwzględnienia i przykładowe wagi. Po dobraniu próby do pliku danych dodano nowo zgromadzoną zmienną, *Current value* (Wartość bieżąca).

12. Breiman, L., i J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, 580-598.

13. Hastie, T., i R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.

- **recidivism.sav.** Jest to plik danych hipotetycznych, które dotyczą dążeń rządowej agencji organów ścigania do zrozumienia wskaźników recydywy w jej obszarze jurysdykcji. Każda obserwacja odpowiada jednej osobie, która wcześniej złamała prawo, i zawiera informacje demograficzne, pewne szczegóły pierwszego przestępstwa oraz czas, jaki upłynął do drugiego aresztowania, jeśli miało ono miejsce w ciągu dwóch lat od pierwszego aresztowania.
- **recidivism_cs_sample.sav.** Jest to plik danych hipotetycznych, które dotyczą dążeń rządowej agencji organów ścigania do zrozumienia wskaźników recydywy w jej obszarze jurysdykcji. Każda obserwacja odpowiada jednej osobie, która wcześniej złamała prawo, zwolnionej z pierwszego aresztu w czerwcu 2003 r., i zawiera jej informacje demograficzne, pewne szczegóły pierwszego przestępstwa oraz dane drugiego aresztowania, jeśli miało ono miejsce przed końcem czerwca 2006 r. Przestępców wybrano z departamentów wyznaczonych zgodnie z planem dobierania próby określonym w pliku *recidivism_cs.csplan*; ponieważ wykorzystano w nim metodę prawdopodobieństwa proporcjonalnego do rozmiaru (PPS), istnieje także plik zawierający prawdopodobieństwa dla łącznego wyboru (*recidivism_cs_jointprob.sav*).
- **rfm_transactions.sav.** Plik z danymi hipotetycznymi, który zawiera dane dotyczące transakcji zamówień, w tym datę zamówienia, elementy zamawiane i wartość pieniężną każdej transakcji.
- **salesperformance.sav.** Jest to plik danych hipotetycznych dotyczący oceny dwóch nowych kursów dla sprzedawców. Sześćdziesięciu pracowników podzielonych na trzy grupy otrzymuje szkolenie standardowe. Dodatkowo grupa 2 uczestniczy w szkoleniu technicznym, a grupa 3 w kursie praktycznym. Na zakończenie kursu przetestowano wszystkich pracowników i zarejestrowano ich oceny. Każda obserwacja w pliku danych odpowiada jednemu kursantowi i zawiera informacje o grupie, do której dany kursant został przydzielony, oraz ocenę, jaką uzyskał na egzaminie.
- **satisf.sav.** Jest to plik danych hipotetycznych, który dotyczy badania zadowolenia klientów przeprowadzonego przez firmę handlową w 4 sklepach. Przebadano łącznie 582 klientów, a każda obserwacja reprezentuje odpowiedzi jednego klienta.
- **screws.sav.** Ten plik danych zawiera informacje o cechach śrub, wkrętów, nakrętek i gwoździ ¹⁴.
- **shampoo_ph.sav.** Jest to plik danych hipotetycznych, który dotyczy procesu kontroli jakości w fabryce produktów do pielęgnacji włosów. W regularnych odstępach czasu dokonywane są pomiary sześciu odrębnych partii produktów i rejestrowane jest ich pH. Docelowy zakres wartości to 4,5–5,5.
- **ships.sav.** Zbiór danych przedstawiony i zanalizowany w innej pracy¹⁵, dotyczący uszkodzeń statków towarowych spowodowanych przez fale. Liczebność wypadków można zamodelować jako zmienną o rozkładzie Poissona, mając dany typ statku, okres jego budowy i serwisowania. Łączna liczba miesięcy pracy dla każdej komórki tabeli utworzonej przez klasyfikację krzyżową czynników stanowi wartość narażenia na ryzyko.
- **site.sav.** Jest to plik danych hipotetycznych, który dotyczy dążeń firmy do wybrania nowych obszarów rozwijania działalności. Firma zatrudniła dwóch konsultantów, aby odrębnie ocenili te lokalizacje. Oprócz rozszerzonego raportu, każdej lokalizacji przyznali oni sumaryczną ocenę, uznając ją za „dobrą”. „średnią” lub „kiepską” możliwość.
- **smokers.sav.** Ten plik danych pochodzi z przeprowadzonego w USA w 1998 r. krajowego badania gospodarstw domowych dotyczącego stosowania używek i stanowi prawdopodobną próbę amerykańskich gospodarstw domowych. (<http://dx.doi.org/10.3886/ICPSR02934>) Pierwszym krokiem analizy tego pliku danych powinno być więc ważenie danych tak, aby odzwierciedlały trendy widoczne w populacji.
- **stocks.sav** Ten plik danych hipotetycznych zawiera ceny akcji i ich ilość dla jednego roku.
- **stroke_clean.sav.** Ten plik danych hipotetycznych zawiera stan bazy danych medycznych po jej wyczyszczeniu za pomocą procedur z modułu Statistics Base Edition.
- **stroke_invalid.sav.** Ten plik danych hipotetycznych zawiera początkowy stan bazy danych medycznych, w tym kilka błędów we wprowadzaniu danych.
- **stroke_survival.** Ten plik danych hipotetycznych zawiera dane o czasie przeżycia pacjentów kończących program rehabilitacji pacjentów po udarze niedokrwiennym, stających przed różnymi wyzwaniem. Notowane są takie zdarzenia, jak udar wtórny, zawał serca, udar niedokrwienny lub krwotoczny, oraz czas ich wystąpienia. Próba jest obcięta z lewej strony, ponieważ obejmuje tylko pacjentów, którzy dożyli do końca programu rehabilitacji zaleconego po udarze.

14. Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.

15. McCullagh, P., i J. A. Nelder. 1989. *Generalized Linear Models*, wydanie drugie. London: Chapman & Hall.

- **stroke_valid.sav.** Ten plik danych hipotetycznych zawiera stan bazy danych medycznych po sprawdzeniu wartości za pomocą procedury Walidacja danych. Nadal zawiera obserwacje, które mogą być anomaliami.
- **survey_sample.sav.** Ten plik danych zawiera wyniki badania obejmujące dane demograficzne i różne miary poglądów. Bazuje on na podzbiorze zmiennych z ogólnych badań opinii społecznej, przeprowadzonych w 1998 roku przez NORC (amerykański ośrodek badania opinii społecznej), mimo to niektóre wartości danych zostały zmodyfikowane i dodatkowe, fikcyjne zmienne zostały dodane do celów demonstracyjnych.
- **tcm_kpi.sav.** Jest to plik danych hipotetycznych zawierający wartości kluczowych wskaźników wydajności (KPI) przedsiębiorstwa za poszczególne tygodnie. Zawiera także dane różnych metryk podlegających kontroli — z poszczególnych tygodni w tym samym okresie, którego dotyczą dane o wskaźnikach wydajności.
- **tcm_kpi_upd.sav.** Ten plik danych jest taki sam jak plik *tcm_kpi.sav*, ale zawiera dane za cztery dodatkowe tygodnie.
- **telco.sav.** Ten plik danych hipotetycznych dotyczy dążeń firmy telekomunikacyjnej do zmniejszenia poziomu odejścia w bazie klientów. Każda obserwacja odpowiada jednemu klientowi i zawiera różne informacje demograficzne oraz dane o wykorzystaniu usług.
- **telco_extra.sav.** Ten plik danych jest podobny do pliku danych *telco.sav*, ale usunięto z niego zmienne „tenure” (tytuł użytkownika) i dane wydatków przedstawione na przekształconej skali logarytmicznej, zastąpione ustandaryzowanymi zmiennymi wydatków klientów, przedstawionymi na przekształconej skali logarytmicznej.
- **telco_missing.sav.** Ten plik danych jest taki sam jak plik danych *telco.sav*, ale niektóre dane demograficzne zastąpiono w nim brakami danych.
- **testmarket.sav.** Ten plik danych hipotetycznych dotyczy planów sieci barów szybkiej obsługi co do wprowadzenia nowej pozycji w menu. Istnieją trzy możliwe kampanie promujące nowy produkt, nowa pozycja jest więc wprowadzana w oddziałach na kilku losowo wybranych rynkach. W każdym oddziale wykorzystuje się inną kampanię promocyjną i rejestruje się tygodniową wielkość sprzedaży nowego produktu. Każda obserwacja odpowiada jednemu tygodniowi w określonej lokalizacji.
- **testmarket_1month.sav.** Ten plik danych hipotetycznych to plik danych *testmarket.sav*, w którym tygodniowe wielkości sprzedaży zostały zagregowane tak, aby każda obserwacja odpowiadała jednej lokalizacji. W rezultacie niektóre zmienne, zmieniające się w skali tygodnia, zniknęły, a zarejestrowana wielkość sprzedaży jest teraz sumą sprzedaży w ciągu czterech tygodni badania.
- **tree_car.sav.** Jest to plik danych hipotetycznych zawierający dane demograficzne i informacje o cenie zakupu pojazdu.
- **tree_credit.sav.** Jest to plik danych hipotetycznych zawierający dane demograficzne i informacje o historii kredytowej.
- **tree_missing_data.sav** Jest to plik danych hipotetycznych zawierający dane demograficzne i informacje o historii kredytowej, z dużą liczbą braków danych.
- **tree_score_car.sav.** Jest to plik danych hipotetycznych zawierający dane demograficzne i informacje o cenie zakupu pojazdu.
- **tree_textdata.sav.** Prosty plik danych, zawierający tylko dwie zmienne, których pierwotnym przeznaczeniem było pokazanie domyślnego stanu zmiennych przed przypisaniem poziomu pomiaru oraz etykiet wartości.
- **tv-survey.sav.** Jest to plik danych hipotetycznych dotyczący badania przeprowadzonego przez studio telewizyjne, które rozważa przedłużenie emisji cieszącego się powodzeniem programu. 906 respondentów zapytano, czy oglądaliby ten program w różnych sytuacjach. Każdy wiersz reprezentuje jednego respondenta, a każda kolumna — jedną sytuację.
- **ulcer_recurrence.sav.** Ten plik zawiera częściowe informacje pochodzące z badania, którego celem było porównanie skuteczności dwóch terapii zapobiegających nawrotom wrzodów. Stanowi dobry przykład danych ocenianych interwałowo i został przedstawiony oraz zanalizowany w innej pracy ¹⁶.

