

IBM SPSS Missing Values 29



Uwaga

Przed użyciem tych informacji i produktu, którego one dotyczą, przeczytaj informacje znajdujące się w sekcji “Uwagi” na stronie 19.

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. Brakujące dane.....	1
Wprowadzenie do braków danych.....	1
Analiza braków danych.....	1
Wyświetlanie układów braków danych.....	3
Wyświetlanie statystyk opisowych braków danych.....	4
Szacowanie statystyk i podstawianie braków danych.....	5
Dodatkowe właściwości komendy MVA.....	7
Wielokrotne podstawienia.....	7
Analiza układów	9
Podstaw brakujące wartości	9
Dodatkowe właściwości komendy MULTIPLE IMPUTATION.....	13
Praca z danymi wielokrotnych podstawień.....	13
Analizowanie danych wielokrotnych podstawień.....	14
Wiele opcji podstawiania.....	18
Uwagi.....	19
Znaki towarowe.....	20
Indeks.....	23

Rozdział 1. Brakujące dane

Następujące funkcje związane z wartościami brakującymi są dostępne w SPSS Statistics Premium Edition oraz w opcji Brakujące wartości.

Wprowadzenie do braków danych

Obserwacje z brakami danych stanowią duże wyzwanie, ponieważ typowe procedury modelowania po prostu usuwają takie obserwacje z analizy. Jeśli istnieje kilka braków danych (w dużym uproszczeniu mniej niż 5% całkowitej liczby obserwacji) i wartości te mogą zostać uznane jako braki występujące losowo (czyli, to, czy wartość jest brakiem danych nie zależy od innych wartości), wówczas typowa metoda usuwania obserwacjami jest relatywnie "bezpieczna". Opcja Brakujące wartości może ułatwić ustalenie, czy usuwanie obserwacjami jest wystarczające i jeśli nie jest, udostępnia metody umożliwiające obsługę braków danych.

Analiza braków danych a wielokrotne podstawienia

Opcja Brakujące wartości oferuje dwa zestawy procedur do obsługi braków danych:

- Procedura Wielokrotne podstawienia umożliwia analizę wzorców braków danych, której celem jest ewentualne wykonanie wielokrotnych podstawień braków danych. Oznacza to, że tworzonych jest kilka wersji zestawów danych, a każdy z nich zawiera własny zestaw podstawionych wartości. Podczas wykonywania analizy statystycznej zbierane są oszacowania parametrów dla wszystkich podstawionych zestawów danych, udostępniając oszacowania, które są zwykle bardziej dokładne niż wówczas, gdyby zostały poddane tylko jednemu podstawieniu.
- Analiza braków danych oferuje nieco inny zestaw narzędzi opisowych do analizowania braków danych (najczęściej MCAR test) i obejmuje różne metody pojedynczego podstawienia. Należy pamiętać, że wielokrotne podstawienia to metoda zwykle traktowana jako nadrzędna względem pojedynczego podstawienia.

Zadania związane z brakami danych

Analizę braków danych można rozpocząć od następujących trzech podstawowych kroków:

1. **Sprawdź braki.** Za pośrednictwem procedury Analiza braków danych i Analiza układów przeprowadź eksplorację wzorów braków danych i ustal, czy konieczne jest wielokrotne podstawienie.
2. **Podstaw braki danych.** Opcja Podstaw brakujące wartości umożliwia wielokrotne podstawienie braków danych.
3. **Przeprowadź analizę kompletnych danych.** Użyj dowolnej procedury, która obsługuje dane wielokrotnych podstawień. Informacje dotyczące analizowania zestawów danych wielokrotnych podstawień oraz listę procedur, które obsługują te dane można znaleźć w sekcji "Analizowanie danych wielokrotnych podstawień" na stronie 14.

Analiza braków danych

Procedura analizy braków danych realizuje trzy podstawowe funkcje:

- Opisuje wzorzec braków danych. Gdzie znajdują się braki danych? Jak obszerne są te braki? Czy pary zmiennych mają tendencję do braku danych w wielu obserwacjach? Czy wartości danych są ekstremalne? Czy braki danych są losowe?
- Oszacowuje średnie, odchylenia standardowe, kowariancje i korelacje dla różnych metod określania braków danych: obserwacjami, parami, regresją lub EM (maksymalizacja wartości oczekiwanych). W przypadku metody wyznaczania parami wyświetlane są także liczby pełnych obserwacji parami.
- Wypełnia (podstawia) wartości oszacowane zamiast braków danych, korzystając z metod regresji lub EM; jednak wielokrotne podstawianie jest na ogół traktowane jako źródło bardziej dokładnych wyników.

Analiza braków danych ułatwia odpowiednie postępowanie w przypadku kilku problemów wywołanych przez niekompletne dane. Jeśli obserwacje z brakami danych systematycznie różnią się od obserwacji bez braków danych, wyniki mogą być mylące. Ponadto braki danych mogą zmniejszać dokładność obliczanych statystyk, ponieważ ilość informacji jest mniejsza niż pierwotnie zakładano. Dodatkowy problem polega na tym, że założenia stanowiące podstawę wielu procedur statystycznych są oparte na kompletnych obserwacjach, a braki danych mogą spowodować nadmierną złożoność wymaganych podstaw teoretycznych.

Przykład. W ramach oceny terapii białaczki mierzonych jest kilka zmiennych. Jednak nie wszystkie pomiary są dostępne u każdego pacjenta. Wyświetlono układy braków danych w formie tabelarycznej i stwierdzono, że są one losowe. Analiza EM stosowana jest do szacowania średnich, korelacji i kowariancji. Umożliwia też określenie, czy braki danych są całkowicie losowe. Następnie brakujące wartości są zastępowane wartościami podstawionymi i zapisywane w nowym pliku danych do dalszej analizy.

Statystyki. Statystyki jednej zmiennej, w tym liczba niebrakujących wartości, średnia, odchylenie standardowe, liczba brakujących wartości i liczba wartości skrajnych. Oszacowane średnie, macierz kowariancji i macierz korelacji przy użyciu metody obserwacjami, parami, EM lub regresji. Test MCAR Little'a z wynikami EM. Podsumowanie średnich różnymi metodami. Dla grup zdefiniowanych przez wartości brakujące i niebrakujące: testy *t*. Dla wszystkich zmiennych: układy braków danych prezentowane obserwacjami wg zmiennych.

Zagadnienia dotyczące danych

Dane. Dane mogą być jakościowe lub ilościowe (skala lub ciągłe). Jednak oszacowania statystyk i podstawianie brakujących danych jest możliwe tylko w przypadku zmiennych ilościowych. Dla każdej zmiennej braki danych niezakodowane jako systemowe muszą być zdefiniowane jako zdefiniowane przez użytkownika. Na przykład, jeśli w kwestionariuszu odpowiedzi *Nie wiem* na jedno z pytań przypisano kod 5 i wartość ta ma być traktowana jako brak danych, to pytanie powinno mieć kod 5 jako brak danych zdefiniowany przez użytkownika.

Wagi liczebności. Wagi częstotliwości (replikacji) są uwzględniane przez tę procedurę. Obserwacje z ujemną lub zerową wagą replikacji są ignorowane. Wagi niecałkowite są obcinane.

Założenia. Oszacowanie obserwacjami, parami i metodą regresji realizowane jest przy założeniu, że układy braków danych nie są zależne od wartości danych. (Ten warunek znany jest jako **braki całkowicie losowe** lub MCAR). Zatem wszystkie metody szacowania (w tym metoda EM) generują spójne i nieobciążone oszacowania korelacji i kowariancji, gdy zmienne spełniają warunek MCAR. Naruszenie założenia MCAR może prowadzić do generowania obciążonych oszacowań obserwacjami, parami lub metodą regresji. Jeśli dane nie spełniają warunku MCAR, konieczne jest użycie oszacowania EM.

Oszacowanie EM jest realizowane przy założeniu, że układ braków danych jest związany wyłącznie z danymi obserwowanymi. (Ten warunek znany jest jako **braki losowe** lub MAR). Założenie to umożliwia skorygowanie oszacowań na podstawie dostępnych informacji. Na przykład w badaniu wykształcenia i dochodów w danych osób mniej wykształconych może być więcej brakujących wartości dochodów. W tym przypadku dane spełniają warunek MAR, a nie MCAR. Innymi słowy, w przypadku braków losowych (MAR) prawdopodobieństwo zarejestrowania przychodu zależy od wykształcenia obiektu. Prawdopodobieństwo może być zmienne i zależne od wykształcenia, a nie od przychodów *osób z tym samym wykształceniem*. Jeśli prawdopodobieństwo zarejestrowania przychodu zależy także od wartości przychodu wśród osób z tym samym wykształceniem (np. osoby o wyższych przychodach nie ujawniają ich), to dane nie spełniają ani warunku MCAR, ani MAR. Nie jest to wcale rzadka sytuacja, a jeśli występuje, to żadna z metod nie jest odpowiednia do analizy.

