

*IBM SPSS Statistics 27 – skrócony  
podręcznik*



**Uwaga**

Przed skorzystaniem z niniejszych informacji oraz produktu, którego one dotyczą, należy zapoznać się z informacjami zamieszczonymi w sekcji [“Informacje” na stronie 87](#).

**Informacje o produkcie**

Niniejsza edycja dotyczy wersji 27, wydania 0, modyfikacji 0 produktu IBM® SPSS Statistics i wszystkich jego późniejszych wersji i modyfikacji, o ile nie zaznaczono inaczej w nowych wydaniach.

© **Copyright International Business Machines Corporation .**

---

# Spis treści

|  |           |
|--|-----------|
| <b>Rozdział 1. Wstęp.....</b>  | <b>1</b>  |
| Pliki przykładowe.....   | 1         |
| Otwieranie pliku danych.....   | 1         |
| Uruchamianie analizy .....   | 2         |
| Tworzenie wykresów.....  | 4         |
| <b>Rozdział 2. Odczytywanie danych.....</b>  | <b>7</b>  |
| Podstawowa struktura plików danych programu IBM SPSS Statistics.....                     | 7         |
| Odczytywanie plików danych programu IBM SPSS Statistics.....                             | 7         |
| Odczytywanie danych programu Excel.....  | 8         |
| Odczytywanie danych z bazy danych.....   | 11        |
| Odczytywanie danych z pliku tekstowego.....  | 14        |
| <b>Rozdział 3. Korzystanie z Edytora danych.....</b>                                     | <b>19</b> |
| Wprowadzanie danych liczbowych.....  | 19        |
| Wprowadzanie danych łańcuchowych.....  | 20        |
| Definiowanie danych.....   | 21        |
| Dodawanie etykiet zmiennych.....   | 21        |
| Zmiana typu i formatu zmiennej.....  | 22        |
| Dodawanie etykiet wartości.....  | 22        |
| Obsługa braków danych.....   | 23        |
| Brakujące wartości zmiennej liczbowej.....   | 23        |
| Brakujące wartości zmiennej łańcuchowej.....   | 24        |
| <b>Rozdział 4. Interpretacja statystyk podsumowujących poszczególnych zmiennych.....</b> | <b>25</b> |
| Poziom pomiaru.....  | 25        |
| Miary podsumowujące danych kategoryjnych.....  | 25        |
| Wykresy dla danych kategoryjnych.....  | 26        |
| Miary podsumowujące zmiennych ilościowych.....   | 27        |
| Histogramy zmiennych ilościowych.....  | 28        |
| <b>Rozdział 5. Tworzenie i edytowanie wykresów.....</b>                                  | <b>31</b> |
| Podstawowe informacje o tworzeniu wykresów.....  | 31        |
| Korzystanie z galerii kreatora wykresów.....   | 31        |
| Definiowanie zmiennych i statystyk.....  | 32        |
| Dodawanie tekstu.....  | 33        |
| Tworzenie wykresu.....   | 34        |
| <b>Rozdział 6. Praca z oknem raportu.....</b>  | <b>37</b> |
| Korzystanie z Edytora raportów.....  | 37        |
| Korzystanie z Edytora tabel przestawnych.....  | 38        |
| Ocena definicji wyników.....   | 38        |
| Przestawianie tabel.....   | 39        |
| Tworzenie i wyświetlanie warstw.....   | 40        |
| Edytowanie tabel.....  | 40        |
| Ukrywanie wierszy i kolumn.....  | 41        |
| Zmiana formatów wyświetlania danych.....   | 42        |
| Szablony TableLook.....  | 43        |

|   |           |
|---|-----------|
| Korzystanie z formatów predefiniowanych.....                                      | 43        |
| Dostosowywanie stylów TableLook.....  | 43        |
| Zmiana domyślnych formatów tabel.....   | 46        |
| Dostosowywanie początkowych ustawień wyświetlania.....                            | 46        |
| Wyświetlanie etykiet zmiennych i wartości.....                                    | 47        |
| Korzystanie z wyników w innych aplikacjach.....                                   | 48        |
| Wklejanie wyników jako tabel do programu Word.....                                | 49        |
| Wklejanie wyników jako tekstu.....  | 49        |
| Eksportowanie wyników do plików programów Microsoft Word, PowerPoint i Excel..... | 50        |
| Eksportowanie wyników do pliku PDF.....   | 55        |
| Eksportowanie wyników do pliku HTML.....  | 57        |
| <b>Rozdział 7. Praca z komendami.....</b>   | <b>59</b> |
| Wklejanie komend.....   | 59        |
| Edytowanie komend.....  | 60        |
| Otwieranie i uruchamianie pliku komend.....                                       | 61        |
| Korzystanie z punktów zatrzymania.....  | 61        |
| <b>Rozdział 8. Modyfikowanie wartości danych.....</b>                             | <b>63</b> |
| Tworzenie zmiennej kategoryjnej na podstawie zmiennej ilościowej.....             | 63        |
| Obliczanie zmiennych wyników.....   | 65        |
| Używanie funkcji w wyrażeniach.....   | 66        |
| Korzystanie z wyrażień warunkowych.....   | 67        |
| Praca z datami i godzinami.....   | 68        |
| Obliczanie długości okresu między dwiema datami.....                              | 68        |
| Dodawanie czasu trwania do daty.....  | 69        |
| <b>Rozdział 9. Sortowanie i wybieranie danych.....</b>                            | <b>71</b> |
| Sortowanie danych.....  | 71        |
| Przetwarzanie pliku dzielonego.....   | 71        |
| Sortowanie obserwacji przy przetwarzaniu pliku dzielonego.....                    | 73        |
| Włączanie i wyłączanie przetwarzania pliku dzielonego.....                        | 73        |
| Wybieranie podzbiorów obserwacji.....   | 73        |
| Wybieranie obserwacji na podstawie wyrażenia warunkowego.....                     | 74        |
| Wybieranie próby losowej.....   | 75        |
| Wybieranie zakresu czasu lub zakresu obserwacji.....                              | 75        |
| Postępowanie z niewybranymi obserwacjami.....                                     | 76        |
| Status wyboru obserwacji.....   | 76        |
| <b>Rozdział 10. Pliki przykładowe.....</b>  | <b>77</b> |
| <b>Informacje.....</b>  | <b>87</b> |
| Znaki towarowe.....   | 88        |
| <b>Indeks.....</b>  | <b>91</b> |

---

# Rozdział 1. Wstęp

W tym podręczniku wyjaśniono, jak korzystać z wielu dostępnych funkcji produktu. Zawiera on praktyczne instrukcje rozpisane na kroki. Wszystkie pliki, o których mowa w przykładach, są instalowane razem z aplikacją, można zatem samodzielnie realizować przykłady, wykonywać te same analizy i uzyskiwać te same wyniki, jakie opisano w podręczniku.

Aby zapoznać się ze szczegółowymi przykładami zastosowania różnych technik analizy statystycznej, można skorzystać ze Studiów przypadków dostępnych w menu Pomoc.

## Pliki przykładowe

---

W większości omawianych tutaj przykładów używamy pliku danych *demo.sav*. Zawiera on dane z fikcyjnego badania ankietowego kilku tysięcy osób, obejmujące podstawowe informacje demograficzne i konsumenne.

Jeśli korzystasz z wersji Student, wersja pliku *demo.sav* jest reprezentatywną próbą pierwotnego pliku danych, zmniejszoną tak, aby mieściła się w limicie 1500 obserwacji. Wyniki uzyskane przy użyciu tego pliku danych będą różnić się od wyników przedstawionych w tej dokumentacji.

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.

## Otwieranie pliku danych

---

Aby otworzyć plik danych:

1. Z menu wybierz:

**Plik > Otwórz > Dane...**

Zostanie wyświetlone okno dialogowe służące do otwierania plików.

Domyślnie wyświetlane są pliki danych programu IBM SPSS Statistics (z rozszerzeniem *.sav*).

W tym przykładzie używany jest plik *demo.sav*.

|    | age | marital | address | income | inccat | car |
|----|-----|---------|---------|--------|--------|-----|
| 1  | 55  | 1       | 12      | 72.00  | 3.00   | 36. |
| 2  | 56  | 0       | 29      | 153.00 | 4.00   | 76. |
| 3  | 28  | 1       | 9       | 28.00  | 2.00   | 13. |
| 4  | 24  | 1       | 4       | 26.00  | 2.00   | 12. |
| 5  | 25  | 0       | 2       | 23.00  | 1.00   | 11. |
| 6  | 45  | 1       | 9       | 76.00  | 4.00   | 37. |
| 7  | 42  | 0       | 19      | 40.00  | 2.00   | 19. |
| 8  | 35  | 0       | 15      | 57.00  | 3.00   | 28. |
| 9  | 46  | 0       | 26      | 24.00  | 1.00   | 12. |
| 10 | 34  | 1       | 0       | 89.00  | 4.00   | 46. |
| 11 | 55  | 1       | 17      | 72.00  | 3.00   | 35. |

Rysunek 1. Plik demo.sav w Edytorze danych

Plik danych zostanie wyświetlony w Edytorze danych. W widoku danych umieszczenie kursora myszy na nazwie zmiennej (nagłówku kolumny) powoduje wyświetlenie bardziej opisowej etykiety zmiennej (jeśli etykieta tej zmiennej została zdefiniowana).

Domyślnie wyświetlane są rzeczywiste wartości danych. Aby wyświetlić etykiety:

2. Z menu wybierz:

**Widok > Etykiety wartości**



Rysunek 2. Przycisk Etykiety wartości

Zamiast tego można użyć przycisku Etykiety wartości na pasku narzędzi.

|    | age | marital   | address | income | inccat      | car |
|----|-----|-----------|---------|--------|-------------|-----|
| 1  | 55  | Married   | 12      | 72.00  | \$50 - \$74 | 36. |
| 2  | 56  | Unmarried | 29      | 153.00 | \$75+       | 76. |
| 3  | 28  | Married   | 9       | 28.00  | \$25 - \$49 | 13. |
| 4  | 24  | Married   | 4       | 26.00  | \$25 - \$49 | 12. |
| 5  | 25  | Unmarried | 2       | 23.00  | Under \$25  | 11. |
| 6  | 45  | Married   | 9       | 76.00  | \$75+       | 37. |
| 7  | 42  | Unmarried | 19      | 40.00  | \$25 - \$49 | 19. |
| 8  | 35  | Unmarried | 15      | 57.00  | \$50 - \$74 | 28. |
| 9  | 46  | Unmarried | 26      | 24.00  | Under \$25  | 12. |
| 10 | 34  | Married   | 0       | 89.00  | \$75+       | 46. |
| 11 | 55  | Married   | 17      | 72.00  | \$50 - \$74 | 35. |

Rysunek 3. Etykiety wartości wyświetlane w Edytorze danych

Teraz wyświetlane są opisowe etykiety wartości ułatwiające interpretację odpowiedzi.

## Uruchamianie analizy

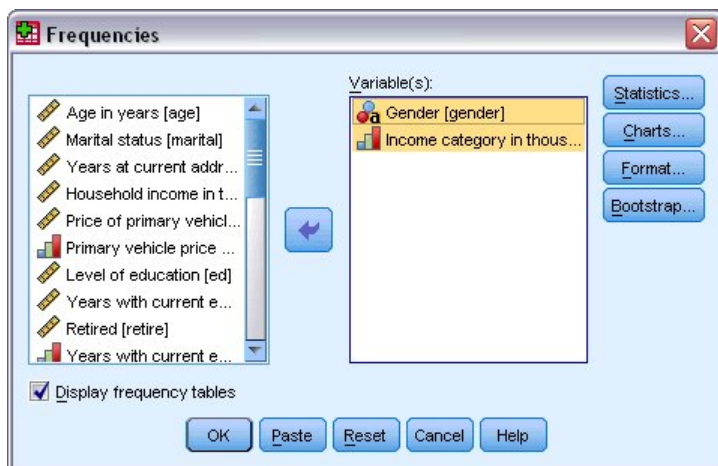
Jeśli zainstalowane są dodatki opcjonalne, menu Analiza zawiera listę kategorii raportowania i analiz statystycznych.

Rozpocniemy od utworzenia prostej tabeli częstości (tabeli liczebności). W tym przykładzie wymagana jest opcja Statistics Base.

1. Z menu wybierz:

**Analiza > Statystyki opisowe > Częstości...**

Zostanie wyświetlone okno dialogowe Częstości.



Rysunek 4. Okno dialogowe Częstości

Ikona obok każdej zmiennej informuje o jej typie danych i poziomie pomiaru.

|                     | Liczba | Łańcuch | Data | Czas |
|---------------------|--------|---------|------|------|
| Zmienna (ilościowa) |        | n/a     |      |      |
| Porządkowa          |        |         |      |      |
| Nominalne           |        |         |      |      |

Jeśli etykieta zmiennej i/lub nazwa jest obcięta na liście, to pełna etykieta/nazwa zostanie wyświetlona po wskazaniu kursorem. Po opisowej etykiecie zmiennej wyświetlana jest nazwa zmiennej *inccat* w nawiasach kwadratowych. *Income category in thousands* to etykieta zmiennej. Gdyby nie było etykiety zmiennej, w polu listy widoczna byłaby tylko nazwa zmiennej.

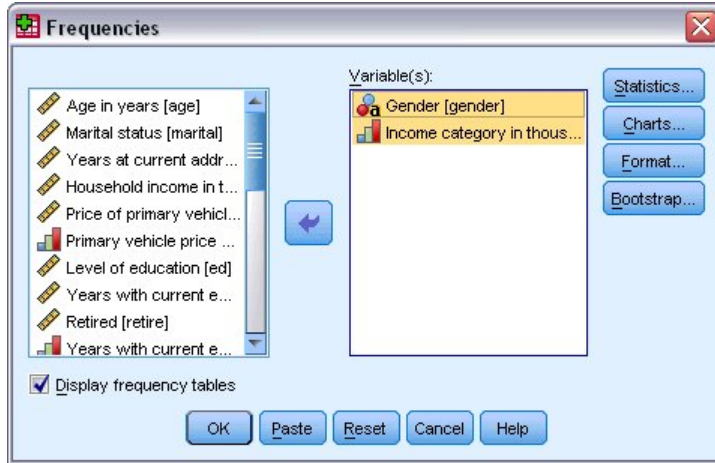
Rozmiar okien dialogowych można zmieniać podobnie jak okien, klikając i przeciągając ich zewnętrzne brzozy lub narożniki. Poszerzenie okna dialogowego spowoduje na przykład poszerzenie list zmiennych.

W oknie dialogowym wybierasz zmienne, które mają być analizowane, z listy źródłowej po lewej stronie, a następnie przeciągasz je na listę Zmienne po prawej stronie. Przycisk **OK**, który uruchamia analizę, jest nieaktywny, dopóki co najmniej jedna zmienna nie zostanie umieszczona na liście Zmienne.

W wielu oknach dialogowych można uzyskać dodatkowe informacje, klikając prawym przyciskiem myszy dowolną nazwą zmiennej na liście i wybierając opcję **Informacje o zmiennej** z menu podręcznego.

2. Kliknij zmienną *Gender [gender]* na liście zmiennych źródłowych i przeciągnij ją na listę zmiennych docelowych.

- Kliknij zmienną *Income category in thousands [inccat]* na liście źródłowej i przeciągnij ją na listę docelową.



Rysunek 5. Zmienne wybrane do analizy

- Kliknij przycisk **OK**, aby uruchomić procedurę.