16. Collett, D. 2003. *Modelling survival data in medical research*, wydanie drugie. Boca Raton: Chapman & Hall/CRC.

- **ulcer_recurrence_recoded.sav.** Ten plik zawiera zreorganizowane dane z pliku *ulcer_recurrence.sav*, aby umożliwić modelowanie prawdopodobieństwa zdarzeń dla każdego interwału badania, zamiast po prostu prawdopodobieństwa zdarzenia na zakończenie badania. Zagadnienie to zostało przedstawione i zanalizowane w innej pracy¹⁷.
- **verd1985.sav.** Ten plik danych dotyczy badania¹⁸. Zarejestrowano w nim odpowiedzi 15 badanych na 8 zmiennych. Interesujące nas zmienne podzielono na trzy zestawy. Zestaw 1 obejmuje *wiek* i *stan cywilny*, zestaw drugi *zwierzę domowe* i *wiadomości*, a zestaw trzeci *muzykę* i *na żywo*. Zmienna *zwierzę* jest typu wielokrotnego nominalnego, a *wiek* typu porządkowego. Wszystkie pozostałe zmienne są typu jednokrotnego nominalnego.
- **virus.sav.** Jest to plik danych hipotetycznych, który dotyczy dążenia dostawcy usług internetowych (ISP) do określenia wpływu wirusów na jego sieci. Prześlędzono (szacowany) udział procentowy ruchu zainfekowanych wiadomości e-mail w sieci dostawcy w pewnym okresie, od momentu wykrycia wirusa do powstrzymania zagrożenia.
- **wheeze_steubenville.sav.** Jest to podzbiór danych z długotrwałego badania wpływu zanieczyszczenia powietrza na zdrowie dzieci¹⁹. Dane zawierają binarne pomiary sapania dzieci w wieku 7, 8, 9 i 10 lat z miejscowości Steubenville w stanie Ohio, a także stałe informacje o tym, czy matka dziecka paliła papierosy w pierwszym roku badania.
- **workprog.sav.** Jest to plik danych hipotetycznych, który dotyczy programu prac rządowych mającego na celu ułatwienie osobom niepełnosprawnym zdobycia lepszych miejsc pracy. Prześlędzono próbę potencjalnych uczestników programu; niektórzy zostali losowo wybrani do udziału w programie, inni zaś nie. Każda obserwacja reprezentuje jednego uczestnika programu.
- **worldsales.sav** Ten plik danych hipotetycznych zawiera przychód ze sprzedaży według kontynentu i produktu.

17. Collett, D. 2003. *Modelling survival data in medical research*, wydanie drugie. Boca Raton: Chapman & Hall/CRC.

18. Verdegaal, R. 1985. *Meer sets analyse voor kwalitatieve gegevens (w jęz. niderlandzkim)*. Leiden: Department of Data Theory, University of Leiden.

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer i B. G. Ferris Jr. 1984. Passive smoking, gas cooking, and respiratory health of children living in six cities. *American Review of Respiratory Diseases*, 129, 366-374.