Procedury pokrewne. Wiele procedur dopuszcza oszacowania obserwacjami lub parami. Procedury Regresja liniowa i Analiza czynnikowa dopuszczają zastępowanie wartości brakujących średnimi. W module dodatkowym Prognozowanie dostępnych jest kilka metod zastępowania braków danych w szeregach czasowych.

Wykonywanie analizy braków danych

1. Z menu wybierz:

Analiza > Analiza braków danych...

- Wybierz co najmniej jedną zmienną ilościową (skali) do statystyki oszacowania i opcjonalnie podstawienie braków danych.

Opcjonalnie można wykonać następujące czynności:

- Wybierz zmienne jakościowe (liczbowe lub łańcuchowe) i wprowadź limit liczby kategorii w polu **Maksimum kategorii**).
- Kliknij opcję **Układy**, aby wygenerować tabelę układów braków danych. Więcej informacji można znaleźć w temacie [“Wyświetlanie układów braków danych”](#) na stronie 3.
- Kliknij opcję **Statystyki opisowe**, aby wyświetlić statystyki opisowe braków danych. Więcej informacji można znaleźć w temacie [“Wyświetlanie statystyk opisowych braków danych”](#) na stronie 4.
- Wybierz metodę szacowania statystyk (średnie, kowariancje i korelacje) oraz ewentualnego podstawiania braków danych. Więcej informacji można znaleźć w temacie [“Szacowanie statystyk i podstawianie braków danych”](#) na stronie 5.
- W przypadku wybrania opcji **EM** lub Regresja kliknij przycisk Zmienne, aby określić podzbiór, który ma być używany dla oszacowania. Więcej informacji można znaleźć w temacie [“Zmienne przewidywane i predyktory”](#) na stronie 6.
- Wybierz zmienną etykiet obserwacji. Wartości tej zmiennej będą używane jako etykiety obserwacji w tabelach układów zawierających poszczególne obserwacje.

Wyświetlanie układów braków danych

Można wyświetlać różne tabele przedstawiające układy i ilości braków danych. Na podstawie tych tabel można zorientować się:

- Gdzie występują braki danych.
- Czy pary zmiennych mają tendencję do braku danych w poszczególnych obserwacjach.
- Czy wartości danych są skrajne.

Wyświetl

Dostępne są trzy typy tabel przedstawiających układy braków danych.

Tabela obserwacji. Układy braków danych w zmiennych analizie są przedstawione w tabelach, a dla każdego układu podane są częstości. Opcja **Sortuj zmienne według układów braków danych** umożliwia określenie, czy liczebności i zmienne mają być posortowane według podobieństwa układów. Opcja **Pomiń układy zawierające mniej niż n% obserwacji**, aby wyeliminować rzadko występujące wzorce.

Obserwacje z brakami danych. Dla każdej zmiennej analizy wymieniona jest każda obserwacja z wartością brakującą lub skrajną. Opcja **Sortuj zmienne według układów braków danych** umożliwia określenie, czy liczebności i zmienne mają być posortowane według podobieństwa układów.

Wszystkie obserwacje. Wymienione są wszystkie obserwacje, a dla każdej zmiennej wskazane są wartości brakujące i skrajne. Obserwacje są wymienione w kolejności, w jakiej występują w pliku danych, chyba że w polu **Sortuj według** określono zmienną.

W tabelach przedstawiających poszczególne obserwacje stosowane są następujące symbole:

- + . Niezwykle wysoka wartość
- . Niezwykle niska wartość
- S.** Systemowy brak danych
- A.** Pierwszy typ brakującej wartości użytkownika
- B.** Drugi typ brakującej wartości użytkownika
- C.** Trzeci typ brakującej wartości użytkownika

Zmienne

Można przedstawić dodatkowe informacje o zmiennych uwzględnionych w analizie. Zmienne dodane do pola **Dodatkowe informacje o** są wyświetlane indywidualnie w tabeli układów braków danych.

W przypadku zmiennych ilościowych (skali) przedstawiona jest średnia; w przypadku zmiennych kategoryalnych przedstawiona jest liczba obserwacji z danym układem w każdej kategorii.

- **Sortuj według.** Obserwacje są wymienione w kolejności rosnącej lub malejącej wartości określonej zmiennej. Opcja dostępna tylko w przypadku, gdy wybrano **Wszystkie obserwacje**.

Aby wyświetlić układy braków danych

1. W głównym oknie dialogowym Analiza braków danych wybierz zmienne, których układy braków danych chcesz wyświetlić.
2. Kliknij opcję **Układy**.
3. Wybierz tabele układów, które chcesz wyświetlić.

Wyświetlanie statystyk opisowych braków danych

Statystyki jednej zmiennej

Statystyki jednej zmiennej mogą pomóc w ogólnym określeniu ilości braków danych. Przedstawiane są następujące informacje o każdej zmiennej:

- Liczba wartości niebrakujących
- Liczba i odsetek braków danych

W przypadku zmiennych ilościowych (skali) wyświetlane są także następujące informacje:

- Średnia
- Odchylenie standardowe
- Liczba wartości skrajnie wysokich i niskich

Statystyki zmiennych wskaźnikowych

Dla każdej zmiennej tworzona jest zmienna wskaźnikowa. Jest to zmienna jakościowa wskazująca, czy w danej obserwacji zmienna jest obecna, czy brakująca. Zmienne wskaźnikowe używane są do tworzenia tabel niedopasowania, testu t i częstości.

Procent niedopasowań. Dla każdej pary zmiennych przedstawia odsetek obserwacji, w których jedna zmienna zawiera wartość brakującą, a druga zmienna zawiera wartość niebrakującą. Każdy element na przekątnej tabeli zawiera odsetek braków danych jednej zmiennej.

Testy t grup zdefiniowanych przez zmienną wskaźnikową. Średnie dwóch grup są porównywane dla każdej zmiennej ilościowej przy użyciu testu t Studenta. Przynależność do grup określa, czy zmienna jest obecna, czy brakująca. Prezentowana jest statystyka t , liczba stopni swobody, liczebność wartości brakujących i niebrakujących oraz średnie dwóch grup. Można też przedstawić dwustronne prawdopodobieństwa związane ze statystyką t . Jeśli w wyniku analizy obliczony zostanie więcej niż jeden test, nie należy używać tych prawdopodobieństw do testowania istotności. Prawdopodobieństwa można wykorzystać tylko wtedy, gdy obliczony został jeden test.

Tabela krzyżowa zmiennej jakościowej ze wskaźnikową. Dla każdej zmiennej jakościowej przedstawiona jest jedna tabela. Dla każdej kategorii tabela zawiera częstość i odsetek wartości niebrakujących pozostałych zmiennych. Przedstawione są także odsetki każdego typu wartości brakujących.

Pomiń zmienne zawierające mniej braków danych niż n% obserwacji. Aby ograniczyć rozmiar tabeli, można pominąć obliczenie statystyk dla niewielkiej liczby obserwacji.

Aby wyświetlić statystyki opisowe

1. W głównym oknie dialogowym Analiza braków danych wybierz zmienne, których statystyki opisowe braków danych chcesz wyświetlić.
2. Kliknij przycisk **Statystyki opisowe**.
3. Wybierz statystyki opisowe, które chcesz wyświetlić.

Szacowanie statystyk i podstawianie braków danych

Istnieje możliwość oszacowania średnich, standardowych odchyłeń, kowariancji i korelacji z użyciem metod usuwania braków danych obserwacjami (tylko kompletne obserwacje), usuwanie braków danych parami, algorytmu EM (maksymalizacja oczekiwań) i/lub regresji. Można również podstawić braki danych (oszacowanie wartości zastępowanych). Należy pamiętać, że procedura Wielokrotne podstawienia jest ogólnie uznawana za nadrzędną wobec pojedynczego podstawia przy rozwiązywaniu problemów z brakami danych. Procedura MCAR test jest nadal przydatna do ustalenia, czy podstawienie jest konieczne.

Metoda usuwania braków danych obserwacjami

Ta metoda ma zastosowanie tylko w przypadku kompletnych obserwacji. Jeśli któraś ze zmiennych analizowanych zawiera braki danych, obserwacja zostanie pominięta w obliczeniach.

Metoda usuwania braków danych parami

Ta metoda sprawdza pary zmiennych analizowanych i używa obserwacji tylko wówczas, gdy nie zawiera braków danych w obu zmiennych. Częstości, średnie i odchylenia standardowe są obliczane osobno dla każdej pary. Ponieważ pozostałe braki danych w obserwacji są ignorowane, korelacje i kowariancje dwóch zmiennych nie zależą od braków danych w innych zmiennych.