Wyniki są wyświetlane w oknie raportów.

|              | Frequency | Percent | Valid Percent | Cumul: Perce |
|--------------|-----------|---------|---------------|--------------|
| Valid Female | 3179      | 49.7    | 49.7          |              |
| Male         | 3221      | 50.3    | 50.3          |              |
| Total        | 6400      | 100.0   | 100.0         |              |

|                  | Frequency | Percent | Valid Percent | Cum Pe |
|------------------|-----------|---------|---------------|--------|
| Valid Under \$25 | 1174      | 18.3    | 18.3          |        |
| \$25 - \$49      | 2388      | 37.3    | 37.3          |        |
| \$50 - \$74      | 1120      | 17.5    | 17.5          |        |
| \$75+            | 1718      | 26.8    | 26.8          |        |
| Total            | 6400      | 100.0   | 100.0         |        |

Rysunek 6. Tabela częstości kategorii dochodów

## Tworzenie wykresów

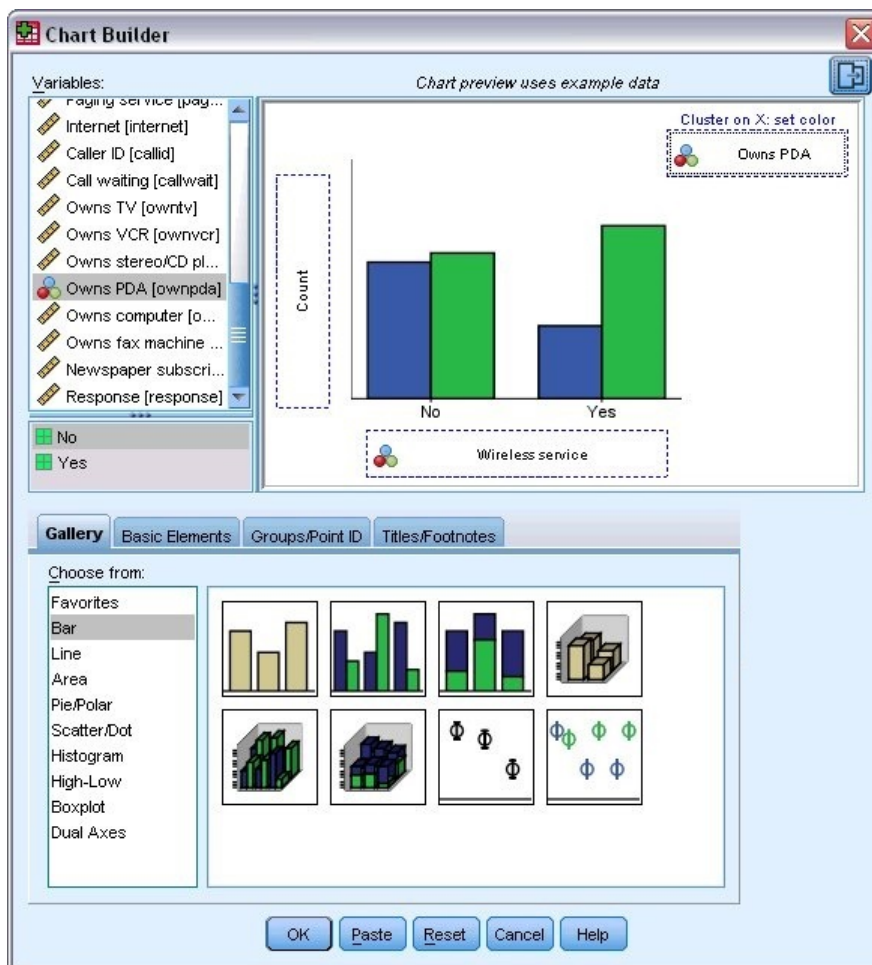
Mimo że niektóre procedury statystyczne tworzą wykresy, można także użyć do tego celu menu Wykresy.

Można na przykład utworzyć wykres przedstawiający zależność między korzystaniem z telefonii komórkowej a posiadaniem urządzeń PDA (osobisty asystent cyfrowy).

- Z menu wybierz:

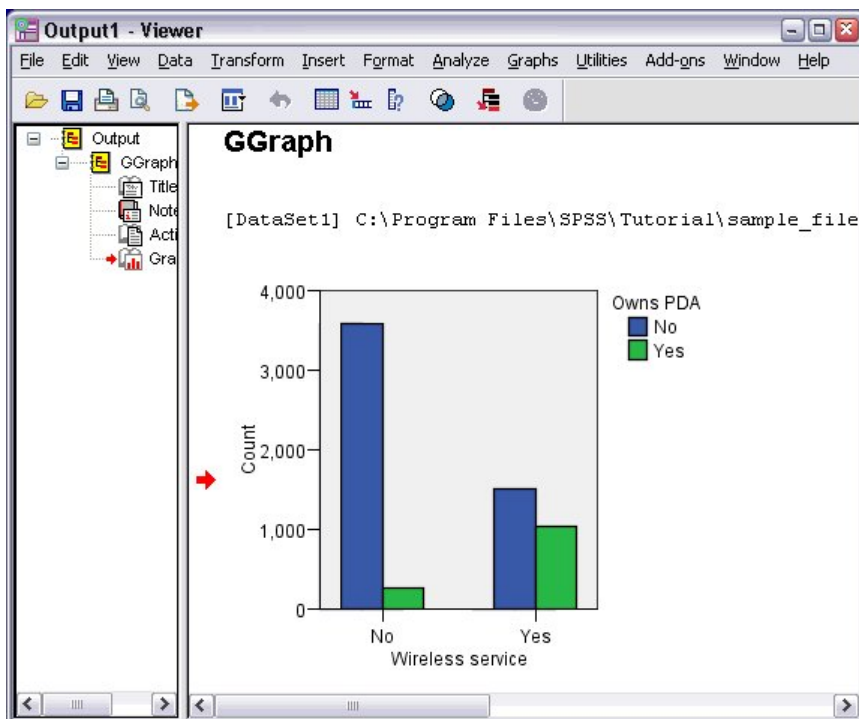
**Wykresy > Kreator wykresów...**





Rysunek 7. Okno dialogowe Kreator wykresów z wypełnionymi strefami upuszczania

2. Kliknij kartę **Galeria** (jeśli nie jest jeszcze wybrana).
3. Kliknij typ **Słupkowy** (jeśli nie jest jeszcze wybrany).
4. Przeciągnij ikonę Zgrupowany słupkowy do obszaru roboczego (dużego obszaru nad Galeria).
5. Przewiń listę zmiennych, kliknij prawym przyciskiem myszy zmienną *Wireless service [wireless]*, a następnie wybierz opcję **Nominalne** jako jej poziom pomiaru.
6. Przeciągnij zmienną *Wireless service [wireless]* do osi *x*.
7. Kliknij prawym przyciskiem myszy zmienną *Owns PDA [ownpda]* i wybierz opcję **Nominalne** jako jej poziom pomiaru.
8. Przeciągnij zmienną *Owns PDA [ownpda]* do strefy upuszczania grupy w prawym górnym rogu obszaru roboczego.
9. Kliknij przycisk **OK**, aby utworzyć wykres.



Rysunek 8. Wykres słupkowy wyświetlany w oknie raportu

Wykres słupkowy zostanie wyświetlony w oknie raportu. Wykres pokazuje, że posiadanie PDA przez osoby korzystające z sieci komórkowej jest bardziej prawdopodobne niż posiadanie PDA przez osoby niekorzystające z sieci komórkowej.

Można edytować wykresy i tabele, klikając je dwukrotnie w panelu zawartości okna raportu. Można także kopiować i wklejać wyniki do innych aplikacji. Tematy te zostaną omówione później.

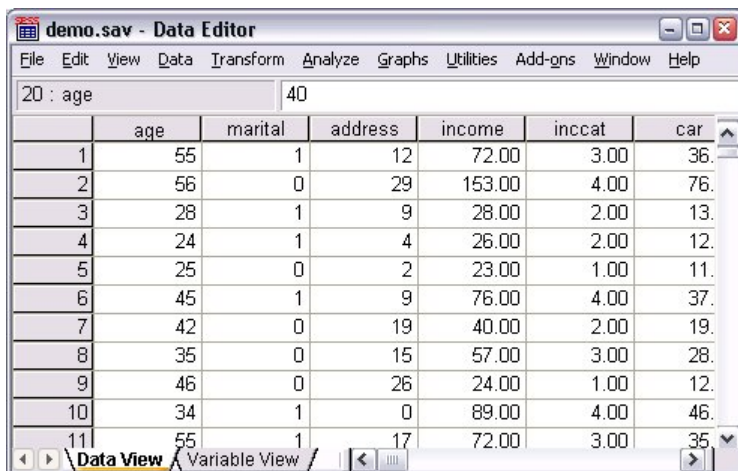
---

## Rozdział 2. Odczytywanie danych

Dane można wprowadzać bezpośrednio lub importować z wielu różnych źródeł. Procesy odczytu danych zapisanych w plikach danych programu IBM SPSS Statistics, arkuszach kalkulacyjnych, takich jak Microsoft Excel, aplikacjach bazodanowych, takich jak Microsoft Access, oraz plikach tekstowych, zostały omówione w niniejszym rozdziale.

### Podstawowa struktura plików danych programu IBM SPSS Statistics

---



|    | age | marital | address | income | inccat | car |
|----|-----|---------|---------|--------|--------|-----|
| 1  | 55  | 1       | 12      | 72.00  | 3.00   | 36. |
| 2  | 56  | 0       | 29      | 153.00 | 4.00   | 76. |
| 3  | 28  | 1       | 9       | 28.00  | 2.00   | 13. |
| 4  | 24  | 1       | 4       | 26.00  | 2.00   | 12. |
| 5  | 25  | 0       | 2       | 23.00  | 1.00   | 11. |
| 6  | 45  | 1       | 9       | 76.00  | 4.00   | 37. |
| 7  | 42  | 0       | 19      | 40.00  | 2.00   | 19. |
| 8  | 35  | 0       | 15      | 57.00  | 3.00   | 28. |
| 9  | 46  | 0       | 26      | 24.00  | 1.00   | 12. |
| 10 | 34  | 1       | 0       | 89.00  | 4.00   | 46. |
| 11 | 55  | 1       | 17      | 72.00  | 3.00   | 35. |

Rysunek 9. Edytor danych

Pliki danych programu IBM SPSS Statistics są zorganizowane w obserwacje (wiersze) i zmienne (kolumny). W tym pliku danych obserwacje odzwierciedlają respondentów ankiety. Zmienne odzwierciedlają odpowiedzi na pytania zadawane w ankiecie.

### Odczytywanie plików danych programu IBM SPSS Statistics

---

Pliki danych programu IBM SPSS Statistics, które mają rozszerzenie `.sav`, zawierają zapisane dane.

1. Z menu wybierz:

**Plik > Otwórz > Dane...**

2. Przejdź do pliku `demo.sav` i otwórz go. Więcej informacji zawiera temat [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

Dane są teraz wyświetlane w Edytorze danych.

demo.sav - Data Editor

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

20 : age 40

|    | age | marital | address | income | inccat | car |
|----|-----|---------|---------|--------|--------|-----|
| 1  | 55  | 1       | 12      | 72.00  | 3.00   | 36. |
| 2  | 56  | 0       | 29      | 153.00 | 4.00   | 76. |
| 3  | 28  | 1       | 9       | 28.00  | 2.00   | 13. |
| 4  | 24  | 1       | 4       | 26.00  | 2.00   | 12. |
| 5  | 25  | 0       | 2       | 23.00  | 1.00   | 11. |
| 6  | 45  | 1       | 9       | 76.00  | 4.00   | 37. |
| 7  | 42  | 0       | 19      | 40.00  | 2.00   | 19. |
| 8  | 35  | 0       | 15      | 57.00  | 3.00   | 28. |
| 9  | 46  | 0       | 26      | 24.00  | 1.00   | 12. |
| 10 | 34  | 1       | 0       | 89.00  | 4.00   | 46. |
| 11 | 55  | 1       | 17      | 72.00  | 3.00   | 35. |

Data View Variable View

Rysunek 10. Otwarty plik danych

## Odczytywanie danych programu Excel

Zamiast wpisywać wszystkie dane bezpośrednio do Edytora danych, można odczytać dane z aplikacji, takiej jak Microsoft Excel. Można także odczytywać nagłówki kolumn jako nazwy zmiennych.

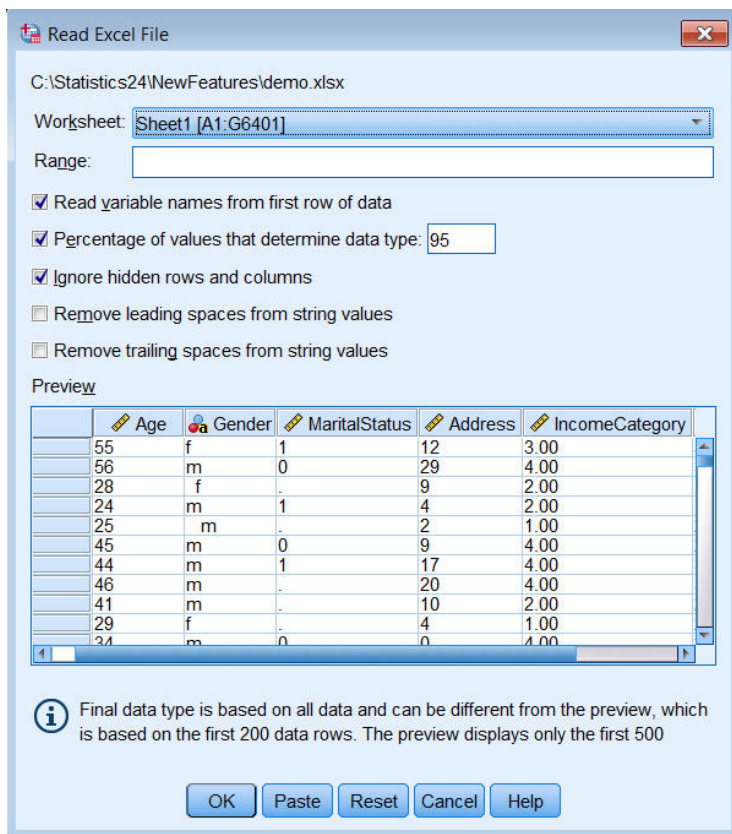
1. Z menu wybierz:

**Plik > Importuj dane > Excel**

2. Przejdź do folderu Samples\English i wybierz plik demo.xlsx.

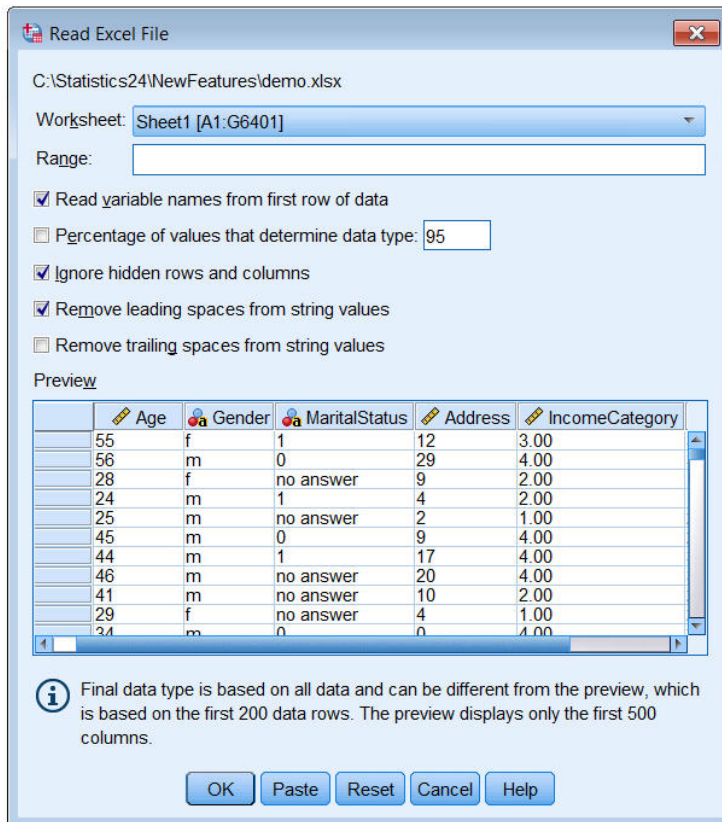
W oknie dialogowym **Odczytaj plik Excel** zostanie wyświetlony podgląd pliku danych. Zostanie wyświetlona zawartość pierwszego arkusza w pliku. Jeśli plik zawiera więcej niż jeden arkusz, można wybrać arkusz listy.

Można zauważyć, że niektóre wartości łańcuchowe zmiennej *Gender* mają początkowe spacje. Niektóre wartości zmiennej *MaritalStatus* są wyświetlane jako kropki (.).



Rysunek 11. Okno dialogowe odczytu pliku programu Excel

3. Upewnij się, że zaznaczona jest opcja **Odczytaj nazwy zmiennych z pierwszego wiersza danych**. Jeśli nagłówki kolumn nie są zgodne z regułami tworzenia nazw zmiennych, zostaną przekształcone w poprawne nazwy zmiennych. Pierwotne nagłówki kolumn zostaną zachowane jako etykiety zmiennych.
4. Wybierz opcję **Usuń początkowe spacje z wartości łańcuchowych**.
5. Usuń zaznaczenie opcji **Odsetek wartości określających typ danych**.



Teraz w komórkach, które były systemowymi brakami danych, wyświetlana jest wartość łańcuchowa „no answer”. Jeśli nie ma parametru odsetka wartości, a kolumna zawiera dane należące do różnych typów, zmienna jest odczytywana jako łańcuchowa. Wszystkie wartości są zachowywane, ale wartości liczbowe są traktowane jako łańcuchy.

6. Wybierz (zaznacz) opcję **Odsetek wartości określających typ danych**, aby traktować *MaritalStatus* jako zmienną liczbową.

7. Kliknij przycisk **OK**, aby odczytać plik programu Excel.

Dane są teraz wyświetlane w Edytorze danych, a nagłówki kolumn są używane jako nazwy zmiennych. Ponieważ, nazwy zmiennych nie mogą zawierać spacji, spacje z pierwotnych nagłówków zostały usunięte. Na przykład nagłówek kolumny „Marital Status” został przekształcony w zmienną *MaritalStatus*. Pierwotny nagłówek kolumny jest zachowany jako etykieta zmiennej.