Informacje

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

IBM może nie oferować w innych krajach produktów, usług lub opcji omawianych w tej publikacji. Informacje o produktach i usługach dostępnych w danym kraju można uzyskać od lokalnego przedstawiciela IBM. Jakakolwiek wzmianka na temat produktu, programu lub usługi IBM nie oznacza, że może być zastosowany jedynie ten produkt, ten program lub ta usługa 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 towary i usługi, o których mowa w niniejszej publikacji. Przedstawienie tej publikacji nie daje żadnych uprawnień licencyjnych do tychże patentów. Pisemne zapytania w sprawie licencji można przysyłać 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 (RĘKOJMIĘ RÓWNIEŻ WYŁĄCZA SIĘ), WYRAŹNYCH LUB DOMNIEMANYCH, A W SZCZEGÓLNOŚCI DOMNIEMANYCH GWARANCJI PRZYDATNOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU ORAZ GWARANCJI, ŻE PUBLIKACJA TA NIE NARUSZA PRAW OSÓB 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 niniejszej publikacji mogą zawierać nieścisłości techniczne lub błędy typograficzne. 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 innych podmiotów zostały wprowadzone wyłącznie dla wygody użytkownika i w żadnym wypadku nie stanowią zachęty do ich odwiedzania. Materiały dostępne na tych stronach nie są częścią materiałów opracowanych dla tego produktu IBM, a użytkownik korzysta z nich na własną odpowiedzialność.

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

Licencjodawcy tego programu, którzy chcieliby uzyskać informacje na temat programu w celu: (i) wdrożenia wymiany informacji między niezależnie utworzonymi programami i innymi programami (łącznie z tym opisywanym) oraz (ii) wspólnego 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 tymi produktami. Pytania dotyczące możliwości produktów innych podmiotów 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 codziennej pracy. W celu kompleksowego ich zilustrowania podane przykłady zawierają nazwiska osób prywatnych, nazwy przedsiębiorstw oraz nazwy 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, handlowym 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 udzielania jakiegokolwiek 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:

© IBM 2019. Fragmenty niniejszego kodu pochodzą z programów przykładowych IBM Corp.

© Copyright IBM Corp. 1989 - 2019. Wszelkie prawa zastrzeżone.

Znaki towarowe

IBM, logo IBM oraz ibm.com są znakami towarowymi lub zastrzeżonymi znakami towarowymi International Business Machines Corp. zarejestrowanymi w wielu systemach prawnych na całym świecie. Nazwy innych produktów i usług mogą być znakami towarowymi IBM lub innych podmiotów. Aktualna lista znaków towarowych IBM dostępna jest w serwisie WWW IBM, 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 oraz Pentium są znakami towarowymi lub zastrzeżonymi znakami towarowymi Intel Corporation lub przedsiębiorstw podporządkowanych Intel Corporation w Stanach Zjednoczonych i/lub 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 The Open Group w Stanach Zjednoczonych i w innych krajach.

Java oraz wszystkie znaki towarowe i logo dotyczące języka Java są znakami towarowymi lub zastrzeżonymi znakami towarowymi Oracle i/lub przedsiębiorstw afiliowanych Oracle.

Indeks

D

dominanta
tabele użytkownika 5

M

maksimum
tabele użytkownika 5
mediana
tabele użytkownika 5
miejsca dziesiętne
ustawianie liczby miejsc dziesiętnych
wyświetlanych w tabelach
użytkownika 3
minimum
tabele użytkownika 5

N

N ważnych
tabele użytkownika 5

O

odchylenie standardowe
tabele użytkownika 5

P

pliki przykładowe
lokalizacja 9
podsumowania
tabele użytkownika 7
poziom pomiaru
zmiana w tabelach użytkownika 1
procenty
w tabelach użytkownika 3, 4
zestawy wielokrotnych odpowiedzi 5
przetwarzanie podzielonych danych
tabele użytkownika 3

R

rozstęp
tabele użytkownika 5

S

suma
tabele użytkownika 5
sumy pośrednie
tabele użytkownika 7

Ś

średnia
tabele użytkownika 5

T

tabele
tabele użytkownika 1
tabele użytkownika
etykiety wartości zmiennych
kategorialnych 1
formaty wyświetlania 3
podsumowania 7
procenty 3, 4
procenty dla zestawów wielokrotnych
odpowiedzi 5
przeliczone kategorie 7
przetwarzanie podzielonych danych 3
statystyki podsumowujące 3, 4, 5
sumy pośrednie 7
testy 7
tworzenie tabeli 2
wykluczanie kategorii 7
zestawy wielokrotnych odpowiedzi 1
zmiana kolejności kategorii 7
zmiana liczby wyświetlanych miejsc
dziesiętnych 3
zmiana poziomu pomiaru 1
zmiennie ilościowe 1
zmiennie kategorialne 1
testy
tabele użytkownika 7
testy istotności
tabele użytkownika 7

U

usuwanie kategorii
tabele użytkownika 7

W

wariancja
tabele użytkownika 5
wykluczanie kategorii
tabele użytkownika 7

Z

zestawy wielokrotnych odpowiedzi
procenty 5
zmiana kolejności kategorii
tabele użytkownika 7



Drukowane w USA