Metoda algorytmu EM (maksymalizacji oczekiwań)

Ta metoda zakłada rozkład częściowych braków danych i opiera wnioski na prawdopodobieństwie tego rozkładu. Każda iteracja składa się z kroku E i kroku M. Krok E znajduje oczekiwania warunkowe dla "braków" danych, zadanych wartości obserwowanych i bieżących oszacowań parametrów. Oczekiwania te są następnie podstawiane za "braki" danych. W kroku M oszacowanie maksymalnego prawdopodobieństwa parametrów jest wyznaczane, tak jakby braki danych były wypełnione. "Brak" jest ujęty w cudzysłów, ponieważ braki danych nie są bezpośrednio wypełniane. Zamiast tego ich funkcje są używane w logarytmie wiarygodności.

Statystyka Roderick J. A. Little's chi-kwadrat do testowania, czy wartości są kompletnymi losowymi brakami danych (ang. missing completely at random (MCAR)), jest drukowana jako przypis w macierzach EM. W przypadku tego testu hipoteza zerowa zakłada, że dane są kompletnymi losowymi brakami danych, a wartość p jest istotna na poziomie 0,05. Jeśli wartość jest mniejsza niż 0,05, dane nie są kompletnymi losowymi brakami danych. Dane mogą być losowymi brakami danych (MAR) lub mogą nie być losowymi brakami danych (NMAR). Nie można założyć jednego lub drugiego przypadku i konieczna jest analiza danych pozwalająca określić typ braków danych.

Metoda regresji

Ta metoda umożliwia obliczenie wielu oszacowań regresji liniowej i udostępnia opcje tworzenia argumentów dla oszacowań z użyciem składników losowych. Dla każdej przewidywanej wartości procedura może dodać resztę z losowo wybranej kompletnej obserwacji, losowe odchylenie normalne lub odchylenie losowe (skalowane wg pierwiastka kwadratowego resztowego kwadratu średniego) z rozkładu t .

Opcje estymacji metodą EM

Metoda EM iteracyjnie estymuje średnie, macierz kowariancji i korelację zmiennych ilościowych (skali) z brakami danych.

Rozkład. Algorytm EM prowadzi wnioskowanie na podstawie wiarygodności przy określonym rozkładzie. Domyślnie zakłada się rozkład normalny. Jeśli wiadomo, że ogony rozkładu są dłuższe niż w rozkładzie normalnym, można zażądać, by procedura zbudowała funkcję wiarygodności na podstawie rozkładu t Studenta z n stopniami swobody. Rozkład normalny mieszany również charakteryzuje się dłuższymi ogonami. Należy określić stosunek odchyłeń standardowych rozkładu normalnego mieszane i proporcję, w jakiej oba rozkłady są wymieszane. Przy rozkładzie normalnym mieszanym zakłada się, że różnią się wyłącznie odchylenia standardowe rozkładów. Średnie muszą być takie same.

Maksymalna liczba iteracji. Określa maksymalną liczbę iteracji potrzebną do oszacowania prawdziwej kowariancji. Procedura kończy działanie po osiągnięciu tej liczby iteracji, nawet jeśli estymacje nie osiągnęły jeszcze zbieżności.

Zapisz kompletne dane. Można zapisać zbiór danych z wartościami podstawionymi w miejsce wartości brakujących. Należy jednak pamiętać, że statystyki oparte na kowariancji z wykorzystaniem wartości podstawionych będą niedoszacowywać wartości odpowiednich parametrów. Stopień niedoszacowania jest proporcjonalny do liczby przypadków łącznie niezaobserwowanych.

Aby określić opcje metody EM

1. W głównym oknie dialogowym Analiza braków danych wybierz zmienne, dla których chcesz oszacować braki danych metodą EM.
2. W grupie Estymacja wybierz metodę **EM**.
3. Aby określić zmienne przewidywane i predyktory, kliknij przycisk **Zmienne**. Więcej informacji można znaleźć w temacie [“Zmienne przewidywane i predyktory”](#) na stronie 6.
4. Kliknij pozycję **EM**.
5. Wybierz żądane opcje EM.

Opcje estymacji metodą regresji

W metodzie regresji brakujące wartości estymowane są przy użyciu regresji liniowej wielokrotnej. Przedstawiane są średnie, macierz kowariancji i macierz korelacji zmiennych przewidywanych.

Korekcja oszacowania. Metoda regresji może dodawać składnik losowy do oszacowań. Można wybrać reszty, zmienne losowe normalne, zmienne losowe *t* Studenta albo brak korekcji.

- *Residuals*. Składniki błędów są wybierane losowo z obserwowanych reszt pełnych obserwacji w celu dodania do oszacowań regresji.
- *Zmienne losowe normalne*. Składniki błędów losowego są losowo pobierane z rozkładu o wartości oczekiwanej 0 i odchyleniu standardowym równym pierwiastkowi średniego błędów kwadratowego regresji.
- *Zmienne losowe t Studenta*. Składniki błędów są losowo pobierane z rozkładu *t*(*n*) o określonej liczbie stopni swobody i skalowane błędem średnim kwadratowym (RMSE).

Maksymalna liczba predyktorów. Określa maksymalną liczbę predyktorów (zmiennych niezależnych) używanych w procesie estymacji.

Zapisz kompletne dane. Powoduje zapisanie zbioru danych w bieżącej sesji lub zewnętrznym pliku danych IBM SPSS Statistics, z wartościami brakującymi zastąpionymi przez wartości oszacowane metodą regresji.

Określanie opcji regresji

1. W głównym oknie dialogowym Analiza braków danych wybierz zmienne, dla których chcesz oszacować braki danych metodą regresji.
2. W grupie Estymacja wybierz metodę **Regresja**.
3. Aby określić zmienne przewidywane i predyktory, kliknij przycisk **Zmienne**. Więcej informacji można znaleźć w temacie [“Zmienne przewidywane i predyktory”](#) na stronie 6.
4. Kliknij opcję **Regresja**.
5. Wybierz żądane opcje regresji.

Zmienne przewidywane i predyktory

Domyślnie w estymacji metodą EM i regresji używane są wszystkie zmienne ilościowe. W razie potrzeby można wybrać konkretne zmienne jako przewidywane i predyktory w estymacjach. Dana zmienna może figurować na obu listach, ale mogą zdarzyć się sytuacje, w których celowe jest ograniczenie użycia zmiennej do jednej roli. Na przykład niektórzy analitycy nie chcą estymować wartości zmiennych wynikowych. Niekiedy celowe jest także użycie różnych zmiennych do różnych estymacji i wielokrotne

uruchamianie procedury. Załóżmy na przykład, że mamy zbiór ocen pielęgniarek i drugi zbiór zawierający oceny lekarzy. W takiej sytuacji celowe może być uruchomienie procedury najpierw na ocenach pielęgniarek, w celu oszacowania braków danych w tym zbiorze, a następnie uruchomienie procedury szacowania braków danych w ocenach lekarzy.

Inna kwestia występuje w przypadku metody regresji. W regresji wielokrotnej użycie dużego podzbioru zmiennych niezależnych może doprowadzić do wygenerowania niższej jakości wartości przewidywanych niż użycie mniejszego podzbioru. Dlatego zmienna, aby została użyta, musi osiągać graniczną wartość F wprowadzania równą 4,0. Tę wartość graniczną można zmienić za pomocą komendy.

Aby określić zmienne przewidywane i predyktory

1. W głównym oknie dialogowym Analiza braków danych wybierz zmienne, dla których chcesz oszacować braki danych metodą regresji.
2. W grupie Estymacja wybierz metodę **EM** lub **Regresja**.
3. Kliknij opcję **Zmienne**.
4. Aby w roli zmiennych przewidywanych i predyktorów użyć wybranych, a nie wszystkich zmiennych, wybierz opcję **Wybór zmiennych** i przenieś zmienne na odpowiednie listy.

Dodatkowe właściwości komendy MVA

Język składni komend umożliwia również:

- Określenie odrębnych zmiennych opisowych dla układów braków wartości, układów braków danych i tabeli układów braków danych za pomocą słowa kluczowego DESCRIBE w opcji komendy MPATTERN, DPATTERN lub TPATTERN.
- Określenie więcej niż jednej zmiennej sortowania dla tabeli układów braków danych za pomocą opcji komendy DPATTERN.
- Określenie więcej niż jednej zmiennej sortowania dla układów braków danych za pomocą opcji komendy DPATTERN.
- Określenie tolerancji i konwergencji za pomocą opcji komendy EM.
- Określenie tolerancji i F -wprowadzenia za pomocą opcji komendy REGRESSION.
- Określenie różnych list zmiennych dla metody EM i regresji za pomocą opcji komend EM i REGRESSION.
- Określenie różnych wartości procentowych dla ukrycia wyświetlanych obserwacji, dla każdej wartości TTESTS, TABULATE i MISMATCH.