\*Untitled2 [DataSet1] - IBM SPSS Statistics Data Editor

File Edit View Data Transform Analyze Direct Marketing Graphs Utilities Extensions Window Help

Visible: 6 of 6 Variables

|    | Age | Gender | MaritalStatus | Address | IncomeCategory | JobCategory |
|----|-----|--------|---------------|---------|----------------|-------------|
| 1  | 55  | f      | 1             | 12      | 3.00           | 3           |
| 2  | 56  | m      | 0             | 29      | 4.00           | 3           |
| 3  | 28  | f      | .             | 9       | 2.00           | 1           |
| 4  | 24  | m      | 1             | 4       | 2.00           | 1           |
| 5  | 25  | m      | .             | 2       | 1.00           | 2           |
| 6  | 45  | m      | 0             | 9       | 4.00           | 2           |
| 7  | 44  | m      | 1             | 17      | 4.00           | 3           |
| 8  | 46  | m      | .             | 20      | 4.00           | 3           |
| 9  | 41  | m      | .             | 10      | 2.00           | 2           |
| 10 | 29  | f      | .             | 4       | 1.00           | 2           |

Data View Variable View

Rysunek 12. Dane zaimportowane z programu Excel

### Informacje pokrewne

“Pliki przykładowe” na stronie 77

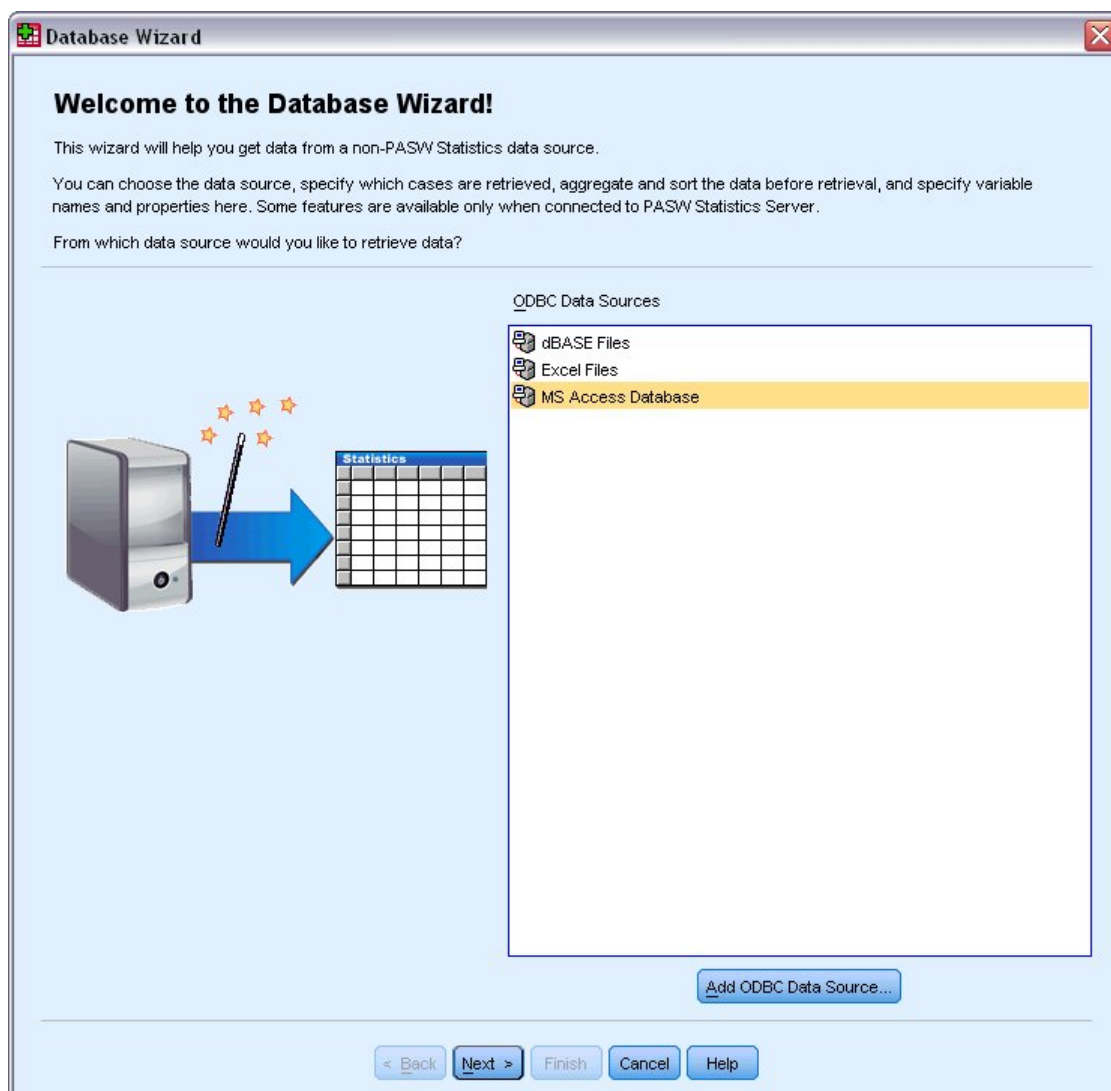
## Odczytywanie danych z bazy danych

Dane ze źródeł bazodanowych można łatwo importować za pomocą Kreatora przejmowania baz danych. Wszystkie bazy danych, które używają sterowników ODBC (Open Database Connectivity), można odczytywać bezpośrednio, o ile tylko zainstalowane są odpowiednie sterowniki. Sterowniki ODBC wielu formatów baz danych są dostarczane na instalacyjnym dysku CD. Dodatkowe sterowniki można uzyskać od innych firm. W tym przykładzie omawiamy jedną z najczęściej spotykanych aplikacji bazodanowych, Microsoft Access.

*Uwaga:* Ten przykład jest właściwy dla systemu Microsoft Windows i wymaga sterownika ODBC dla programu Access. Sterownik ODBC programu Microsoft Access działa tylko z 32-bitową wersją produktu IBM SPSS Statistics. Kroki na innych platformach są podobne, ale mogą wymagać uzyskania i użycia sterownika ODBC innej firmy.

1. Z menu wybierz:

**Plik > Importuj dane > Baza danych > Nowa kwerenda...**



Rysunek 13. Okno powitalne Kreator przejmowania baz danych

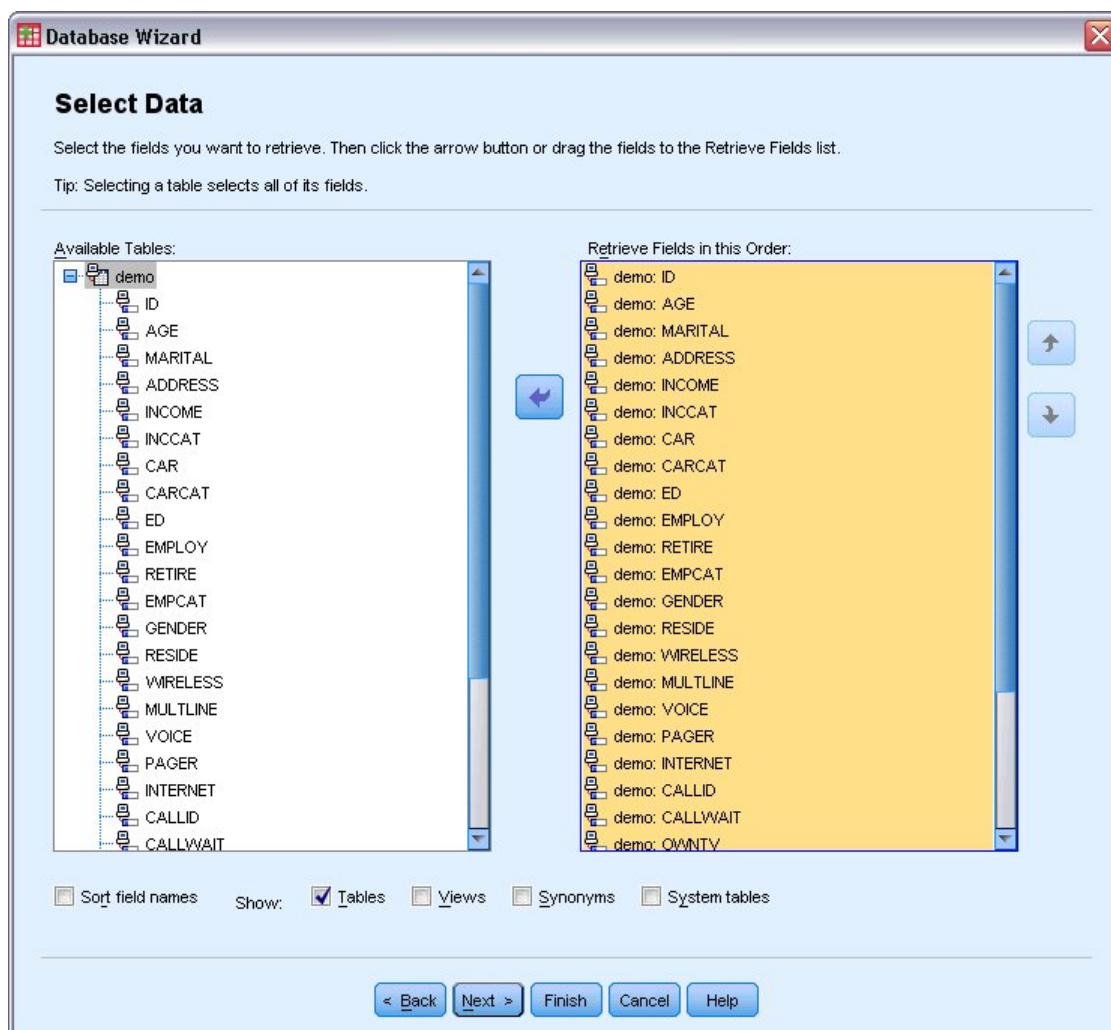
2. Z listy źródeł danych wybierz pozycję **MS Access Database** i kliknij przycisk **Dalej**.

*Uwaga:* W zależności od instalacji po lewej stronie kreatora może być również widoczna lista źródeł danych OLEDB (tylko w systemach operacyjnych Windows), ale w tym przykładzie używana jest lista źródeł danych ODBC wyświetlana po prawej stronie.

3. Kliknij przycisk **Przełączaj**, aby przejść do pliku bazy danych Access, który ma zostać otwarty.
4. Otwórz plik *demo.mdb*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.
5. Kliknij przycisk **OK** w oknie dialogowym logowania.

W następnym kroku można określić tabele i zmienne, które mają zostać zaimportowane.





Rysunek 14. Krok Wybierz dane

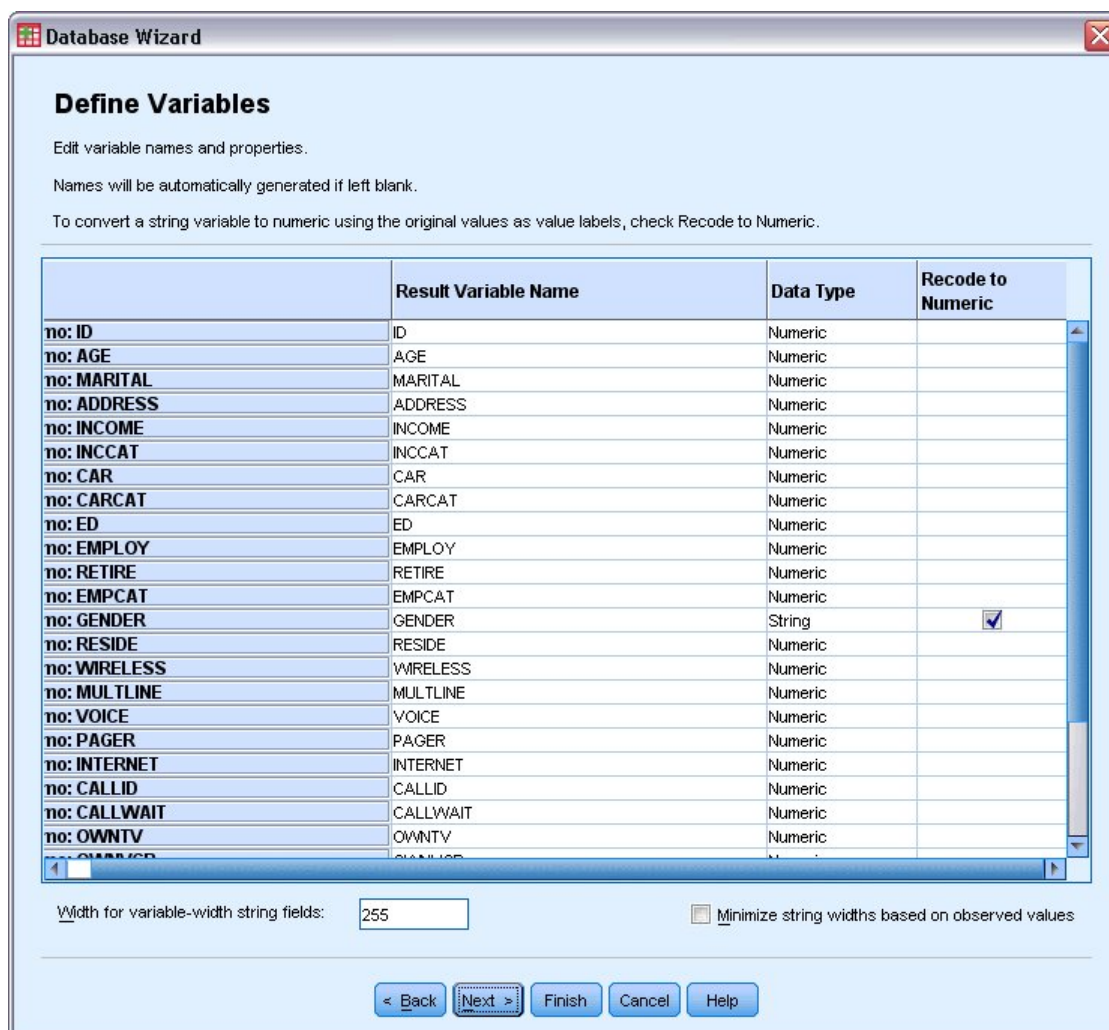
6. Przeciągnij całą tabelę **demo** do listy Pola pobierane będą w przedstawionym porządku.
7. Kliknij przycisk **Dalej**.

W następnym kroku można wybrać rekordy (obserwacje) do zaimportowania.

Jeśli nie chcesz importować wszystkich obserwacji, możesz zaimportować ze źródła danych podzbiór (np. rekordy mężczyzn w wieku powyżej 30 lat) lub losową próbę obserwacji. W przypadku dużych źródeł danych wskazane może być ograniczenie liczby obserwacji do małej, reprezentatywnej próby w celu skrócenia czasu przetwarzania.

8. Kliknij przycisk **Dalej**, aby kontynuować.

Nazwy pól w bazie danych są używane do tworzenia nazw zmiennych. W razie potrzeby nazwy pól są przekształcane w prawidłowe nazwy zmiennych. Pierwotne nazwy pól są zachowywane jako etykiety zmiennych. Można także zmienić nazwy zmiennych przed zaimportowaniem bazy danych.



Rysunek 15. Krok Definiowanie zmiennych

9. Kliknij komórkę **Rekoduj na numeryczną** w polu Gender. Ta opcja przekształca zmienne łańcuchowe w zmienne całkowite i zachowuje pierwotną wartość jako etykietę wartości dla nowej zmiennej.
10. Kliknij przycisk **Dalej**, aby kontynuować.

W kroku Wyniki zostanie wyświetlona instrukcja SQL utworzona na podstawie opcji wybranych w Kreatorze przejmowania baz danych. Ta instrukcja może zostać wykonana teraz lub zapisana w pliku do późniejszego użycia.

11. Kliknij przycisk **Zakończ**, aby zaimportować dane.

Wszystkie dane z bazy danych Access, które zostały wybrane do zaimportowania, są teraz dostępne w Edytorze danych.

## Odczytywanie danych z pliku tekstowego

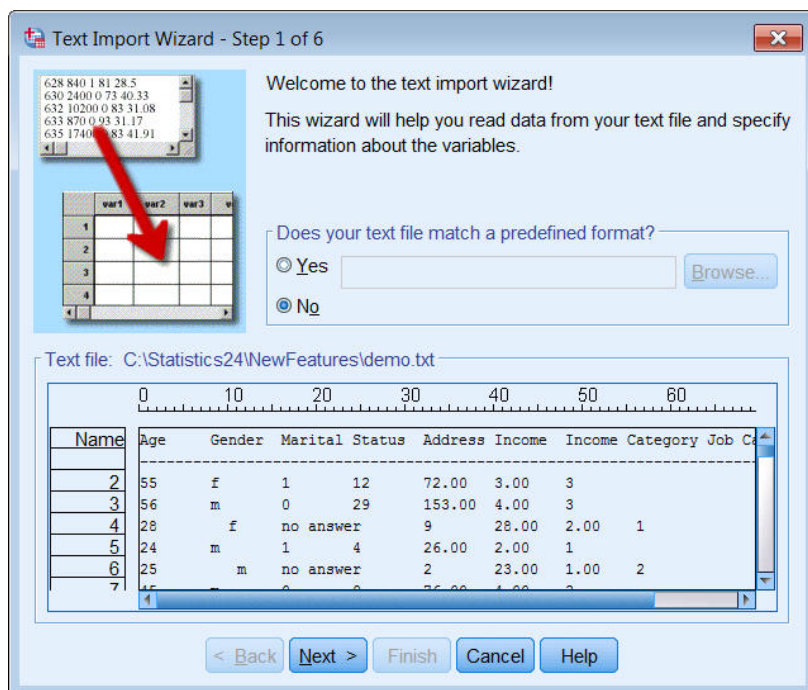
Pliki tekstowe są kolejnym często spotykanym źródłem danych. Wiele programów do obsługi arkuszy kalkulacyjnych i baz danych może zapisywać swoje dane w różnych formatach tekstowych. Pliki rozdzielane przecinkami lub tabulatorami zawierają wiersze danych, w których zmienne są rozdzielone przecinkami lub znakami tabulacji. W tym przykładzie dane są oddzielane znakami tabulacji.