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

Wielokrotne podstawienia

Celem podstawień wielokrotnych jest wygenerowanie możliwych wartości dla braków danych i utworzenie kilku „kompletnych” zbiorów danych. Procedury analityczne, które działają ze zbiorami danych z wielokrotnymi podstawieniami generują dane wyjściowe dla każdego „kompletnego” zbioru danych, a dodatkowo generują wynik kombinowany, który oszacowuje, jakie byłyby wyniki, gdyby oryginalny zbiór danych nie zawierał braków danych. Te wyniki kombinowane są zwykle bardziej dokładne niż udostępniane przez metody podstawień pojedynczych.

Wymagania dotyczące danych wielokrotnych podstawień












Zmienne analizowane. Możliwe zmienne analizowane to:

- *Nominalny.* Zmienna może być traktowana jako nominalna, gdy jej wartości reprezentują kategorie bez wewnętrznego rankingu (na przykład dział przedsiębiorstwa, w którym pracuje pracownik). Przykładami zmiennych nominalnych są: region, kod pocztowy lub wyznanie.
- *Porządkowy.* Zmienna może być traktowana jako porządkowa, gdy jej wartości reprezentują kategorie z jakimś nieodłącznym rangą (na przykład poziomy zadowolenia z usługi z bardzo niezadowolonego do bardzo zadowolonego). Przykładami zmiennych porządkowych mogą być oceny opinii reprezentujące stopień satysfakcji lub przekonania oraz oceny preferencji.

- **scale.** Zmienna może być traktowana jako skala (ilościowa), gdy jej wartości reprezentują uporządkowane kategorie z miarodajnym pomiarem, tak aby porównania odległości między wartościami były odpowiednie. Przykładami zmiennych ilościowych mogą być: wiek w latach lub przychód w tysiącach złotych.

W procedurze przyjmuje się, że odpowiedni poziom pomiaru został przypisany do wszystkich zmiennych, można jednak tymczasowo zmienić poziom pomiaru dla zmiennej, klikając prawym przyciskiem myszy zmienną na liście zmiennych źródłowych i wybierając poziom pomiaru z menu kontekstowego. Informacje na temat trwałej zmiany poziomu pomiaru dla zmiennej

Ikona obok każdej zmiennej na liście zmiennych określa poziom pomiaru oraz typ danych:

Tabela 1. Ikony poziomu pomiaru				
	Liczbowy	Łańcuch	Data	Czas
Zmienna (ilościowa)		nie dotyczy		
Porządkowy				
Nominalna				

Wagi liczebności. Wagi częstotliwości (replikacji) są uwzględniane przez tę procedurę. Obserwacje z ujemną lub zerową wagą replikacji są ignorowane. Wagi niecałkowite są zaokrąglane do najbliższej liczby całkowitej.

Waga analizy. Wagi analizy (regresji lub losowania) są włączane do podsumowań braków danych oraz do modeli dopasowania podstawień. Obserwacje z ujemną lub zerową wagą analizy są wykluczane.

Próby złożone. Procedura Wielokrotne podstawienia nie obsługuje w sposób jawny warstw, skupień lub innych złożonych struktur losowania, jednak może akceptować finalne wagi losowania w postaci zmiennej ważącej analizy. Należy również pamiętać, że procedury złożonego losowania aktualnie nie umożliwiają analizy zestawów danych wielokrotnie podstawianych. Pełną listę procedur obsługujących łączenie w pulę można znaleźć w temacie [“Analizowanie danych wielokrotnych podstawień”](#) na stronie 14.

Braki danych. Braki danych użytkownika i systemowe braki danych są traktowane jako wartości nieprawidłowe; oznacza to, że oba typy braków danych są zastępowane w czasie podstawiania wartości i oba typy są traktowane jako wartości nieprawidłowe zmiennych używanych jako predyktory w modelach podstawień. Braki danych użytkownika i systemowe braki danych są również traktowane jako braki w analizie braków danych.

Replikacja wyników (podstawianie braków danych). Aby przeprowadzić dokładną replikację wyników podstawień, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych, tej samej kolejności danych oraz tej samej kolejności zmiennych oraz użyć tych samych ustawień procedury.

- **Generowanie liczb losowych.** Procedura ta używa generowania liczb losowych podczas obliczania podstawionych wartości. Aby ponownie utworzyć takie same losowe wyniki w przyszłości, należy użyć tej samej wartości inicjalizacji dla generatora liczb losowych przed każdym uruchomieniem procedury podstawiania braków danych. .
- **Kolejność obserwacji.** Wartości są podstawiane w kolejności obserwacji.
- **Kolejność zmiennych.** Metoda całkowicie warunkowej specyfikacji (FCS) podstawia wartości w kolejności określonej na liście Zmienne analizowane.

Istnieją dwie okna dialogowe przeznaczone do pracy z wielokrotnymi podstawieniami.

- **Analiza układów** oferuje miary opisowe wzorów braków danych i może być przydatna jako krok badawczy przed rozpoczęciem podstawiania.

- **Podstaw brakujące wartości** to procedura używana do generowania wielokrotnych podstawień. Procedury dopuszczające wiele zbiorów danych podstawień pozwalają na analizę kompletnych zbiorów danych. Sekcja [“Analizowanie danych wielokrotnych podstawień”](#) na stronie 14 zawiera informacje o analizowaniu zbiorów danych z wielokrotnymi podstawieniami oraz listę procedur obsługujących takie dane.

Analiza układów

Analiza układów wyznacza opisowe miary układów brakujących wartości danych i może być przydatna jako krok eksploracyjny przed podstawieniem wartości.

Przykład. Operator telekomunikacyjny chce na podstawie bazy danych klientów przeanalizować sposób korzystania przez nich z usług. Operator dysponuje kompletem danych o wykorzystaniu usług przez klientów, ale w danych demograficznych dot. klientów jest wiele braków. Analiza układów braków danych może pomóc w określeniu następnych kroków podstawiania.

Z menu wybierz:

Analiza > Wielokrotne podstawienia > Analiza układów...

1. Wybierz co najmniej dwie zmienne analizy. Procedura przeanalizuje układy braków danych w tych zmiennych.

Ustawienia opcjonalne

Waga analizy. Ta zmienna zawiera wagi analizy (regresji lub doboru prób). Procedura uwzględnia wagi analizy w podsumowaniach braków danych. Obserwacje z ujemną lub zerową wagą analizy są wykluczane.

Wynik. Dostępne są następujące wyniki opcjonalne:

- **Podsumowanie braków danych.** Podzielony na panele wykres kołowy przedstawiający liczby i odsetki zmiennych analizy, obserwacji lub poszczególnych wartości danych z co najmniej jednym brakiem danych.
- **Układy braków danych.** Przedstawia układy braków danych w formie tabelarycznej. Każdy układ odpowiada grupie obserwacji z tym samym układem danych kompletnych i niekompletnych w zmiennej analizy. Można wykorzystać ten wynik do ustalenia, czy względem danych można użyć monotonicznej metody podstawiania, a jeśli nie, to jak dokładnie dane odzwierciedlają układ monotoniczny. Procedura ta określa kolejność zmiennych analizy tak, by ujawnić lub przybliżyć układ monotoniczny. Jeśli po zmianie kolejności zmiennych nie ujawnia się układ niemonotoniczny, można wysnuć wniosek, że przy takiej kolejności zmiennych analizy dane mają układ monotoniczny.
- **Zmienne o najwyższej częstotliwości braków danych.** Przedstawia tabelę zmiennych analizy posortowaną według odsetka braków danych, w kolejności malejącej. Tabela zawiera statystyki opisowe (średnią i odchylenie standardowe) zmiennych ilościowych.

Można wpływać na maksymalną liczbę wyświetlanych zmiennych i minimalny odsetek braków danych, przy którym zmienna będzie uwzględniana w wyniku. Prezentowany jest zbiór zmiennych spełniających oba te kryteria. Na przykład, gdy maksymalna liczba zmiennych zostanie określona jako 50, a minimalny odsetek braków danych jako 25, to tabela będzie zawierać maksymalnie 50 zmiennych mających co najmniej 25% braków danych. Jeśli istnieje 60 zmiennych analizy, ale tylko 15 ma co najmniej 25% braków danych, to wyniki będą zawierały tylko 15 zmiennych.

Podstaw brakujące wartości

Procedura **Podstaw brakujące wartości** służy do generowania podstawień wielokrotnych. Procedury dopuszczające wiele zbiorów danych podstawień pozwalają na analizę kompletnych zbiorów danych. Sekcja [“Analizowanie danych wielokrotnych podstawień”](#) na stronie 14 zawiera informacje o analizowaniu zbiorów danych z wielokrotnymi podstawieniami oraz listę procedur obsługujących takie dane.

Przykład. Operator telekomunikacyjny chce na podstawie bazy danych klientów przeanalizować sposób korzystania przez nich z usług. Operator dysponuje kompletem danych o wykorzystaniu usług przez

klientów, ale w danych demograficznych dot. klientów jest wiele braków. Co więcej, braki te nie występują całkowicie losowo, dlatego zbiór danych zostanie uzupełniony poprzez podstawienia wielokrotne.

Z menu wybierz:

Analiza > Wielokrotne podstawienia > Podstaw brakujące wartości...

1. Wybierz co najmniej dwie zmienne w modelu podstawień. Procedura podstawia wiele brakujących wartości danych w tych zmiennych.
2. Określ liczbę podstawień do obliczenia. Domyślnie wartość ta wynosi 5.
3. Określ zbiór danych lub plik danych w formacie IBM SPSS Statistics, w którym mają być zapisane podstawione dane.

Wynikowy zbiór danych zawierać będzie pierwotne dane obserwacji z brakami danych plus zbiór obserwacji z podstawionymi wartościami dla każdego podstawienia. Na przykład, jeśli pierwotny zbiór danych zawiera 100 obserwacji i wybrano pięć podstawień, wynikowy zbiór danych będzie zawierał 600 obserwacji. Wszystkie zmienne w wejściowym zbiorze danych są uwzględnione w wynikowym zbiorze danych. Właściwości słownika (nazwy, etykiety itp.) istniejących zmiennych są kopiowane do nowego zbioru danych. Plik zawiera też nową zmienną *Imputation_*. Jest to zmienna liczbowa wskazująca, czy dane pochodzą z podstawienia (0 oznacza dane pierwotne, a 1..n oznacza obserwacje z wartościami podstawionymi).

Procedura automatycznie definiuje zmienną *Imputation_* jako zmienną dzielącą) podczas tworzenia wynikowego zbioru danych. Jeśli podczas wykonywania procedury obowiązują podziały, wynikowy zbiór danych zawiera po jednym zbiorze podstawień dla każdej kombinacji wartości zmiennych dzielących.

Ustawienia opcjonalne

Waga analizy. Ta zmienna zawiera wagi analizy (regresji lub doboru prób). Procedura uwzględnia wagi analizy w modelach regresji i klasyfikacji używanych do podstawiania braków danych. Wagi analizy są też używane w podsumowaniach podstawionych wartości, na przykład średniej, odchylenia standardowego i błędu standardowego. Obserwacje z ujemną lub zerową wagą analizy są wykluczane.

Zmienne z nieznanym poziomem pomiaru

Alert poziomu pomiaru wyświetla się, gdy poziom pomiaru dla jednej lub większej ilości zmiennych w zbiorze danych jest nieznan. Ponieważ poziom pomiaru wpływa na wyliczenie wyników dla tej procedury, wszystkie zmienne muszą mieć zdefiniowany poziom pomiaru.

Skanowanie danych. Odczytuje dane w aktywnym zbiorze danych i przypisuje domyślny poziom pomiaru do wszystkich zmiennych, które mają aktualnie nieznaną poziom pomiaru. Jeśli zbiór danych jest duży, może to zająć trochę czasu.

Przypisz ręcznie. Otwiera okno dialogowe, które zestawia wszystkie zmienne z nieznanym poziomem pomiaru. Można użyć tego okna dialogowego do przypisania poziomu pomiaru do tych zmiennych. Można również przypisać poziom pomiaru w Widoku zmiennych Edytora danych.

Ponieważ poziom pomiaru jest ważny dla tej procedury, nie można wejść do tego okna dialogowego w celu uruchomienia tej procedury, dopóki wszystkie zmienne nie będą miały zdefiniowanego poziomu pomiaru.

Metoda

Na karcie Metoda określa się sposób podstawiania brakujących wartości, w tym typy używanych modeli. Predyktory jakościowe są kodowane przy użyciu wskaźników (wartości sztucznych).

Metoda podstawiania

W metodzie **Automatycznie** dane są przeszukiwane i podstawienia realizowane są monotonicznie, jeśli dane wykazują monotoniczny układ braków; w przeciwnym razie stosowana jest całkowicie warunkowa specyfikacja. Jeśli wiesz, która metoda ma być używana, określ ją jako metodę **Użytkownika**.

Całkowicie warunkowa specyfikacja

Jest to iteracyjna metoda próbkowania Monte Carlo łańcuchami Markowa (MCMC), która może być używana w przypadku dowolnego wzorca brakujących danych (monotonicznego lub niemonotonicznego).

Dla każdej iteracji i każdej zmiennej w kolejności określonej na liście zmiennych metoda całkowicie warunkowej specyfikacji dopasowuje model jednej zmiennej (z jedną zmienną zależną), używając wszystkich pozostałych zmiennych dostępnych w modelu jako predyktorów, a następnie podstawia brakujące wartości dopasowywanej zmiennej. Proces ten jest kontynuowany aż do osiągnięcia maksymalnej liczby iteracji, a podstawione wartości z maksymalnej iteracji są zapisywane w zbiorze danych z podstawieniami.

Maksymalna liczba iteracji

Określa liczbę iteracji lub „kroków” w łańcuchu Markowa używanym w metodzie całkowicie warunkowej specyfikacji. Jeśli metoda całkowicie warunkowej specyfikacji została wybrana automatycznie, przyjmuje się domyślną liczbę 10 iteracji. W przypadku jawnego wybrania tej metody użytkownik może jawnie określić liczbę iteracji. Jeśli łańcuch Markowa nie osiąga zbieżności, konieczne może być zwiększenie liczby iteracji. Na karcie wyników można zapisać dane historii iteracji metody całkowicie warunkowej specyfikacji oraz wykreślić te dane, aby ocenić zbieżność.

Monotoniczna

Jest to metoda nieiteracyjna, która może być używana tylko w przypadku, gdy dane mają monotoniczny wzorec brakujących wartości. Wzorec monotoniczny istnieje wtedy, gdy można uporządkować zmienne w taki sposób, że jeśli zmienna ma niebrakującą wartość, wszystkie poprzedzające je zmienne również mają niebrakujące wartości. Wybierając tę metodę jako metodę **Użytkownika**, należy wymienić zmienne na liście w kolejności charakteryzującej się układem monotonicznym.

Dla każdej zmiennej w porządku monotonicznym metoda monotoniczna dopasowuje model jednej zmiennej (z jedną zmienną zależną), używając wszystkich poprzednich zmiennych dostępnych w modelu jako predyktorów, a następnie podstawia brakujące wartości dopasowywanej zmiennej. Podstawione wartości są zapisywane w zbiorze danych podstawionych.

Uwzględnij interakcje dwukierunkowe

Gdy metoda podstawiania została wybrana automatycznie, model podstawiania każdej zmiennej zawiera składnik stały i efekty główne predyktorów. Wybierając metodę jawnie, można opcjonalnie uwzględnić wszystkie możliwe interakcje drugiego rzędu między predyktorami jakościowymi.

Typ modelu dla zmiennych ilościowych

Regresja liniowa

Gdy metoda podstawiania została wybrana automatycznie, jako model jednej zmiennej dla zmiennych ilościowych używana jest regresja liniowa.

Dopasowanie średnimi predykcyjnymi (PMM)

Wybierając metodę jawnie, można wybrać dopasowywanie średnimi predykcyjnymi (PMM) jako model dla zmiennych ilościowych. Metoda PMM jest wariantem regresji liniowej, która zapewnia, że podstawiane wartości są wiarygodne. W przypadku metody PMM wartość podstawiana jest oparta na wartości zdefiniowanej dla **losowo wybranej pełnej obserwacji z najbliższych (k) predykcji**, gdzie (k) jest dodatnią liczbą całkowitą o wartości domyślnej 5.

Dla zmiennych jakościowych jako model jednej zmiennej zawsze używana jest regresja logistyczna. Niezależnie od typu modelu predyktory jakościowe są kodowane wskaźnikami (wartościami sztucznymi).

Tolerancja osobliwości

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

Ograniczenia

Karta Ograniczenia umożliwia ograniczenie roli zmiennej podczas podstawiania i ograniczenie zakresu podstawianych wartości zmiennej ilościowej, tak aby były one wiarygodne. Ponadto można ograniczyć analizę do zmiennych z mniejszym od maksymalnego odsetkiem braków danych.

Skanowanie danych dla podsumowania zmiennej. Kliknięcie opcji **Skanowanie danych** powoduje, że na liście wyświetlane są zmienne analizy i dla każdej z nich zaobserwowane: odsetek brakujących, minimum oraz maksimum. Podsumowania mogą być oparte na skanowaniu pierwszych n obserwacji, zgodnie z wpisem w polu tekstowym Obserwacje. Kliknięcie opcji **Zeskanuj dane ponownie** powoduje zaktualizowanie podsumowań rozkładu.