1. Z menu wybierz:

**Plik > Importuj dane > Dane tekstowe**

2. Przejdź do folderu Samples\English i wybierz plik demo.txt.

Kreator importu tekstu prowadzi użytkownika przez proces definiowania sposobu interpretowania określonego pliku tekstowego.



Rysunek 16. Kreator importu tekstu: krok 1 z 6

3. W kroku 1 można wybrać format predefiniowany lub utworzyć nowy format w kreatorze. Wybierz opcję **Nie**.

4. Kliknij przycisk **Dalej**, aby kontynuować.

Jak wspomniano wcześniej, w tym pliku zmienne są rozdzielane znakami tabulacji. Ponadto nazwy zmiennych są zdefiniowane w pierwszym wierszu tego pliku.

5. W kroku 2 kreatora wybierz opcję **Separowane**, aby wskazać, że plik ma strukturę danych rozdzielanych ogranicznikami.

6. Wybierz opcję **Tak**, aby wskazać, że plik zawiera nazwy zmiennych w pierwszym wierszu.

7. Kliknij przycisk **Dalej**, aby kontynuować.

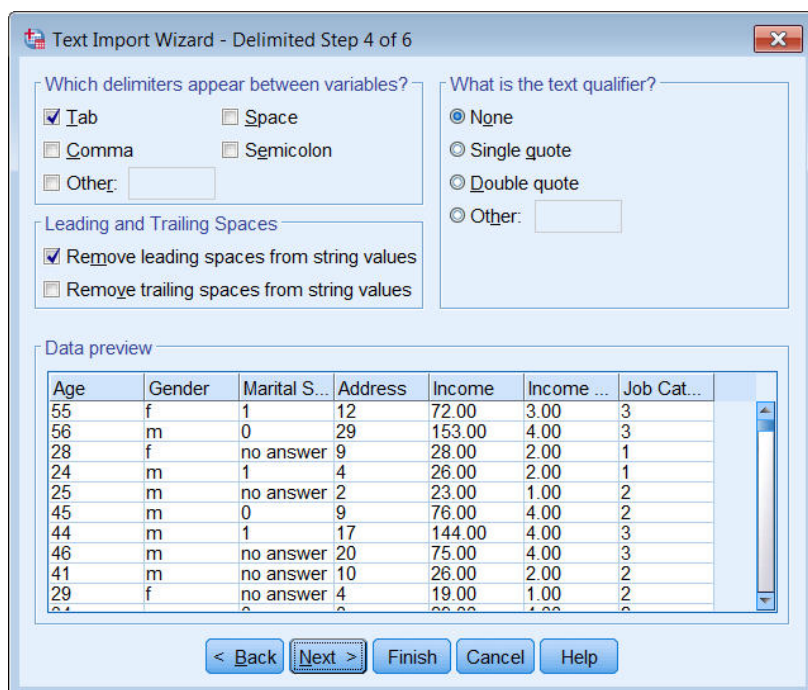
8. W kroku 3 wprowadź wartość 2 jako numer wiersza, w którym rozpoczyna się pierwsza obserwacja (ponieważ nazwy zmiennych znajdują się w pierwszym wierszu).

9. Zachowaj wartości domyślne w pozostałych polach tego kroku, a następnie kliknij przycisk **Dalej**, aby kontynuować.

Podgląd danych w kroku 4 umożliwia szybkie sprawdzenie, czy plik został poprawnie odczytany

10. Wybierz opcję **Tabulator** i usuń zaznaczenie innych opcji ograniczników. Domyślnie wybrana jest **Spacja**, ponieważ plik zawiera spacje. W tym pliku spacje są częścią wartości danych, a nie ogranicznikami. Aby poprawnie odczytać plik, należy usunąć zaznaczenie opcji **Spacja**.

11. Zaznacz opcję **Usuń początkowe spacje z wartości łańcuchowych**. Spacje na początku wartości łańcuchowych mają wpływ na sposób wyznaczania wartości łańcuchowych w wyrażeniach. W tym pliku niektóre wartości zmiennej *Gender* mają początkowe spacje, które nie są częścią wartości. Jeśli te spacje nie zostaną usunięte, wartość " f " będzie traktowana jako inna niż "f".



Rysunek 17. Kreator importu tekstu: krok 4 z 6

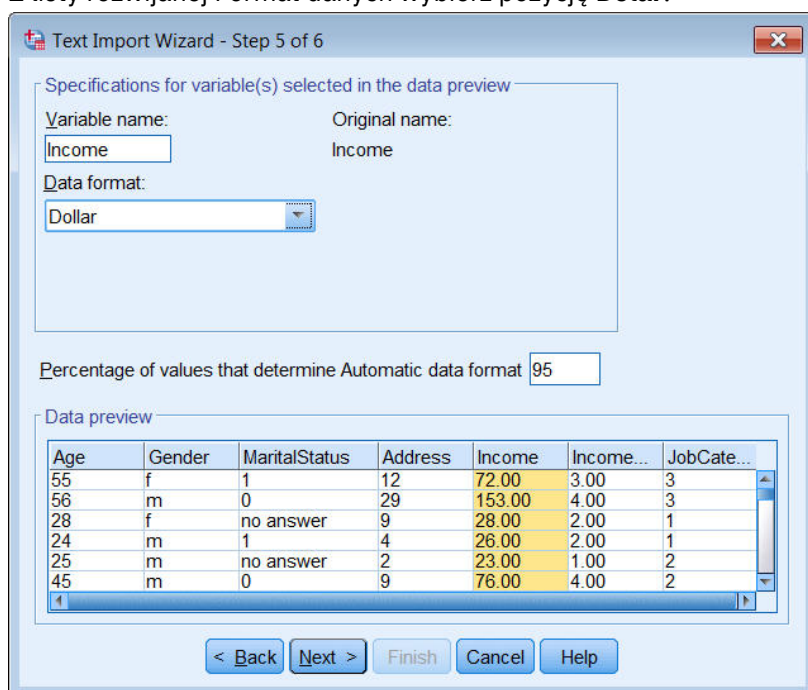
12. Kliknij przycisk **Dalej**, aby kontynuować.

Ze względu na to, że nazwy zmiennych są modyfikowane w sposób zgodny z regułami nazewnictwa, krok 5 daje możliwość edycji wszelkich niepożądanych nazw.

W tym kroku można również określić typy danych. Na przykład można zmienić typ zmiennej *Income* na walutowy (Dolar).

Aby zmienić typ danych:

13. W oknie **Podgląd danych** wybierz zmienną *Income*.  
 14. Z listy rozwijanej Format danych wybierz pozycję **Dolar**.



Rysunek 18. Zmiana typu danych

Zmienna *MaritalStatus* zawiera zarówno wartości łańcuchowe, jak i liczbowe. Mniej niż pięć procent zmiennych jest łańcuchami. Przy domyślnej wartości 95% parametru **Odsetek wartości określających typ danych** zmienna jest traktowana jako liczbowa, a wartości łańcuchowe są przekształcane w systemowe braki danych. Jeśli żaden format danych nie jest zgodny z parametrem odsetka, zmienna jest traktowana jako łańcuchowa. Jeśli wartość odsetka zostanie zmieniona na 100, wszystkie wartości zostaną zachowane, ale wszystkie wartości liczbowe będą traktowane jako łańcuchy.

15. Kliknij przycisk **Dalej**, aby kontynuować.
16. Pozostaw domyślne wybory w ostatnim kroku, a następnie kliknij przycisk **Zakończ**, aby zaimportować dane.



## Rozdział 3. Korzystanie z Edytora danych

W Edytorze danych wyświetlana jest zawartość aktywnego pliku danych. Informacje w Edytorze danych obejmują zmienne i obserwacje.

- W widoku danych kolumny reprezentują zmienne, a wiersze reprezentują obserwacje.
- W widoku zmiennych każdy wiersz jest zmienną, a każda kolumna jest atrybutem powiązaniem z tą zmienną.

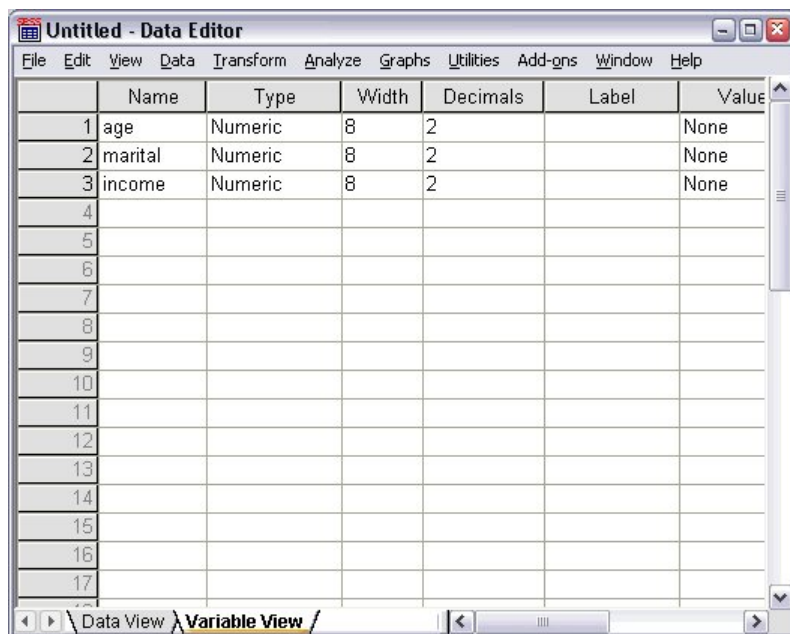
Zmienne są używane do reprezentowania różnych typów zebranych danych. Typową analogią jest badanie ankietowe. Odpowiedź na każde pytanie w ankiecie jest równoważne zmiennej. Zmienne mogą należeć do wielu różnych typów, takich jak liczby, łańcuchy, waluty i daty.

### Wprowadzanie danych liczbowych

Dane można wprowadzać w Edytorze danych, co bywa przydatne w przypadku mniejszych plików danych lub konieczności wprowadzenia drobnych modyfikacji do większych plików.

1. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.

Należy zdefiniować zmienne, które będą używane. W tym przypadku potrzebne są tylko trzy zmienne: *wiek*, *stancywi dochód*.



Rysunek 19. Nazwy zmiennych w widoku zmiennych

2. W pierwszym wierszu pierwszej kolumny wpisz *wiek*.
3. W drugim wierszu wpisz *stan cywilny*.
4. W trzecim wierszu wpisz *dochód*.

Nowym zmienne jest automatycznie nadawany typ liczbowy.

Jeśli nazwy zmiennych nie zostaną wprowadzone, automatycznie tworzone są unikalne nazwy. Jednak nazwy te nie są opisowe i nie są zalecane w przypadku dużych plików danych.

5. Kliknij kartę **Widok danych**, aby kontynuować wprowadzanie danych.

Nazwy wprowadzone w widoku zmiennych są teraz nagłówkami pierwszych trzech kolumn w widoku danych.

Rozpocznij wprowadzanie danych w pierwszym wierszu, rozpoczynając od pierwszej kolumny.

The screenshot shows the 'Untitled - Data Editor' window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The title bar shows '19 : age'. The data table has the following structure:

|    | age   | marital | income    | var | var | var |
|----|-------|---------|-----------|-----|-----|-----|
| 1  | 55.00 | 1.00    | 72000.00  |     |     |     |
| 2  | 53.00 | .00     | 153000.00 |     |     |     |
| 3  |       |         |           |     |     |     |
| 4  |       |         |           |     |     |     |
| 5  |       |         |           |     |     |     |
| 6  |       |         |           |     |     |     |
| 7  |       |         |           |     |     |     |
| 8  |       |         |           |     |     |     |
| 9  |       |         |           |     |     |     |
| 10 |       |         |           |     |     |     |
| 11 |       |         |           |     |     |     |
| 12 |       |         |           |     |     |     |
| 13 |       |         |           |     |     |     |
| 14 |       |         |           |     |     |     |
| 15 |       |         |           |     |     |     |
| 16 |       |         |           |     |     |     |

At the bottom, there are tabs for 'Data View' and 'Variable View', with 'Data View' selected.

Rysunek 20. Wartości wprowadzone w widoku danych

6. W kolumnie *wiek* wpisz 55.
7. W kolumnie *stancyw* wpisz 1.
8. W kolumnie *dochód* wpisz 72000.
9. Przesuń kursor do drugiego wiersza w pierwszej kolumnie, aby dodać dane następnego obiektu.
10. W kolumnie *wiek* wpisz 53.
11. W kolumnie *stancyw* wpisz 0.
12. W kolumnie *dochód* wpisz 153000.

W tej chwili kolumny *wiek* i *stancyw* przedstawiają dane z separatorami dziesiętnymi, mimo że ich wartości mają być liczbami całkowitymi. Aby ukryć separatory dziesiętne w tych zmiennych:

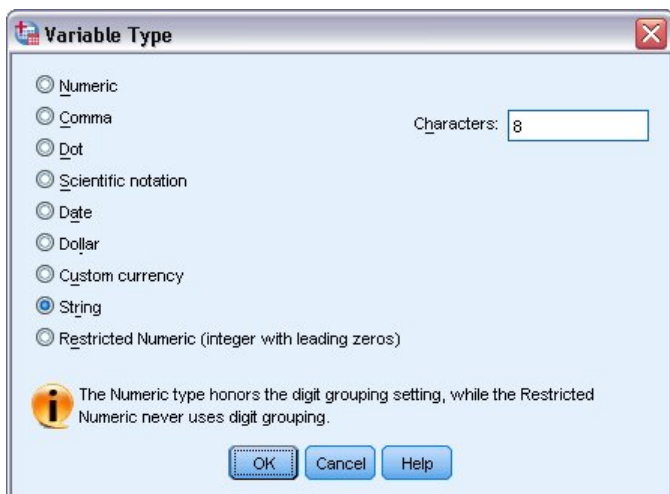
13. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.
14. W kolumnie *Dziesiętne* w wierszu *wiek* wpisz 0, aby ukryć miejsca dziesiętne.
15. W kolumnie *Dziesiętne* w wierszu *stancyw* wpisz 0, aby ukryć miejsca dziesiętne.

## Wprowadzanie danych łańcuchowych

Dane nieliczbowe, takie jak łańcuchy tekstowe, także mogą być wprowadzane w Edytorze danych.

1. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.
2. W pierwszej komórce pierwszego pustego wiersza wpisz *płec* jako nazwę zmiennej.
3. Kliknij komórkę *Typ* znajdującą się obok tego wpisu.
4. Kliknij przycisk znajdujący się z prawej strony komórki *Typ*, aby otworzyć okno dialogowe *Typ zmiennej*.
5. Wybierz **Łańcuch**, aby określić typ zmiennej.
6. Kliknij przycisk **OK**, aby zapisać wybór i powrócić do Edytora danych.





Rysunek 21. Okno dialogowe Typ zmiennej

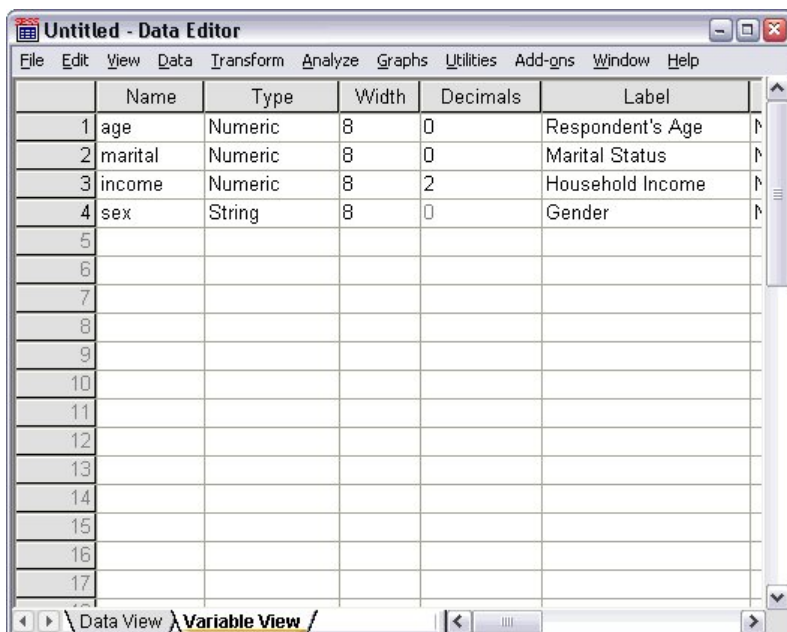
## Definiowanie danych

Oprócz definiowania typów danych można również definiować opisowe etykiety zmiennych oraz etykiety wartości zmiennych i danych. Te etykiety opisowe są używane w raportach i na wykresach statystycznych.

### Dodawanie etykiet zmiennych

Etykiety są opisami zmiennych. Opisy te są często dłuższymi wersjami nazw zmiennych. Etykiety mogą mieć długość maksymalnie 255 bajtów. Etykiety te są używane w wynikach do identyfikacji różnych zmiennych.

1. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.
2. W kolumnie *Etykieta* wiersza *wiek* wpisz *Wiek* respondenta.
3. W kolumnie *Etykieta* wiersza *stancyw* wpisz *Stan* cywilny.
4. W kolumnie *Etykieta* wiersza *dochód* wpisz *Dochód* gospodarstwa domowego.
5. W kolumnie *Etykieta* wiersza *pleć* wpisz *Płeć* osoby.

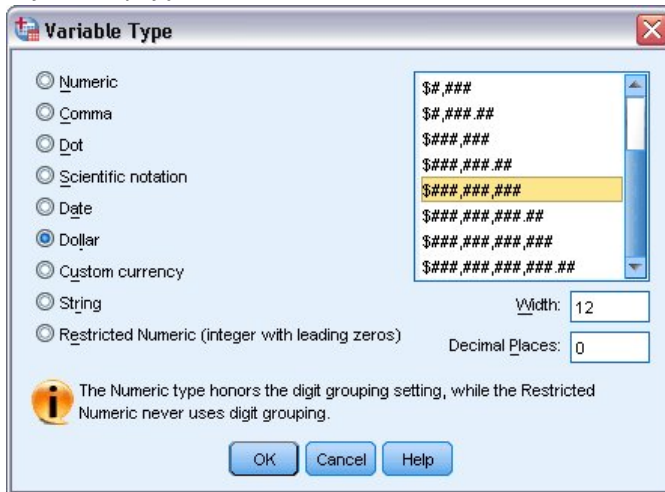


Rysunek 22. Etykiety zmiennych wprowadzone w widoku zmiennych

## Zmiana typu i formatu zmiennej

W kolumnie *Typ* wyświetlany jest bieżący typ danych każdej zmiennej. Najczęściej spotykane typy danych to typ liczbowy i łańcuchowy, ale obsługiwane są także liczne inne formaty. W bieżącym pliku danych zmienna *dochód* jest zdefiniowana jako liczbowy.

1. Kliknij komórkę *Typ* dla wiersza *dochód*, a następnie kliknij przycisk znajdujący się z prawej strony komórki, aby otworzyć okno dialogowe Typ zmiennej.
2. Wybierz opcję **Dolar**.



Rysunek 23. Okno dialogowe Typ zmiennej

Wyświetlane zostaną opcje formatowania wybranego typu danych.

3. Jako format waluty w tym przykładzie należy wybrać opcję **\$###,###,###**.
4. Kliknij przycisk **OK**, aby zapisać zmiany.

## Dodawanie etykiet wartości

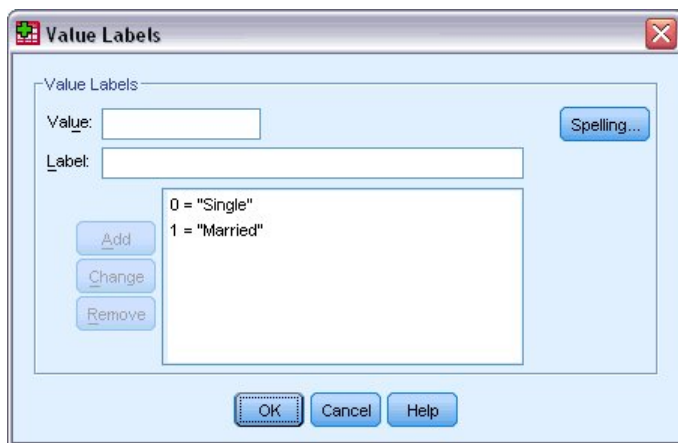
Etykiety wartości umożliwiają odwzorowanie wartości zmiennych na etykiety łańcuchowe. W tym przykładzie istnieją dwie akceptowalne wartości zmiennej *stancyw*. Wartość 0 oznacza, że obiekt jest stanu wolnego, a wartość 1 oznacza, że jest osobą żoną/zamężną.

1. Kliknij komórkę *Wartości* dla wiersza *stancyw*, a następnie kliknij przycisk z prawej strony komórki, aby otworzyć okno dialogowe Etykiety wartości.

**Wartość** jest rzeczywistą wartością liczbową.

**Etykieta wartości** to etykieta łańcuchowa przypisana do określonej wartości liczbowej.

2. W polu *Wartość* wpisz wartość 0.
3. Wpisz *Wolny* w polu *Etykieta*.
4. Kliknij przycisk **Dodaj**, aby dodać tę etykietę do listy.



Rysunek 24. Okno dialogowe Etykiety wartości

5. W polu Wartość wpisz 1, a w polu Etykieta wpisz Żonaty / Zamężna.
6. Kliknij przycisk **Dodaj**, a następnie kliknij przycisk **OK**, aby zapisać zmiany i powrócić do Edytora danych.

Etykiety te mogą być również wyświetlane w widoku danych, sprawiając, że dane będą bardziej czytelne.

7. Kliknij kartę **Dane** w dolnej części okna Edytora danych.
8. Z menu wybierz:

#### **Widok > Etykiety wartości**

Etykiety będą teraz wyświetlane na liście podczas wprowadzania wartości w Edytorze danych. Takie rozwiązanie sugeruje poprawne odpowiedzi i sprawia, że są one bardziej zrozumiałe.

Jeśli element menu Etykiety wartości jest już aktywny (ma znacznik wyboru), ponowne wybranie opcji **Etykiety wartości** spowoduje *wyłączenie* wyświetlania etykiet wartości.

## **Obsługa braków danych**

Braki danych lub niepoprawne dane występują zbyt często, aby można je było ignorować. Respondenci mogą odmówić odpowiedzi na niektóre pytania, nie znać odpowiedzi lub udzielić odpowiedzi w nieoczekiwanym formacie. Jeśli nie odfiltrujesz lub nie zidentyfikujesz takich danych, wyniki analizy mogą być niedokładne.

W przypadku danych liczbowych puste pola danych lub pola zawierające niepoprawne wpisy są przekształcane w systemowe braki danych, oznaczone pojedynczymi kropkami.

Powód, dla którego brakuje wartości, może być istotny dla wyników analizy. Na przykład przydatne może się okazać rozróżnienie między respondentami, którzy odmówili odpowiedzi na pytanie, a respondentami, którzy nie odpowiedzieli na pytanie, ponieważ nie miało ono zastosowania w ich przypadku.

## **Brakujące wartości zmiennej liczbowej**

1. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.
2. Kliknij komórkę *Brakujące wartości* w wierszu *age*, a następnie kliknij przycisk z prawej strony komórki, aby otworzyć okno dialogowe Brakujące wartości.

W tym oknie dialogowym można określić maksymalnie trzy odrębne brakujące wartości lub można określić zakres wartości plus jedną dodatkową wartość dyskretną.



Rysunek 25. Okno dialogowe Brakujące wartości

3. Wybierz opcję **Wartości dyskretne braków**.
4. Wpisz 999 w pierwszym polu tekstowym i pozostaw dwa pozostałe pola tekstowe puste.
5. Kliknij przycisk **OK**, aby zapisać zmiany i powrócić do Edytora danych.  
Teraz, gdy została dodana brakująca wartość danych, można zastosować etykietę do tej wartości.
6. Kliknij komórkę *Wartości* w wierszu *wiek*, a następnie kliknij przycisk znajdujący się z prawej strony komórki, aby otworzyć okno dialogowe Etykiety wartości.
7. W polu Wartość wpisz wartość 999.
8. W polu Etykieta wpisz Brak odpowiedzi.
9. Kliknij przycisk **Dodaj**, aby dodać tę etykietę do pliku danych.
10. Kliknij przycisk **OK**, aby zapisać zmiany i powrócić do Edytora danych.

## Brakujące wartości zmiennej łańcuchowej

Brakujące wartości zmiennych łańcuchowych są obsługiwane podobnie, jak brakujące wartości zmiennych liczbowych. Jednak w przeciwieństwie do zmiennych liczbowych puste zmienne łańcuchowe nie są oznaczane jako systemowe braki danych. Są one interpretowane po prostu jako puste łańcuchy.

1. Kliknij kartę **Zmienne** w dolnej części okna Edytora danych.
2. Kliknij komórkę *Brak danych* w wierszu *wiek*, a następnie kliknij przycisk z prawej strony komórki, aby otworzyć okno dialogowe Brakujące wartości.
3. Wybierz opcję **Wartości dyskretne braków**.
4. Wpisz B0 w pierwszym polu tekstowym.  
W przypadku braków danych w zmiennych łańcuchowych rozróżniana jest wielkość liter. Zatem wartość *bo* nie będzie traktowana jako brakująca wartość.
5. Kliknij przycisk **OK**, aby zapisać zmiany i powrócić do Edytora danych.  
Teraz można dodać etykietę brakującej wartości.
6. Kliknij komórkę *Wartości* w wierszu *pleć*, a następnie kliknij przycisk znajdujący się z prawej strony komórki, aby otworzyć okno dialogowe Etykiety wartości.
7. W polu Wartość wpisz B0.
8. W polu Etykieta wpisz Brak odpowiedzi.
9. Kliknij przycisk **Dodaj**, aby dodać tę etykietę do projektu.
10. Kliknij przycisk **OK**, aby zapisać zmiany i powrócić do Edytora danych.

---

# Rozdział 4. Interpretacja statystyk podsumowujących poszczególnych zmiennych

W tej sekcji omówiono proste miary podsumowujące oraz sposób, w jaki poziom pomiaru zmiennej wpływa na typy statystyk, które powinny być używane. Wykorzystamy plik danych *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

---

## Poziom pomiaru

Różne miary są odpowiednie dla różnych typów danych, w zależności od poziomu pomiaru:

**Jakościowy rozkład kategorii.** Dane z ograniczoną liczbą odrębnych wartości lub kategorii (na przykład płeć lub stan cywilny). Nazywane są również **danymi jakościowymi**. Zmienne kategoryjne mogą być zmiennymi łańcuchowymi (alfanumerycznymi) lub liczbowymi, wykorzystującymi kody liczbowe reprezentujące kategorie (np. 0 = *wolny* i 1 = *żonaty/zamężna*). Istnieją dwa podstawowe typy danych kategoryjnych:

- **Nominalne.** Dane kategoryjne, w których nie istnieje naturalny porządek kategorii. Na przykład kategoria pracy *sprzedaż* nie jest wyższa ani niższa niż kategoria pracy *marketing* lub *badania*.
- **Porządkowe.** Dane kategoryjne, w przypadku których istnieje znaczący porządek kategorii, ale nie ma mierzalnej odległości między kategoriami. Na przykład istnieje porządek wartości *wysoki*, *średni* i *niski*, ale nie można obliczyć „odległości” między tymi wartościami.

**Ilościowe.** Dane mierzone na skali interwałowej lub ilorazowej, których wartości określają zarówno ich porządek, jak i odległość między nimi. Na przykład wynagrodzenie w wysokości 7219 zł jest wyższe niż wynagrodzenie w wysokości 5239 zł, a odległość pomiędzy tymi dwiema wartościami wynosi 1980 zł. Zwane również **danymi ciągłymi**.

---

## Miary podsumowujące danych kategoryjnych

W przypadku danych kategoryjnych najbardziej typową miarą podsumowującą jest liczba lub odsetek obserwacji w każdej kategorii. **Dominanta** to kategoria, w której znajduje się największa liczba obserwacji. W przypadku danych porządkowych **mediana** (wartość, powyżej i poniżej której znajduje się po połowie obserwacji), może być również użyteczną miarą podsumowującą, jeśli liczba kategorii jest duża.

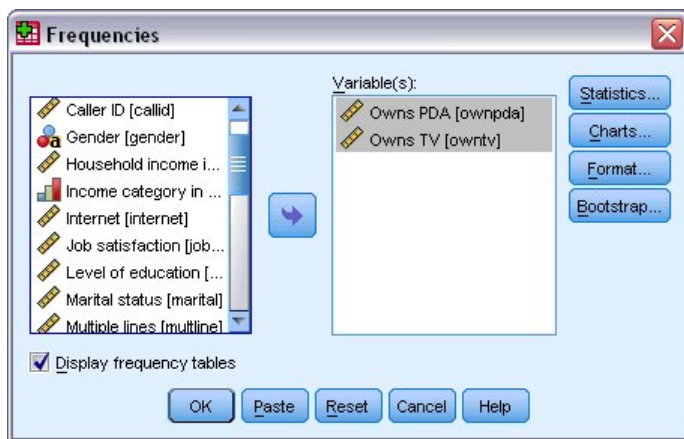
Procedura Częstości tworzy tabele częstości, które zawierają zarówno liczbę, jak i odsetek dla każdej obserwowanej wartości zmiennej.

1. Z menu wybierz:

**Analiza > Statystyki opisowe > Częstości...**

Uwaga: ta funkcja wymaga opcji Statistics Base.

2. Wybierz zmienne *Owns PDA [ownpda]* i *Owns TV [owntv]*, po czym przenieś je na listę Zmienne.



Rysunek 26. Zmienne kategoryjne wybrane do analizy

3. Kliknij przycisk **OK**, aby uruchomić procedurę.

**Frequency Table**

**Owns PDA**

|       |     | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----|-----------|---------|---------------|--------------------|
| Valid | No  | 5093      | 79.6    | 79.6          | 79.6               |
|       | Yes | 1307      | 20.4    | 20.4          | 100.0              |
| Total |     | 6400      | 100.0   | 100.0         |                    |

**Owns TV**

|       |     | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----|-----------|---------|---------------|--------------------|
| Valid | No  | 63        | 1.0     | 1.0           | 1.0                |
|       | Yes | 6337      | 99.0    | 99.0          | 100.0              |
| Total |     | 6400      | 100.0   | 100.0         |                    |

Rysunek 27. Tabele częstości

Tabele częstości zostaną wyświetlone w oknie raportu. Z tabel częstości wynika, że tylko 20,4% badanych osób posiada urządzenia PDA, ale prawie każdy ma telewizor (99,0%). To spostrzeżenie niekoniecznie jest interesujące, ale być może warto byłoby dowiedzieć się czegoś więcej o małej grupie osób, które nie posiadają telewizorów.

## Wykresy dla danych kategoryjnych

Informacje zawarte w tabeli częstości można przedstawić graficznie na wykresie słupkowym lub kołowym.

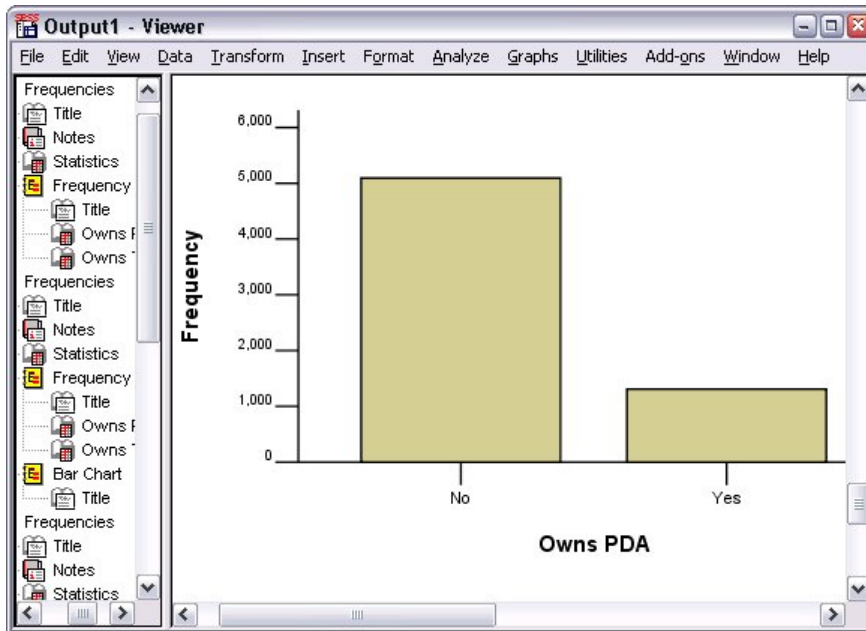
1. Ponownie otwórz okno dialogowe Częstości. (Dotychczasowe dwie zmienne nadal powinny być wybrane).

Aby szybko wracać do ostatnio używanych procedur, można używać przycisku Przywołaj okno dialogowe na pasku narzędzi.



Rysunek 28. Przycisk Przywołaj okno dialogowe

2. Kliknij opcję **Wykresy**.
3. Wybierz opcję **Wykresy słupkowe**, a następnie kliknij przycisk **Kontynuuj**.
4. Kliknij przycisk **OK** w głównym oknie dialogowym, aby uruchomić procedurę.



Rysunek 29. Wykres słupkowy

Te same informacje co w tabelach częstości są teraz wyświetlane w postaci wykresów słupkowych, dzięki czemu łatwo zauważyć, że większość osób nie ma urządzeń PDA, ale prawie wszyscy są właścicielkami telewizorów.