Zdefiniuj ograniczenia

- **Role.** Umożliwia zmodyfikowanie zbioru zmiennych, które mają być podstawiane i/lub traktowane jako predyktory. Zwykle każda zmienna analizy jest w modelu podstawiania traktowana zarówno jako zmienna zależna, jak i jako predyktor. Opcję **Rola** można wykorzystać do wyłączenia podstawiania zmiennych, które mają być używane tylko jako predyktor (**Tylko użyj jako predyktora**), lub wykluczenia zmiennych z użycia w roli predyktorów (**Tylko podstaw**) – po to, by model predykcji był bardziej zwarty. Jest to jedyne ograniczenie, które można określić dla zmiennych jakościowych lub dla zmiennych używanych wyłącznie jako predyktory.
- **Min. i Maks.** Te kolumny umożliwiają określenie minimalnej i maksymalnej dozwolonej wartości podstawianej zmiennych ilościowych. Jeśli podstawiana wartość wykracza poza ten przedział, procedura próbuje wybrać następną wartość, dopóki nie znajdzie wartości mieszczącej się w przedziale lub nie zostanie osiągnięta maksymalna liczba prób (patrz **Maksymalna liczba prób parametrów**, poniżej). Te kolumny są dostępne tylko wtedy, gdy jako typ modelu zmiennej ilościowej na karcie Metoda wybrano metodę **Regresja liniowa**.
- **Zaokrąglanie.** Niektóre zmienne mogą być używane jako ilościowe, ale ich wartości są w naturalny sposób ograniczone; na przykład liczba osób w gospodarstwie domowym musi być liczbą całkowitą, a kwota wydana w sklepie spożywczym nie może mieć ułamkowej liczby groszy. Ta kolumna umożliwia określenie najmniejszego dozwolonego nominału. Na przykład, aby wymagać wartości całkowitych, należy podać 1 w polu zaokrąglania; aby wymagać wartości zaokrąglonych do grosza, należy podać 0,01. Co do zasady wartości są zaokrąglane do najbliższej całkowitej wielokrotności nominału zaokrąglania. W poniższej tabeli przedstawiono wpływ różnych wartości zaokrągleń na podstawioną wartość 6,64823 (przed zaokrągleniem).

Nominał zaokrąglania	Wartość, do której zaokrąglona zostanie liczba 6,64832
10	10
1	7
0,25	6,75
0,1	6,6
0,01	6,65

Wyklucz zmienne z dużą ilością braków danych. Zwykle zmienne analizy są podstawiane i używane jako predyktory niezależnie od tego, ile braków danych zawierają, jeśli tylko zawierają ilość danych wystarczającą do oszacowania modelu podstawiania. Można nakazać wykluczenie zmiennych o wysokim odsetku braków danych. Na przykład, jeśli jako **Maksymalny procent braków** zostanie określona liczba 50, to zmienne analizy z więcej niż 50% braków danych nie będą podstawiane ani używane jako predyktory w modelach podstawiania.

Maksymalna liczba prób parametrów. Jeśli określono minimalną i maksymalną wartość podstawianych wartości zmiennych ilościowych (patrz **Min. i Maks.**, powyżej), procedura próbuje wybierać wartości obserwacji, dopóki nie znajdzie zbioru wartości mieszczących się w zadanym przedziale. Jeśli po określonej liczbie prób przypadającej na każdą obserwację nie uda się uzyskać zbioru wartości, procedura

wybiera następny zbiór parametrów modelu i powtarza proces wybierania wartości dla obserwacji. Jeśli po określonej liczbie prób wyboru dla obserwacji i wyboru parametrów nie uda się znaleźć zbioru wartości mieszczących się w przedziałach, występuje błąd.

Należy zauważyć, że zwiększenie tych wartości może wydłużyć czas przetwarzania. Jeśli wykonywanie procedury trwa bardzo długo lub nie udaje się wybrać odpowiednich wartości, sprawdź, czy określone minimum i maksimum jest odpowiednie.

Wynik

Pokaż. Steruje prezentacją wyników. Zawsze prezentowane jest ogólne podsumowanie podstawienia, które zawiera tabele dotyczące specyfikacji podstawiania, iteracji (w metodzie całkowicie warunkowej specyfikacji), podstawione zmienne zależne, zmienne zależne wykluczone z podstawiania oraz kolejność podstawiania. Jeśli określone są ograniczenia zmiennych analizy, to również są prezentowane w wynikach.

- **Model podstawiania.** Przedstawia model podstawiania zmiennych zależnych i predyktorów. Informacje obejmują typ modelu jednej zmiennej, efekty modelu i liczbę podstawionych wartości.
- **Statystyki opisowe.** Przedstawia statystyki opisowe zmiennych zależnych, których wartości są podstawiane. W przypadku zmiennych ilościowych statystyki opisowe obejmują średnią, liczebność, odchylenie standardowe, minimum i maksimum pierwotnych danych wejściowych (przed podstawieniem), wartości podstawione (dla każdego podstawienia) oraz dane kompletne (pierwotne i podstawione — dla każdego podstawienia). W przypadku zmiennych jakościowych statystyki opisowe obejmują liczebność i odsetki według kategorii pierwotnych danych wejściowych (przed podstawieniem), wartości podstawione (dla każdego podstawienia) oraz dane kompletne (pierwotne i podstawione — dla każdego podstawienia).

Przebieg iteracji. Gdy stosowana jest metoda całkowicie warunkowej specyfikacji, można zażądać przedstawienia zbioru danych zawierającego historię iteracji wykonywanych w ramach tej metody. Taki zbiór danych zawiera średnie i odchylenia standardowe dla każdej iteracji i podstawienia każdej zmiennej ilościowej, której wartości są podstawiane. Można wykreślić dane, aby ułatwić sobie ocenę zbieżności modelu.

Dodatkowe właściwości komendy MULTIPLE IMPUTATION

Język składni komend umożliwia również:

- Określenie podzbioru zmiennych, dla których pokazane są statystyki opisowe (opcja komendy IMPUTATIONSUMMARIES).
- Określanie analizy brakujących wzorów oraz podstawień w pojedynczym przebiegu procedury.
- Określanie maksymalnej liczby parametrów modelu dozwolonych w przypadku podstawiania dowolnej zmiennej (słowo kluczowe MAXMODELPARAM).

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

Praca z danymi wielokrotnych podstawień

Jeśli tworzony jest zestaw danych wielokrotnych podstawień (MI), dodawana jest zmienna o nazwie *Imputation_*, z etykietą zmiennej *Numer podstawienia*, w zestaw danych zostanie posortowany w kolejności rosnącej. Obserwacje z oryginalnego zbioru danych mają wartość 0. Obserwacje dla podstawianych wartości są numerowane od 1 do *M*, gdzie *M* jest liczbą podstawień.

Po otwarciu zestawu danych obecność zmiennej *Imputation_* określa zestaw danych jako możliwy zestaw danych MI.

Aktywowanie zestawu danych wielokrotnych podstawień na potrzeby analizy

Zestaw danych należy podzielić, używając opcji **Porównaj grupy** i zmiennej *Imputation_* jako zmiennej grupowania, dzięki czemu będzie traktowana w analizie jak zestaw danych MI. . Można również zdefiniować podziały dla innych zmiennych.

Z menu wybierz:

Dane > Podziel na podzbiory...

1. Wybierz opcję **Porównaj grupy**.
2. Wybierz wartość *Numer podstawienia [Imputation_]* jako zmienną do pogrupowania obserwacji.

Alternatywnie, po włączeniu oznaczeń (patrz poniżej) plik zostaje podzielony zgodnie z wartością *Numer podstawienia [Imputation_]*.

Rozróżnianie podstawionych wartości od obserwowanych wartości

Można rozróżnić podstawione wartości od obserwowanych po kolorze tła komórki, po czcionce i typie pogrubienia (w przypadku wartości podstawionych). Podczas tworzenia nowego zestawu danych w bieżącej sesji za pomocą opcji Uwzględnij braki danych, oznaczenia są włączane domyślnie. Po otwarciu zapisanego pliku danych, który zawiera podstawienia, oznaczenia zostają wyłączone.

Aby włączyć oznaczenia, z menu Okno edytora danych wybierz opcje:

Widok > Oznaczenie danych podstawionych...

Alternatywnie, można włączyć oznaczenia, klikając przycisk oznaczenia podstawień przy prawej krawędzi paska edytora w widoku danych w oknie edytora danych.

Przesunięcia pomiędzy podstawieniami

1. Wybierz z menu następującą opcję:

Edytuj > Przejdź do podstawiania...

2. Wybierz imputację (lub Dane źródłowe) z listy rozwijanej.

Można też wybrać imputację z listy rozwijanej paska edycji w Podglądzie danych Edytora danych.

Podczas wyboru imputacji względna pozycja imputacji zostaje zachowana. Jeśli na przykład w oryginalnym zbiorze danych znajduje się 1000 obserwacji, obserwacja 1034 (34 obserwacja w pierwszym podstawieniu) jest wyświetlana na górze siatki. Jeśli z menu rozwijanego wybrana zostanie wartość **2**, obserwacja 2034, 34. obserwacja 2. imputacji, zostanie wyświetlona w górnej części siatki. Po wybraniu w menu rozwijanym pozycji **Dane źródłowe** obserwacja 34 zostanie wyświetlona w górnej części siatki. Podczas nawigacji pomiędzy imputacjami zachowana jest również pozycja kolumny, aby ułatwić porównanie wartości pomiędzy imputacjami.

Przekształcanie i edytowanie podstawionych wartości

Niekiedy konieczne jest przekształcenie podstawionych danych. Przykładowo, można użyć dziennika ze wszystkimi wartościami zmiennej wynagrodzenia i zapisać wynik w nowej zmiennej. Wartość obliczona z użyciem podstawionych danych będzie traktowana jako podstawiona, jeśli różni się od wartości obliczonej z użyciem oryginalnych danych.

Jeśli edycja podstawionej wartości odbywa się w komórce edytora danych, ta komórka jest nadal traktowana jako podstawiona. Nie zaleca się edytowania podstawionych wartości w taki sposób.

Analizowanie danych wielokrotnych podstawień

Wiele procedur obsługuje zestawianie wyników z analizy wielowarstwowych podstawianych zestawów danych. Jeśli włączone są oznaczenia podstawiania, obok procedur obsługujących łączenie w pule wyświetlana jest specjalna ikona. W podmenu Statystyki opisowe menu Analiza na przykład procedury Tabela częstości, Statystyki opisowe, Eksploracja i Tabele krzyżowe obsługują łączenie w pule, natomiast procedury Ilorazowe, Wykresy prawdopodobieństwo-prawdopodobieństwo i Wykresy kwantyl-kwantyl nie.

W pule można łączyć wyniki tabelaryczne i model PMML. Nie ma żadnej nowej procedury tworzenia żądań wyników połączonych w pule; natomiast okno dialogowe Opcje umożliwia ogólną kontrolę wyników wielokrotnych podstawień.

- **Łączenie w pule wyników tabelarycznych.** Domyślnie, gdy uruchamiana jest obsługiwana procedura w zestawie danych MI (multiple imputation), wyniki są automatycznie tworzone dla każdego

podstawienia, pierwotne (nieokreślone) dane i zbiorcze (końcowe) wyniki, które uwzględniają różnice w imputacji. Statystyki łączone w pule różnią się w zależności od procedury.

- **Łączenie w pule PMML.** Możliwe jest również uzyskanie połączonych w pule PMML; umożliwiają to obsługiwane procedury, które obsługują eksport PMML. Zestaw PMML w puli jest wymagany w taki sam sposób, jak i jest składowany zamiast w puli PMML, który nie jest w puli.

Nieobsługiwane procedury tworzą wynik połączony w pule lub połączone w pule pliki PMML.

Poziomy łączenia w pule

Dane wyjściowe są umieszczane w puli przy użyciu jednego z dwóch poziomów:

- **Kombinacja naiwna.** Dostępny jest tylko parametr zebrany w puli.
- **Kombinacja jednej zmiennej.** Parametr zebrany w puli, jego błąd standardowy, test statystyczny i efektywne stopnie swobody, wartość p , przedział ufności i diagnostyka łączenia w pule (frakcja brakujących informacji, względna wydajność, względny wzrost wariancji) są wyświetlane, jeśli są dostępne.

Zwykle w pule łączone są współczynniki (regresji i korelacji), średnie (i różnice średnich) i liczebność. Jeśli jest dostępny standardowy błąd statystyki, używana jest pula univariate; w przeciwnym razie używana jest zestawianie naïve.

Procedury obsługujące łączenie w pule

Poniższe procedury obsługują zestawy danych MI, na poziomach zestawiania określonych dla każdego elementu danych wyjściowych.

Częstości. Obsługiwane są następujące funkcje:

- Tabela Statystyki obsługuje średnie w przypadku łączenia w pule z jedną zmienną (jeśli średnia błędu standardowego jest także wymagana) oraz N ważnych i brakujących N w przypadku naiwnego łączenia w pule.
- Tabela częstości obsługuje częstotliwość w przypadku naiwnego łączenia w pule.

Statystyki opisowe. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnie w przypadku łączenia w pule z jedną zmienną (jeśli średnia błędu standardowego jest także wymagana) i N ważnych w przypadku naiwnego łączenia w pule.

Tabele krzyżowe. Obsługiwane są następujące funkcje:

- Tabela krzyżowa obsługuje liczebność w przypadku naiwnego łączenia w pule.

Średnie. Obsługiwane są następujące funkcje:

- Tabela Raport obsługuje średnie w przypadku łączenia w pule z jedną zmienną (jeśli średnia błędu standardowego jest także wymagana) i N w przypadku naiwnego łączenia w pule.

Test t dla jednej próby. Obsługiwane są następujące funkcje:

- Tabela Statystyki obsługuje średnią w przypadku łączenia w pule z jedną zmienną oraz N w przypadku naiwnego łączenia w pule.
- Tabela Test obsługuje różnicę średniej w przypadku łączenia w pule z jedną zmienną.

Test t dla prób niezależnych. Obsługiwane są następujące funkcje:

- Tabela Statystyki dla grup obsługuje średnie w przypadku łączenia w pule z jedną zmienną i N w przypadku naiwnego łączenia w pule.
- Tabela Test obsługuje różnicę średniej w przypadku łączenia w pule z jedną zmienną.

Test t dla prób zależnych. Obsługiwane są następujące funkcje:

- Tabela Statystyki obsługuje średnie w przypadku łączenia w pule z jedną zmienną i N w przypadku naiwnego łączenia w pule.
- Tabela Korelacje obsługuje korelacje i N w przypadku naiwnego łączenia w pule.
- Tabela Test obsługuje średnią w przypadku łączenia w pule z jedną zmienną.

Jednoczynnikowa ANOVA. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnią w przypadku łączenia w pulę z jedną zmienną i N w przypadku naiwnego łączenia w pulę.
- Tabela Testy kontrastu obsługuje wartości kontrastu w przypadku łączenia w pulę z jedną zmienną.

GLM Univariate i Multivariate. Obsługiwane są następujące funkcje:

- Tabela Czynniki międzyobiektowe obsługuje N w puli Naive.
- Tabela Statystyki opisowe obsługuje średnią i N w puli Naive.
- Tabela Oszacowania parametrów obsługuje współczynnik, B, w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oszacowanie średnie brzegowe: Oszacowania obsługuje średnią w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oszacowania średnie brzegowe: Porównania parami obsługuje różnicę średniej w przypadku łączenia w pulę z jedną zmienną.

Liniowe modele mieszane. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnią i N w przypadku naiwnego łączenia w pulę.
- Tabela Oceny parametrów efektów stałych obsługuje oszacowanie w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oceny parametrów kowariancji obsługuje oszacowanie w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oszacowanie średnie brzegowe: Oszacowania obsługuje średnią w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oszacowania średnie brzegowe: Porównania parami obsługuje różnicę średniej w przypadku łączenia w pulę z jedną zmienną.

Uogólnione modele liniowe i Uogólnione równania estymujące. Te procedury obsługują PMML zebrane w puli.

- Tabela Informacja o zmiennych jakościowych obsługuje N i procenty w przypadku naiwnego łączenia w pulę.
- Tabela Informacja o zmiennej ciągłej obsługuje N i średnią w przypadku naiwnego łączenia w pulę.
- Tabela Oszacowania parametrów obsługuje współczynnik, B, w przypadku łączenia w pulę z jedną zmienną.
- Oszacowane średnie brzegowe: Współczynniki estymacji obsługuje średnią w przypadku naiwnego łączenia w pulę.
- Tabela Oszacowanie średnie brzegowe: Oszacowania obsługuje średnią w przypadku łączenia w pulę z jedną zmienną.
- Tabela Oszacowania średnie brzegowe: Porównania parami obsługuje różnicę średniej w przypadku łączenia w pulę z jedną zmienną.

Korelacje parami. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnią i N w przypadku naiwnego łączenia w pulę.
- Tabela Korelacje obsługuje korelacje i N w przypadku łączenia w pulę z jedną zmienną. Należy zauważyć, że korelacje są transformowane przy użyciu transformacji z Fishera przed zestawieniem, a następnie z powrotem po zestawieniu.

Korelacje cząstkowe. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnią i N w przypadku naiwnego łączenia w pulę.
- Tabela Korelacje obsługuje korelacje w przypadku naiwnego łączenia w pulę.

Regresja liniowa. Ta procedura obsługuje PMML zebrane w pulę.

- Tabela Statystyki opisowe obsługuje średnią i N w przypadku naiwnego łączenia w pulę.

- Tabela Korelacje obsługuje korelacje i N w przypadku naiwnego łączenia w pule.
- Tabela Współczynniki obsługuje B w przypadku łączenia w pule z jedną zmienną i korelacje w przypadku naiwnego łączenia w pule.
- Tabela Współczynniki korelacji obsługuje korelacje w przypadku naiwnego łączenia w pule.
- Tabela Statystyki reszt obsługuje średnia i N w przypadku naiwnego łączenia w pule.