## Miary podsumowujące zmiennych ilościowych

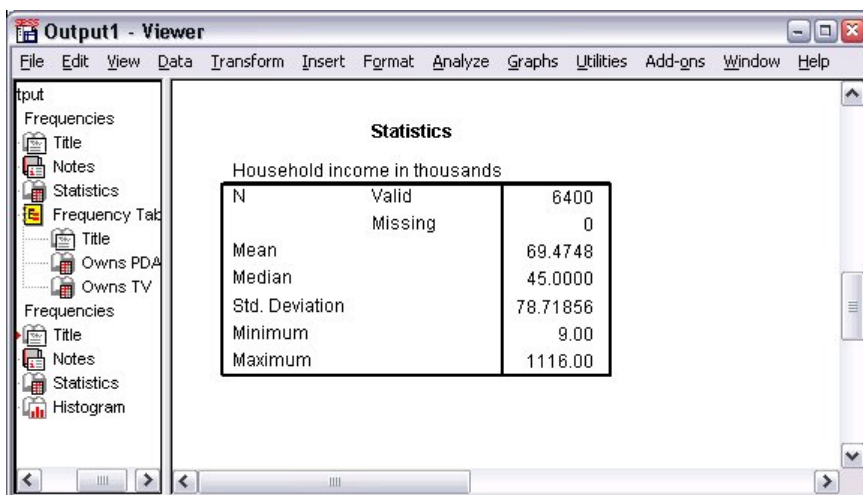
Dostępnych jest wiele miar podsumowujących odpowiednich dla zmiennych ilościowych, w tym:

- **Miary tendencji centralnej.** Najczęściej stosowanymi miarami tendencji centralnej są: **średnia** (arytmetyczna) i **mediana** (wartość, powyżej i poniżej której znajduje się po połowie obserwacji).
- **Miary rozproszenia.** Statystyki, które mierzą zmienność lub rozrzut danych, obejmują odchylenie standardowe, minimum i maksimum.

1. Ponownie otwórz okno dialogowe Częstości.
2. Kliknij przycisk **Wyzeruj**, aby usunąć poprzednie ustawienia.
3. Wybierz zmienną *Household income in thousands [income]* i przenieś ją na listę Zmienne.
4. Kliknij przycisk **Statystyka**.
5. Wybierz opcję **Średnia, Mediana, Odchylenie standardowe, Minimum i Maksimum**.
6. Kliknij przycisk **Kontynuuj**.
7. Anuluj zaznaczenie opcji **Pokaż tabele częstości** w głównym oknie dialogowym. (Tabele częstości zwykle nie są użyteczne w przypadku zmiennych ilościowych, ponieważ w pliku danych może występować prawie tyle odrębnych wartości, ile jest obserwacji).
8. Kliknij przycisk **OK**, aby uruchomić procedurę.



Tabela Statystyki częstości jest wyświetlana w oknie raportu.

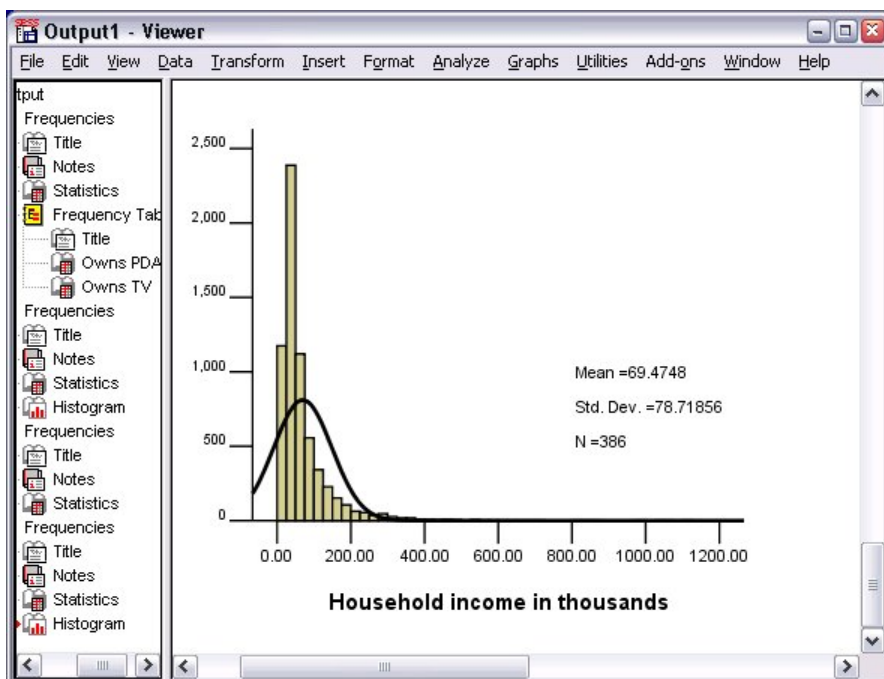


Rysunek 30. Tabela Statystyki częstości

W tym przykładzie występuje duża różnica między średnią a medianą. Średnia jest o prawie 25 000 większa od mediany, co oznacza, że wartości nie mają rozkładu normalnego. Rozkład można sprawdzić wzrokowo, korzystając z histogramu.

## Histogramy zmiennych ilościowych

1. Ponownie otwórz okno dialogowe Częstości.
2. Kliknij opcję **Wykresy**.
3. Wybierz opcję **Histogramy i Krzywa normalna**.
4. Kliknij przycisk **Kontynuuj**, a następnie kliknij przycisk **OK** w głównym oknie dialogowym, aby uruchomić procedurę.



Rysunek 31. Histogram

Większość obserwacji jest zgrupowana w dolnej części skali, przy czym większość z nich wypada poniżej 100 000. Istnieje jednak kilka obserwacji z zakresu od 500 000 wzwyż (jest ich za mało, aby były



widoczne bez modyfikacji histogramu). Te wysokie wartości kilku obserwacji mają istotny wpływ na średnią, ale w niewielkim lub żadnym stopniu wpływają na medianę, co powoduje, że mediana jest lepszym wskaźnikiem tendencji centralnej w tym przykładzie.



# Rozdział 5. Tworzenie i edytowanie wykresów

Istnieje możliwość tworzenia i edytowania wielu różnych typów wykresów. W tym rozdziale utworzymy i zmodyfikujemy wykresy słupkowe. Możesz zastosować te same zasady do dowolnego typu wykresu.

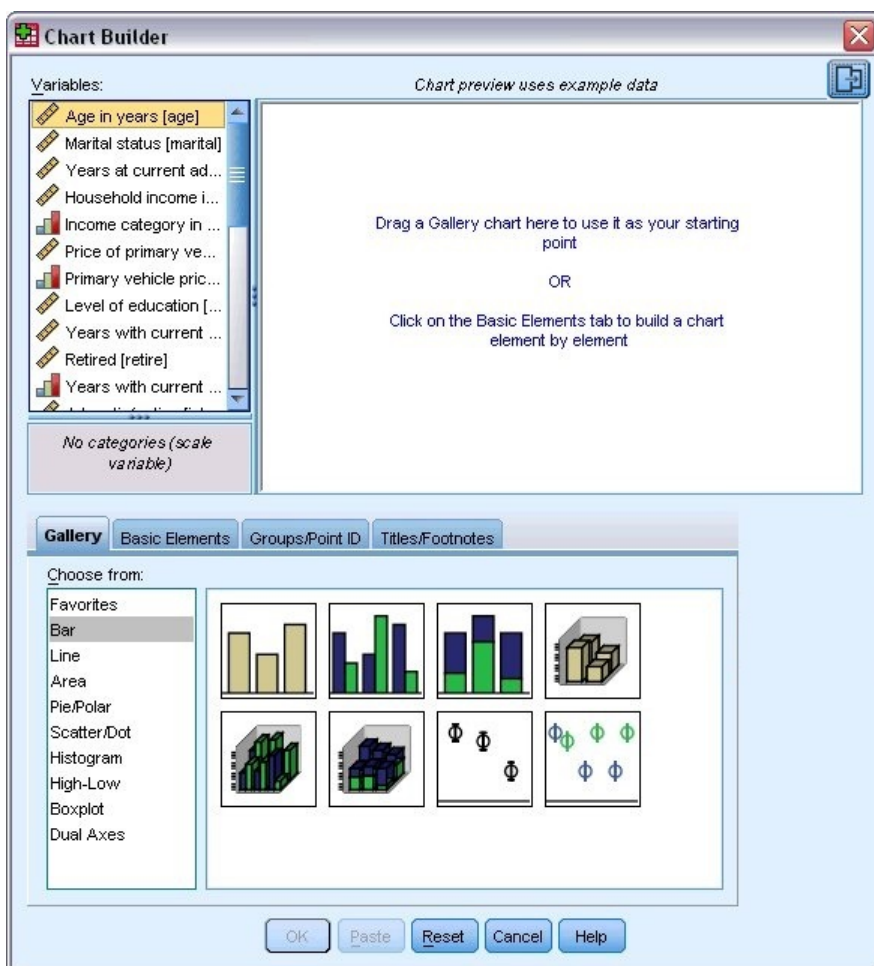
## Podstawowe informacje o tworzeniu wykresów

Aby zademonstrować podstawy tworzenia wykresów, stworzymy wykres słupkowy średnich dochodów w funkcji poziomu zadowolenia z pracy. W tym przykładzie używany jest plik danych *demo.sav*. Więcej informacji zawiera temat Rozdział 10, "Pliki przykładowe", na stronie 77.

1. Z menu wybierz:

### Wykresy > Kreator wykresów...

Okno dialogowe Kreator wykresów to interaktywne okno, które umożliwia podgląd wykresu podczas jego tworzenia.



Rysunek 32. Okno dialogowe Kreator wykresów

### Korzystanie z galerii kreatora wykresów

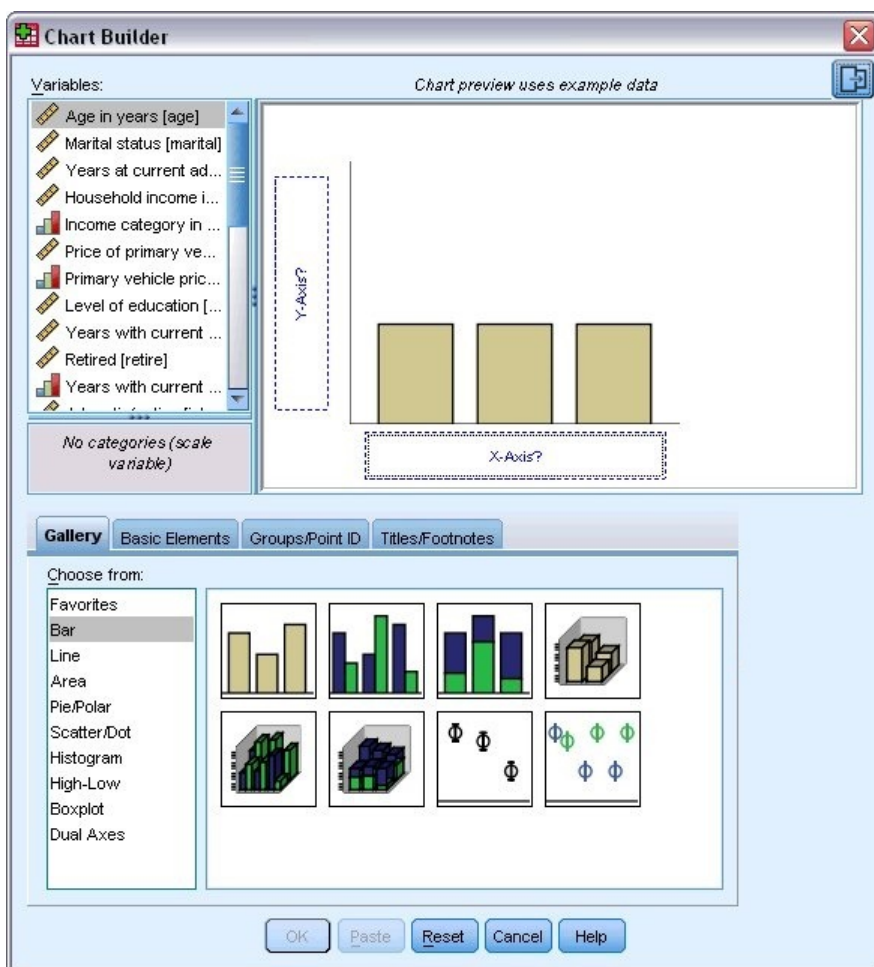
1. Kliknij kartę **Galeria**, jeśli nie jest jeszcze wybrana.

Galeria zawiera wiele różnych predefiniowanych wykresów, które są zorganizowane według typu. Karta Elementy podstawowe zawiera także elementy podstawowe (takie jak osie i elementy graficzne) do tworzenia wykresów od podstaw, jednak łatwiej jest skorzystać z galerii.

2. Kliknij typ **Słupkowy**, jeśli nie jest wybrany.

W oknie dialogowym pojawią się ikony reprezentujące wykresy słupkowe dostępne w galerii. Ikony powinny wystarczyć do identyfikacji konkretnych rodzajów wykresów. Jeśli potrzebujesz więcej informacji, możesz także wyświetlić podpowiedź do wykresu, zatrzymując kursor nad ikoną.

3. Przeciągnij ikonę prostego wykresu słupkowego na obszar roboczy (duży obszar nad galerią). Kreator wykresów wyświetli podgląd wykresu w obszarze roboczym. Dane użyte do narysowania wykresu nie są danymi rzeczywistymi. Są to dane przykładowe.



Rysunek 33. Wykres słupkowy w obszarze roboczym kreatora wykresów.

## Definiowanie zmiennych i statystyk

Chociaż w obszarze roboczym znajduje się wykres, nie jest on kompletny, ponieważ nie ma zmiennych lub statystyk, które określałyby, jak wysokie mają być słupki, oraz które kategorie zmiennych odpowiadają poszczególnym słupkom. Wykres faktycznie nie istnieje bez zmiennych i statystyk. Zmienne można dodawać, przeciągając je z listy Zmienne, która znajduje się po lewej stronie obszaru roboczego.

W Kreatorze wykresów istotny jest poziom pomiaru zmiennej. Zmiennej *Zadowolenie z pracy* chcemy użyć na osi x. Jednak ikona (która wygląda jak linijka) obok zmiennej wskazuje, że jej poziom pomiaru jest zdefiniowany jako ilościowy. Aby utworzyć poprawny wykres, musimy użyć jakościowego (kategorialnego) poziomu pomiaru. Zamiast wracać do widoku zmiennych i tam zmieniać poziom pomiaru, można tymczasowo zmienić poziom pomiaru w Kreatorze wykresów.

1. Kliknij prawym przyciskiem pozycję *Zadowolenie z pracy* na liście Zmienne i wybierz opcję **Porządkowy**. Poziom Porządkowy jest odpowiednim poziomem pomiaru, ponieważ kategorie zmiennej *Zadowolenie z pracy* mogą być uporządkowane według poziomu zadowolenia. Zwróć uwagę, że ikona zmieni się po zmianie poziomu pomiaru.
2. Teraz przeciągnij zmienną *Zadowolenie z pracy* z listy Zmienne do strefy upuszczania osi x.  
Strefa upuszczania osi y jest domyślnie ustawiona na statystykę *Liczebność*. Jeśli chcesz użyć innej statystyki (jak na przykład procent lub średnia), możesz ją łatwo zmienić. W tym przykładzie nie użyjemy żadnej z tych statystyk, ale omówimy proces zmiany statystyki, na wypadek gdyby taka konieczność zaszła w przyszłości.
3. Kliknij zakładkę **Właściwości elementu** na pasku bocznym Kreatora wykresów. (Jeśli pasek boczny nie jest widoczny, kliknij przycisk w prawym górnym rogu Kreatora wykresów, aby wyświetlić pasek boczny).



Rysunek 34. Właściwości elementu

Właściwości elementu pozwalają na zmianę właściwości różnych elementów wykresu. Należą do nich elementy graficzne (takie jak słupki na wykresie słupkowym) i osie na wykresie. Aby zmienić właściwości powiązane z tym elementem, należy wybrać jeden z elementów w obszarze Edytuj właściwości listy. Zwróć również uwagę na czerwony symbol X znajdujący się z prawej strony listy. Ten przycisk usuwa element graficzny z obszaru roboczego. Ponieważ wybrany jest element **Słupek1**, wyświetlane właściwości mają zastosowanie do elementów graficznych, a w szczególności do słupka.

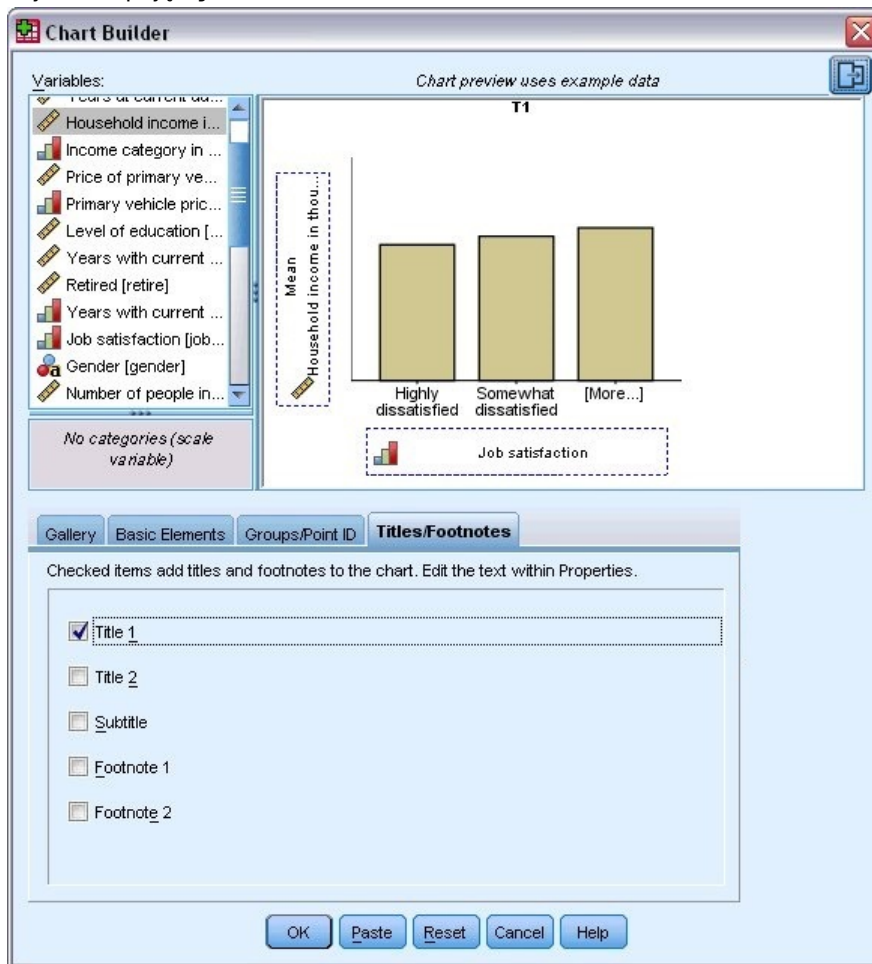
Na liście rozwijanej Statystyka wyświetlane są dostępne statystyki. Dla każdego typu wykresu są zwykle dostępne te same statystyki. Należy pamiętać, że niektóre statystyki wymagają, aby strefa upuszczania osi y zawierała zmienną.

4. Przeciągnij zmienną *Dochód gospodarstwa domowego w tysiącach* z listy Zmienne do strefy upuszczania osi y. Ponieważ zmienna na osi y jest ilościowa, a zmienna osi x jest jakościowa (poziom porządkowy jest typem jakościowym, czyli kategoryjnym), w strefie upuszczania osi y domyślnie przyjmowana jest statystyka *Średnia*. Są to zmienne i statystyki, które nas interesują, więc nie trzeba zmieniać właściwości elementów.

## Dodawanie tekstu

Do wykresu można również dodawać tytuły i przypisy.

1. Kliknij kartę **Tytuły/stopki**.
2. Wybierz opcję **Tytuł 1**.



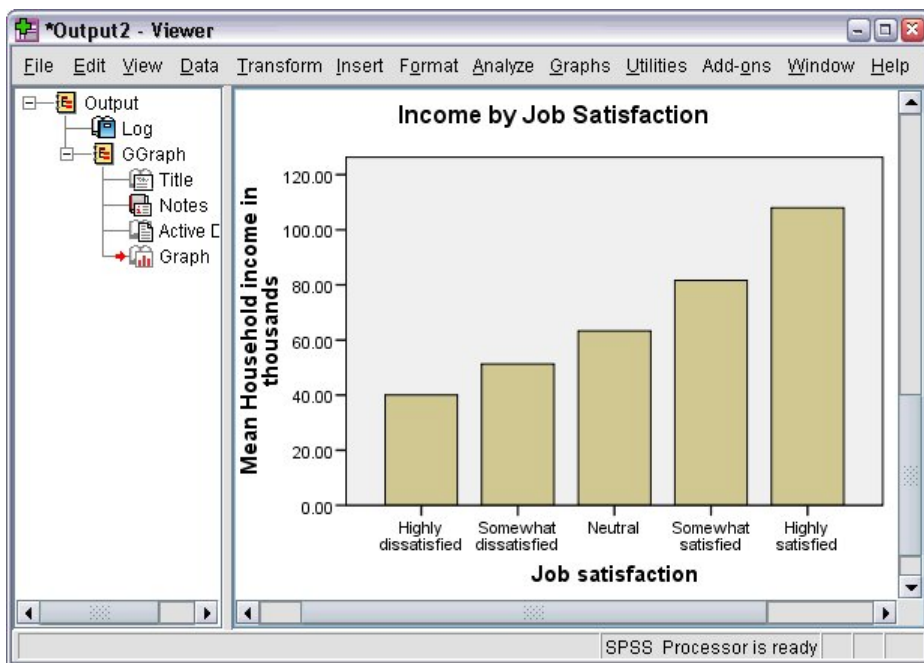
Rysunek 35. Tytuł 1 wyświetlony w obszarze roboczym

Tytuł jest widoczny w obszarze roboczym z etykietą **T1**.

3. Na karcie **Właściwości elementu** wybierz pozycję **Tytuł 1** w obszarze Edytuj właściwości listy.
4. W polu tekstowym Zawartość wpisz Dochód w zależności od zadowolenia z pracy. Jest to tekst, który będzie wyświetlany w tytule.

## Tworzenie wykresu

1. Kliknij przycisk **OK**, aby utworzyć wykres słupkowy.



Rysunek 36. Wykres słupkowy

Z wykresu słupkowego wynika, że respondenci, którzy są bardziej zadowoleni ze swojej pracy, mają zwykle wyższe dochody w gospodarstwie domowym.

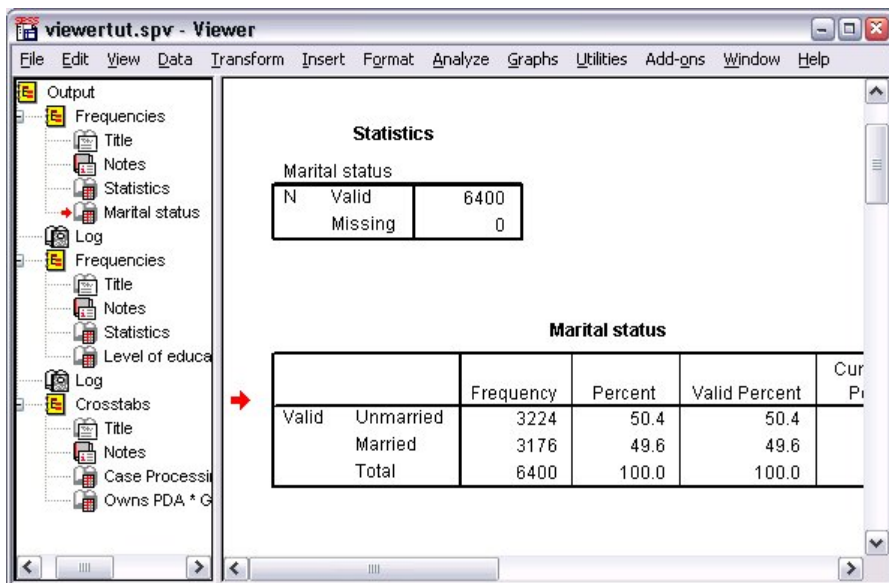




## Rozdział 6. Praca z oknem raportu

Wyniki wykonania procedury statystycznej są wyświetlane w oknie raportu. Wygenerowane wyniki mogą być tabelami statystycznymi, diagramami, wykresami lub tekstem, w zależności od wyboru dokonanego podczas wykonywania procedury. W tej sekcji używane są pliki *viewertut.spv* i *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

### Korzystanie z Edytora raportów



Rysunek 37. Okno raportu

Okno raportu jest podzielone na dwa panele. **Struktura raportu** zawiera schemat struktury informacji zapisanych w oknie raportu. **Zawartość raportu** zawiera tabele, wykresy i wyniki tekstowe.

Za pomocą pasków przewijania można poruszać się po zawartości okna, zarówno w pionie, jak i w poziomie. Aby ułatwić sobie nawigację, kliknij element w strukturze raportu, aby wyświetlić go w panelu zawartości.

1. Kliknij i przeciągnij prawą krawędź panelu struktury raportu, aby zmienić jego szerokość.

Ikona otwartej książki w panelu struktury raportu wskazuje, że element aktualnie jest widoczny w oknie raportu, chociaż może nie znajdować się w widocznej części panelu zawartości.

2. Aby ukryć tabelę lub wykres, kliknij dwukrotnie jego ikonę książki w strukturze raportu.

Ikona otwartej książki zmieni się w ikonę zamkniętej książki, oznaczającą, że powiązane z nią informacje są teraz ukryte.

3. Aby ponownie wyświetlić ukryty wynik, kliknij dwukrotnie ikonę zamkniętej książki.

Można również ukryć wszystkie wyniki z określonej procedury statystycznej lub wszystkie wyniki w oknie raportu.

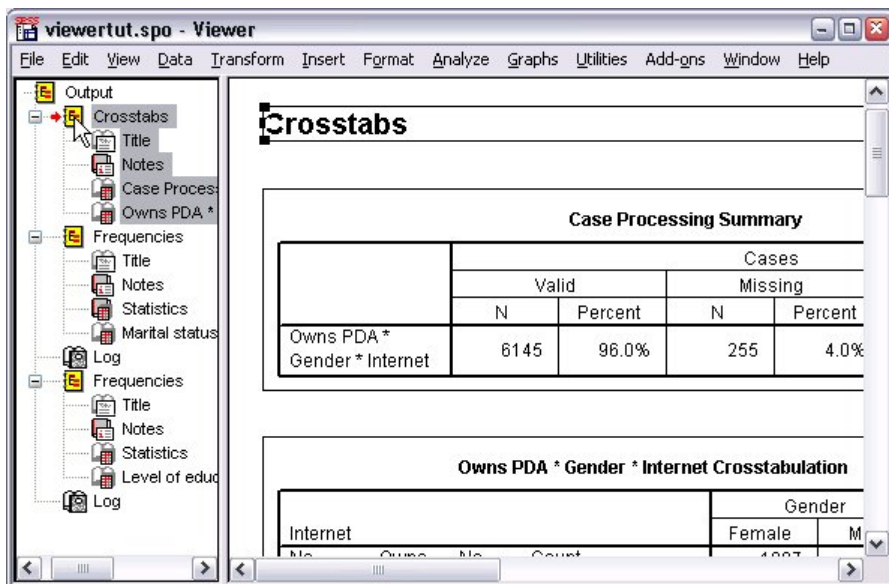
4. Kliknij pole ze znakiem minus (–) po lewej stronie procedury, której wyniki mają zostać ukryte, lub kliknij pole obok elementu najwyższego poziomu w strukturze raportu, aby ukryć wszystkie wyniki.

Struktura zwinie się, wskazując, że odpowiednie wyniki są ukryte.

Można również zmienić kolejność wyświetlania wyników.

5. W strukturze raportu kliknij elementy, które chcesz przenieść.

6. Przeciągnij wybrane elementy do nowego miejsca w strukturze.



Rysunek 38. Zmieniona kolejność wyników w oknie raportu

Elementy wyników można też przenosić, klikając i przeciągając je w panelu zawartości.

## Korzystanie z Edytora tabel przestawnych

Wyniki większości procedur statystycznych są wyświetlane w **tabelach przestawnych**.

### Ocena definicji wyników

Wyniki zawierają wiele terminów statystycznych. Definicje tych terminów można Wyświetlać bezpośrednio z okna raportu.

1. Kliknij dwukrotnie tabelę *Owns PDA \* Gender \* Internet Crosstabulation*.
2. Kliknij prawym przyciskiem pozycję *Liczebność oczekiwana* i z menu podręcznego wybierz opcję **Co to jest?**.

Definicja zostanie wyświetlona w oknie kontekstowym.

**Owns PDA \* Gender \* Internet Crosstabulation**

| Internet |                   |                | Gender |        | Total  |
|----------|-------------------|----------------|--------|--------|--------|
|          |                   |                | Female | Male   |        |
| No       | Owns PDA          | Count          | 1897   | 1962   | 3859   |
|          |                   | Expected Count | 1903.4 | 1955.6 | 3859.0 |
|          | % within Owns PDA | 50.3%          | 49.7%  | 100.0% |        |
|          | Total             | Count          | 2224   | 2285   | 4509   |
| Yes      | Owns PDA          | Count          | 513    | 512    | 1025   |
|          |                   | Expected Count | 512.5  | 512.5  | 1025.0 |
|          | % within Owns PDA | 50.0%          | 50.0%  | 100.0% |        |
|          | Total             | Count          | 305    | 306    | 611    |

Rysunek 39. Definicja w oknie kontekstowym

## Przestawianie tabel

Czytelność wygenerowanych tabel domyślnych bywa czasem niezadowolająca. Za pomocą tabel przestawnych można transponować wiersze i kolumny (zamienić je miejscami), zmienić kolejność danych w tabeli i zmodyfikować tabelę na wiele innych sposobów. Na przykład można zmienić krótką, szeroką tabelę w długą i wąską, transponując wiersze i kolumny. Zmiana układu tabeli nie wpływa na wyniki. Pozwala natomiast wyświetlić informacje w inny lub bardziej pożądany sposób.

1. Kliknij dwukrotnie tabelę *Owns PDA \* Gender \* Internet Crosstabulation*, aby ją aktywować, jeśli nie jest jeszcze aktywna.
2. Jeśli okno Panel przestawiania nie jest widoczne, z menu wybierz:

### Przestaw > Panel przestawiania

Panel przestawiania umożliwia przenoszenie danych między kolumnami, wierszami i warstwami.

| Internet |                   |                | Gender |        | Total  |
|----------|-------------------|----------------|--------|--------|--------|
|          |                   |                | Female | Male   |        |
| No       | Owns PDA          | Count          | 1897   | 1962   | 3859   |
|          |                   | Expected Count | 1903.4 | 1955.6 | 3859.0 |
|          | % within Owns PDA | 49.2%          | 50.8%  | 100.0% |        |
|          | Count             | 327            | 323    | 650    |        |
| Yes      | Owns PDA          | Count          | 513    | 512    | 1025   |
|          |                   | Expected Count | 512.5  | 512.5  | 1025.0 |
|          | % within Owns PDA | 50.0%          | 50.0%  | 100.0% |        |
|          | Count             | 305            | 306    | 611    |        |

Rysunek 40. Panel przestawiania

- Przeciągnij element *Statistics* z wymiaru Wiersz do wymiaru Kolumna, poniżej elementu *Gender*. Tabela zostanie natychmiast zrekonfigurowana, aby odzwierciedlić wprowadzone zmiany. Kolejność elementów w panelu przestawiania odzwierciedla kolejność elementów w tabeli.
- Przeciągnij i upuść element *Owens PDA* przed elementem *Internet* w wymiarze wierszy, aby zamienić kolejność tych dwóch wierszy.

|           |     |     | Gender |                    |        |                |                    | Total  |                    |
|-----------|-----|-----|--------|--------------------|--------|----------------|--------------------|--------|--------------------|
|           |     |     | Female |                    | Male   |                |                    |        |                    |
|           |     |     | Count  | % within Owens PDA | Count  | Expected Count | % within Owens PDA | Count  | % within Owens PDA |
| Owens PDA | No  | No  | 1897   | 49.2%              | 1903.4 |                |                    | 3859.0 | 100.0%             |
|           | Yes | Yes | 1025.0 | 100.0%             | 650.0  | 611.0          | 100.0%             | 4509.0 | 100.0%             |
| Total     | No  | Yes | 2822.0 |                    | 2553.4 | 1666.0         | 100.0%             | 5375.4 | 100.0%             |

Rysunek 41. Zamiana wierszy

## Tworzenie i wyświetlanie warstw

Warstwy mogą być przydatne w przypadku dużych tabel z zagnieżdżonymi kategoriami informacji. Tworząc warstwy, upraszcza się wygląd tabeli i poprawia jej czytelność.

- Przeciągnij element *Gender* z wymiaru Kolumna do wymiaru Warstwa.

|           |     |     | Gender |                |                    |
|-----------|-----|-----|--------|----------------|--------------------|
|           |     |     | Count  | Expected Count | % within Owens PDA |
| Owens PDA | No  | No  | 1897   | 1903.4         | 49.2%              |
|           | Yes | Yes |        |                |                    |
| Total     | No  | Yes |        |                |                    |

Rysunek 42. Ikona przestawiania Gender w wymiarze Warstwa

Aby wyświetlić inną warstwę, wybierz kategorię z listy rozwijanej w tabeli.

## Edytowanie tabel


Jeśli nie utworzono niestandardowego szablonu TableLook, tabele przestawne są tworzone z formatowaniem standardowym. Można zmienić format dowolnego tekstu w tabeli. Elementy

formatowania, które można zmieniać, obejmują nazwę czcionki, wielkość czcionki, styl czcionki (pogrubienie lub kursywa) oraz kolor.