Dwumianowa regresja logistyczna. Ta procedura obsługuje PMML zebrane w pule.

- Tabela Zmienne w modelu obsługuje B w przypadku łączenia w pule z jedną zmienną.

Wielomianowa regresja logistyczna. Ta procedura obsługuje PMML zebrane w pule.

- Tabela Oszacowania parametrów obsługuje współczynnik, B, w przypadku łączenia w pule z jedną zmienną.

Regresja porządkowa. Obsługiwane są następujące funkcje:

- Tabela Oszacowania parametrów obsługuje współczynnik, B, w przypadku łączenia w pule z jedną zmienną.

Analiza dyskryminacyjna. Ta procedura obsługuje model XML zebrany w puli.

- Tabela Statystyki dla grup obsługuje średnią i N ważnych w przypadku naiwnego łączenia w pule.
- Tabela Macierze potężonych wewnątrzgrupowych obsługuje korelacje w przypadku naiwnego łączenia w pule.
- Tabela Współczynniki kanonicznych funkcji dyskryminacyjnych obsługuje współczynniki niestandardyzowane w przypadku naiwnego łączenia w pule.
- Tabela Funkcje w środkach ciężkości grup obsługuje współczynniki niestandardyzowane w przypadku naiwnego łączenia w pule.
- Tabela Współczynniki funkcji klasyfikacyjnej obsługuje współczynniki w przypadku naiwnego łączenia w pule.

Test chi-kwadrat. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnią i N w przypadku naiwnego łączenia w pule.
- Tabela częstości obsługuje obserwowane N w przypadku naiwnego łączenia w pule.

Test dwumianowy. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnie i N w przypadku naiwnego łączenia w pule.
- Tabela Test obsługuje N, obserwowane proporcje i test podziału w przypadku naiwnego łączenia w pule.

Test serii. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnie i N w przypadku naiwnego łączenia w pule.

Test Kołmogorowa-Smirnowa dla jednej próby. Obsługiwane są następujące funkcje:

- Tabela Statystyki opisowe obsługuje średnie i N w przypadku naiwnego łączenia w pule.

Testy dla dwóch prób niezależnych. Obsługiwane są następujące funkcje:

- Tabela Rangów obsługuje średnią rangę i N w przypadku naiwnego łączenia w pule.
- Tabela częstości obsługuje N w przypadku naiwnego łączenia w pule.

Testy dla kilku prób niezależnych. Obsługiwane są następujące funkcje:

- Tabela Rangów obsługuje średnią rangę i N w przypadku naiwnego łączenia w pule.
- Tabela częstości obsługuje liczebności w przypadku naiwnego łączenia w pule.

Testy dla dwóch prób zależnych. Obsługiwane są następujące funkcje:

- Tabela Rangów obsługuje średnią rangę i N w przypadku naiwnego łączenia w pule.
- Tabela częstości obsługuje N w przypadku naiwnego łączenia w pule.

Testy dla kilku prób zależnych. Obsługiwane są następujące funkcje:

- Tabela Rangii obsługuje średnią rangę w przypadku naiwnego łączenia w pule.

Regresja Coxa. Ta procedura obsługuje PMML zebrane w pule.

- Tabela Zmienne w modelu obsługuje B w przypadku łączenia w pule z jedną zmienną.
- Tabela Średnie współzmiennych obsługuje średnie w przypadku naiwnego łączenia w pule.

Proporcje jedn. próbne. Obsługiwane są następujące funkcje:

- Tabela Przedziały ufności dla jednej próby dla proporcji między próbkami obsługuje wartość Profragment w zestawach Univariate.
- Tabela Test Proporcji dla jednej próby obsługuje wartość Obserwowane-Wartość testowa w zestawianych pulach Univariate.

Proporcje dla prób niezależnych. Obsługiwane są następujące funkcje:

- Tabela Statystyki grup próbnych dla prób niezależnych obsługuje wartość Profragment w zestawionym zestawionym poziomie.
- Tabela Przedziały ufności dla prób niezależnych obsługuje różnicę w proporcjach w zestawach Naïve.
- Tabela Test Proporcji dla prób niezależnych obsługuje różnicę w Procjach w zestawianych pulach Univariate.

Proporcje dla prób zależnych. Obsługiwane są następujące funkcje:

- Tabela Statystyki proporcje dla prób zależnych (Paired-Samples Proporcji Statistics) obsługuje wartość Profragment w puli Univariate.
- Tabela przedziały ufności dla prób zależnych obsługuje różnicę w proporcjach w zestawieniu nie wcześniej niż w przypadku naïve.
- Tabela Test Proporcji dla prób zależnych (Paired-Samples Test table) obsługuje różnicę w Procjach w zestawianych pulach Univariate.

Wiele opcji podstawiania

Karta Podstawiania wielokrotnie umożliwia kontrolowanie dwóch rodzajów preferencji związanych z podstawieniami wielokrotnymi.

Oznaczenie podstawianych danych. Domyślnie komórki zawierające podstawiane dane będą miały inny kolor niż komórki zawierające dane niepodstawiane. Charakterystyczny wygląd podstawianych danych powinien ułatwić przewijanie zbioru danych w celu lokalizacji takich komórek. Możliwa jest zmiana domyślnego koloru tła komórki i czcionek oraz wyświetlanie danych podstawianych przy użyciu czcionki pogrubionej.

Wyniki analizy. Ta grupa kontroluje wyniki Edytora raportów, utworzone podczas analizy zestawu zbioru wielokrotnie podstawianych. Domyślnie wynik zostanie wygenerowany w oryginalnym zbiorze danych (przed podstawieniem) dla każdego zbioru danych podstawianych. Dodatkowo dla procedur obsługujących łączenie danych podstawianych zostaną wygenerowane końcowe wyniki kombinowane. W przypadku wykonywania łączenia jednej zmiennej wyświetlana jest również diagnostyka łączenia. Można jednak wyłączyć zbędne wyniki.

W celu ustawienia opcji Imputacji wielokrotnych

Wybierz z menu następujące opcje:

Edytuj > Opcje

Kliknij kartę Imputacje wielokrotne.

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 EM
 - w Analizie braków danych [5](#)
- Analiza braków danych
 - Algorytm EM [5](#)
 - dotatkowe właściwości komendy [7](#)
 - maksymalizacja oczekiwań [6](#)
 - metody [5](#)
 - podstawianie braków danych [5](#)
 - regresja [6](#)
 - statystyki opisowe [4](#)
 - szacowanie statystyk [5](#)
 - test MCAR [5](#)
 - układy [3](#)
- Analiza układów [9](#)

B

- brakujące wartości
 - statystyki jednej zmiennej [4](#)
- brakujące zmienne wskaźnikowe
 - w Analizie braków danych [4](#)

C

- całkowicie warunkowa specyfikacja
 - w podstawieniach wielokrotnych [10](#)

K

- kategorie w tabelach
 - w Analizie braków danych [4](#)
- korelacje
 - w Analizie braków danych [5](#), [6](#)
- kowariancja
 - w Analizie braków danych [5](#), [6](#)

L

- liczebności wartości skrajnych
 - w Analizie braków danych [4](#)

N

- niekompletne dane
 - patrz Analiza braków danych [1](#)
- niezgodność
 - w Analizie braków danych [4](#)

O

- odchylenie standardowe
 - w Analizie braków danych [4](#)

P

- Podstaw brakujące wartości danych
 - wyniki [13](#)
- Podstawianie brakujących wartości danych
 - metoda podstawiania [10](#)
 - ograniczenia [12](#)
- podstawianie monotoniczne
 - w podstawieniach wielokrotnych [10](#)
- podstawienia wielokrotne
 - analiza układów [9](#)
 - podstawianie brakujących wartości [9](#)
- przebieg iteracji
 - w podstawieniach wielokrotnych [13](#)

R

- regresja
 - w Analizie braków danych [6](#)
- reszty
 - w Analizie braków danych [6](#)

S

- sortowanie obserwacji
 - w Analizie braków danych [3](#)

Ś

- średnia
 - w Analizie braków danych [4–6](#)

T

- tabele częstości
 - w Analizie braków danych [4](#)
- tabele obserwacji
 - w Analizie braków danych [3](#)
- test MCAR
 - w Analizie braków danych [1](#)
- test MCAR Little'a
 - w Analizie braków danych [1](#)
- test t
 - w Analizie braków danych [4](#)
- test t Studenta
 - w Analizie braków danych [6](#)

U

- usuwanie obserwacjami
 - w Analizie braków danych [1](#)
- usuwanie parami
 - w Analizie braków danych [1](#)

W

Wielokrotne podstawienia [13](#), [14](#)

Z

zmiennie losowe normalne

w Analizie braków danych [6](#)

zmiennie wskaźnikowe

w Analizie braków danych [4](#)