1. Kliknij dwukrotnie tabelę *Level of education*.
2. Jeśli pasek narzędzi Formatowanie nie jest widoczny, z menu wybierz:

**Widok > Pasek narzędzi**

3. Kliknij tekst tytułu, *Level of education*.
4. Z listy rozwijanej wielkości czcionki na pasku narzędzi wybierz wartość **12**.
5. Aby zmienić kolor tekstu tytułu, kliknij narzędzie do zmiany koloru tekstu i wybierz nowy kolor.



The screenshot shows the 'Formatting Toolbar' with the font set to 'Arial', size '12', and bold, italic, underline, and color options. Below the toolbar is a table with the title 'Level of education' in red, bold font.

|       |                              | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|------------------------------|-----------|---------|---------------|--------------------|
| Valid | Did not complete high school | 1390      | 21.7    | 21.7          | 21.7               |
|       | High school degree           | 1936      | 30.3    | 30.3          | 52.0               |
|       | Some college                 | 1360      | 21.3    | 21.3          | 73.2               |
|       | College degree               | 1355      | 21.2    | 21.2          | 94.4               |
|       | Post-undergraduate degree    | 359       | 5.6     | 5.6           | 100.0              |
|       | Total                        | 6400      | 100.0   | 100.0         |                    |

Rysunek 43. Przeformatowany tekst tytułu w tabeli przestawnej

Można także edytować zawartość tabel i etykiet. Na przykład można zmienić tytuł tej tabeli.

6. Kliknij dwukrotnie tytuł.
7. Wpisz Poziom wykształcenia jako nową etykietę.

*Uwaga:* Jeśli zmienisz wartości w tabeli, podsumowania i inne statystyki nie zostaną ponownie obliczone.

## Ukrywanie wierszy i kolumn

Niektóre z danych w tabeli mogą nie być użyteczne lub niepotrzebnie komplikować tabelę. Na szczęście można ukryć całe wiersze i kolumny bez utraty danych.

1. Kliknij dwukrotnie tabelę *Education Level*, aby ją aktywować, jeśli nie jest jeszcze aktywna.
2. Kliknij etykietę kolumny *Valid Percent*, aby ją zaznaczyć.
3. W menu Edycja lub w menu podręcznym dostępnym po kliknięciu prawym przyciskiem myszy wybierz:

**Wybierz > Komórki danych i opis**

4. Z menu Widok wybierz opcję **Ukryj** lub kliknij prawym przyciskiem myszy i z menu podręcznego wybierz opcję **Ukryj kategorię**.

Kolumna jest teraz ukryta, ale nie została usunięta.

**Education Level**

|       |                              | Frequency | Percent | Cumulative Percent |
|-------|------------------------------|-----------|---------|--------------------|
| Valid | Did not complete high school | 1390      | 21.7    | 21.7               |
|       | High school degree           | 1936      | 30.3    | 52.0               |
|       | Some college                 | 1360      | 21.3    | 73.2               |
|       | College degree               | 1355      | 21.2    | 94.4               |
|       | Post-undergraduate degree    | 359       | 5.6     | 100.0              |
|       | Total                        | 6400      | 100.0   |                    |

Rysunek 44. Kolumna *Valid Percent* ukryta w tabeli

Aby ponownie wyświetlić kolumnę:

5. Z menu wybierz:

## Widok > Pokaż wszystko

Wiersze można ukrywać i uwidaczniać w taki sam sposób, jak kolumny.

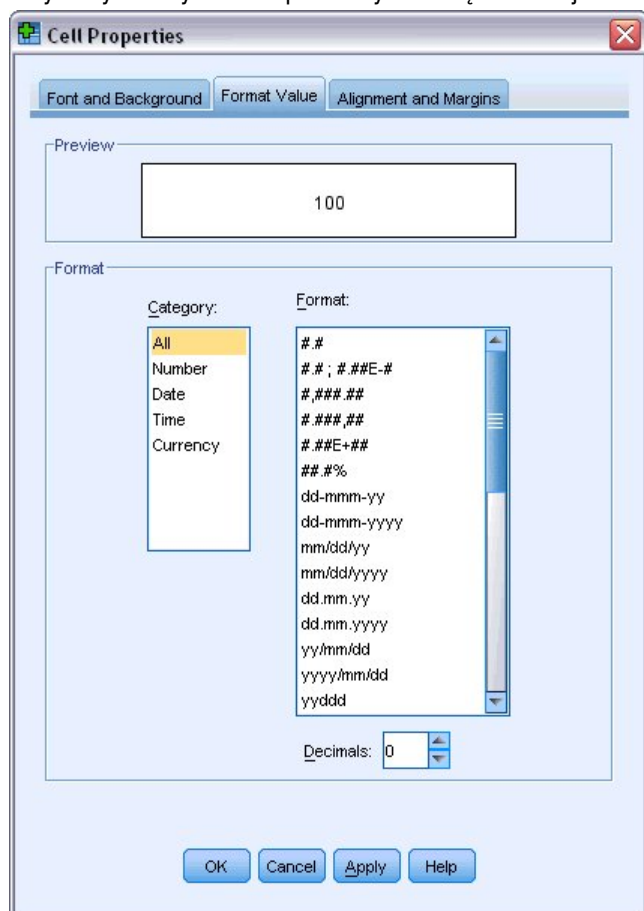
## Zmiana formatów wyświetlania danych

Można łatwo zmieniać format wyświetlania danych w tabelach przestawnych.

1. Kliknij dwukrotnie tabelę *Education Level*, aby ją aktywować, jeśli nie jest jeszcze aktywna.
2. Kliknij etykietę kolumny *Percent*, aby ją zaznaczyć.
3. W menu Edycja lub w menu podręcznym dostępnym po kliknięciu prawym przyciskiem myszy wybierz:

### Wybierz pozycję > Komórki danych

4. Z menu Format lub z menu wywoływanego prawym przyciskiem myszy wybierz opcję **Właściwości komórki**.
5. Kliknij kartę **Format wartości**.
6. Aby ukryć wszystkie separatory dziesiętne w tej kolumnie, wpisz 0 w polu Miejsca dziesiętne.



Rysunek 45. Właściwości komórki, karta Format wartości

W tym oknie dialogowym można również zmienić typ i format danych.

7. Wybierz typ z listy Kategoria, a następnie wybierz format dla tego typu na liście Format.
8. Kliknij przycisk **OK** lub **Zastosuj**, aby wprowadzić zmiany.

## Education Level

|       |                              | Frequency | Percent | Cumulative Percent |
|-------|------------------------------|-----------|---------|--------------------|
| Valid | Did not complete high school | 1390      | 22      | 21.7               |
|       | High school degree           | 1936      | 30      | 52.0               |
|       | Some college                 | 1360      | 21      | 73.2               |
|       | College degree               | 1355      | 21      | 94.4               |
|       | Post-undergraduate degree    | 359       | 6       | 100.0              |
| Total |                              | 6400      | 100     |                    |

Rysunek 46. Separatory dziesiętne ukryte w kolumnie Percent

Separatory dziesiętne w kolumnie *Percent* są teraz ukryte.

## Szablony TableLook

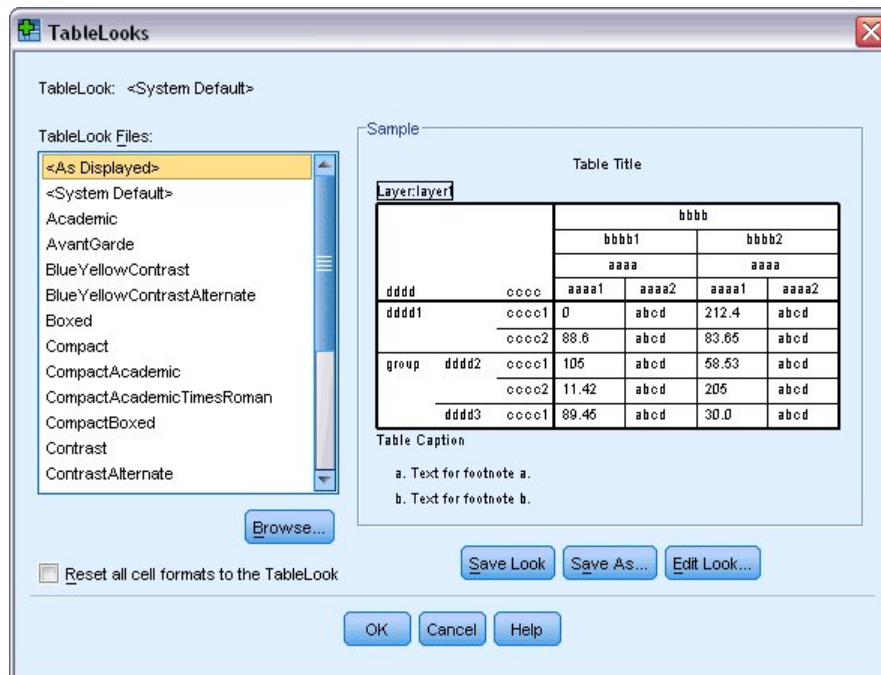
Format tabel ma zasadnicze znaczenie dla czytelności, zwięzłości i użyteczności wyników. Jeśli tabela jest trudna do odczytania, informacje w niej zawarte będą trudne do interpretacji.

### Korzystanie z formatów predefiniowanych

1. Kliknij dwukrotnie tabelę *Marital status*.
2. Z menu wybierz:

**Format > Szablony TableLook...**

Okno dialogowe Szablony TableLook zawiera listę różnych predefiniowanych stylów. Wybierz styl z listy, aby wyświetlić jego podgląd w oknie Przykład po prawej stronie.



Rysunek 47. Okno dialogowe Szablony TableLook

Można użyć stylu bez zmian lub edytować istniejący styl, aby lepiej dostosować go do swoich potrzeb.

3. Aby użyć istniejącego stylu, wybierz jeden z nich i kliknij przycisk **OK**.

### Dostosowywanie stylów TableLook

Istnieje możliwość dostosowania formatu do konkretnych potrzeb. Można dostosować prawie wszystkie aspekty tabeli, od koloru tła po style obramowania.



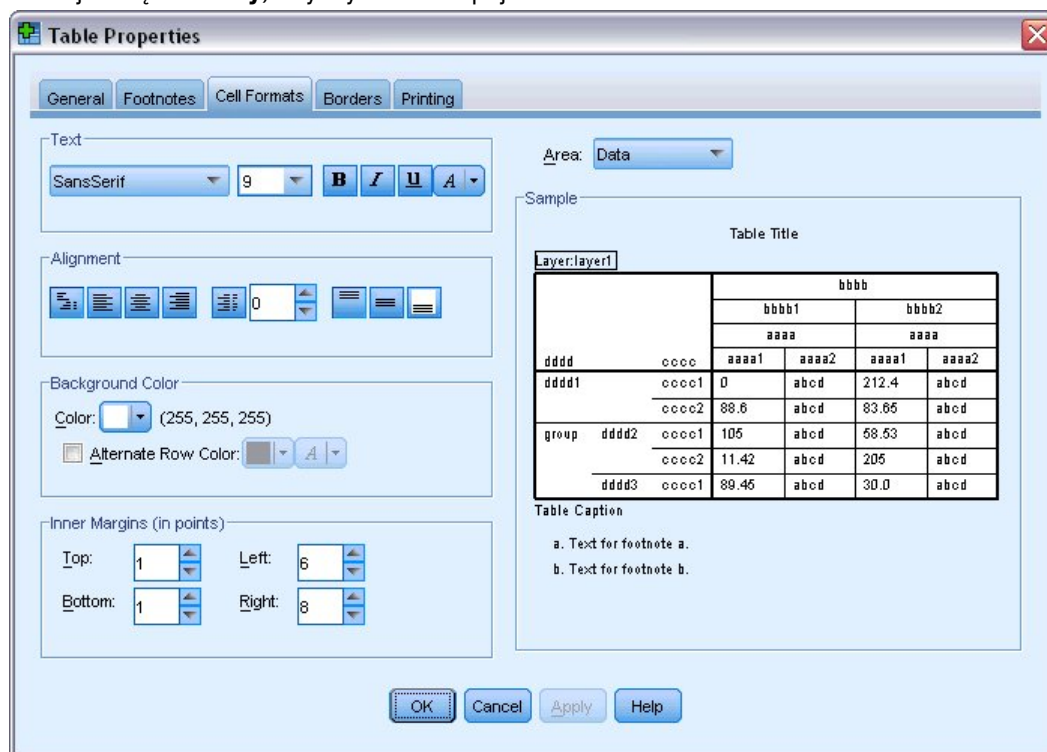
1. Kliknij dwukrotnie tabelę *Marital status*.

2. Z menu wybierz:

**Format > Szablony TableLook...**

3. Wybierz styl, który jest najbliższy żadanemu formatowi, a następnie kliknij opcję **Edytuj wzorzec**.

4. Kliknij kartę **Formaty**, aby wyświetlić opcje formatowania.



Rysunek 48. Okno dialogowe Właściwości tabeli

Opcje formatowania obejmują nazwę czcionki, rozmiar czcionki, styl i kolor. Dodatkowe opcje to wyrównanie, kolor tekstu i tła oraz wielkości marginesów.

Okno Przykład z prawej strony umożliwia podgląd wpływu zmian formatowania na tabelę. Każdy obszar tabeli może mieć inny styl formatowania. Na przykład zwykle tytuł powinien mieć inny styl niż dane. Aby wybrać obszar tabeli do edycji, można wybrać jego nazwę na liście rozwijanej Obszar lub kliknąć obszar, który ma zostać zmieniony, w oknie Przykład.

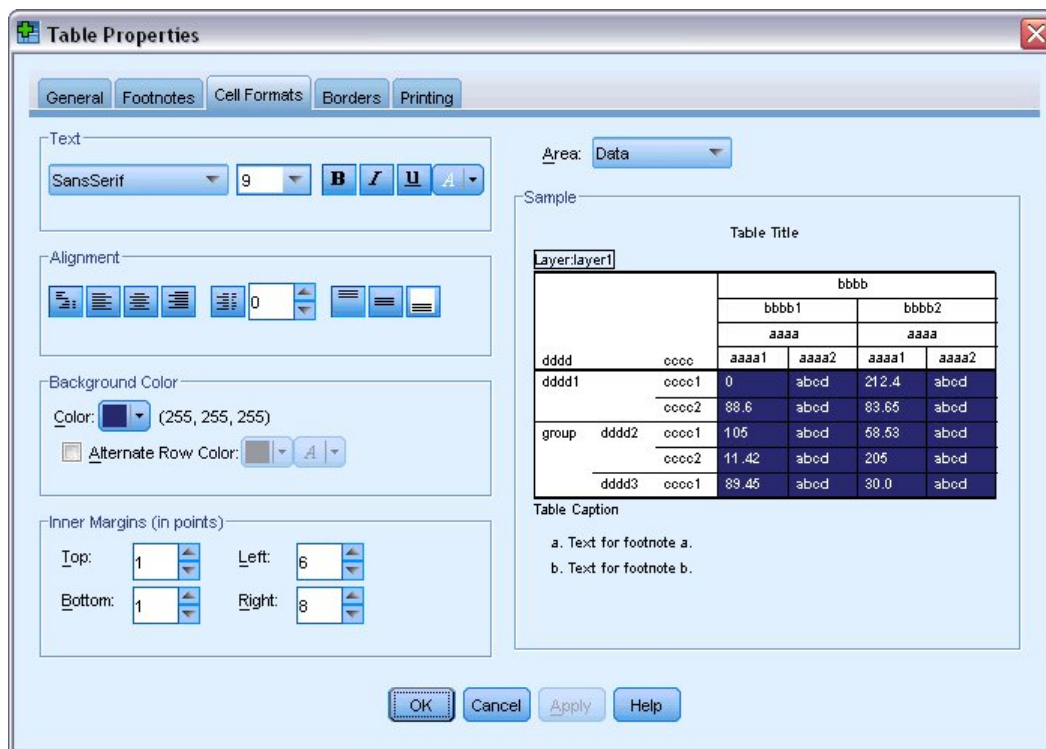
5. Z rozwijanej listy Obszar wybierz opcję **Dane**.

6. Wybierz nowy kolor z palety rozwijanej Tło.

7. Następnie wybierz nowy kolor tekstu.

Nowy styl zostanie uwzględniony w oknie Przykład.





Rysunek 49. Zmiana formatów komórek tabeli

8. Kliknij przycisk **OK**, aby powrócić do okna dialogowego Szablony TableLook.

Można zapisać nowy styl, co pozwala na łatwe zastosowanie go do innych tabel w przyszłości.

9. Kliknij opcję **Zapisz jako**.

10. Przejdź do katalogu docelowego i wprowadź nazwę nowego stylu w polu tekstowym Nazwa pliku.

11. Kliknij przycisk **Zapisz**.

12. Kliknij przycisk **OK**, aby zastosować zmiany i powrócić do okna raportu.

Tabela jest teraz wyświetlana zgodnie z określonym formatem niestandardowym.

|       |           | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-----------|-----------|---------|---------------|--------------------|
| Valid | Unmarried | 3224      | 50.4    | 50.4          | 50.4               |
|       | Married   | 3176      | 49.6    | 49.6          | 100.0              |
|       | Total     | 6400      | 100.0   | 100.0         |                    |

Rysunek 50. Niestandardowy szablon TableLook

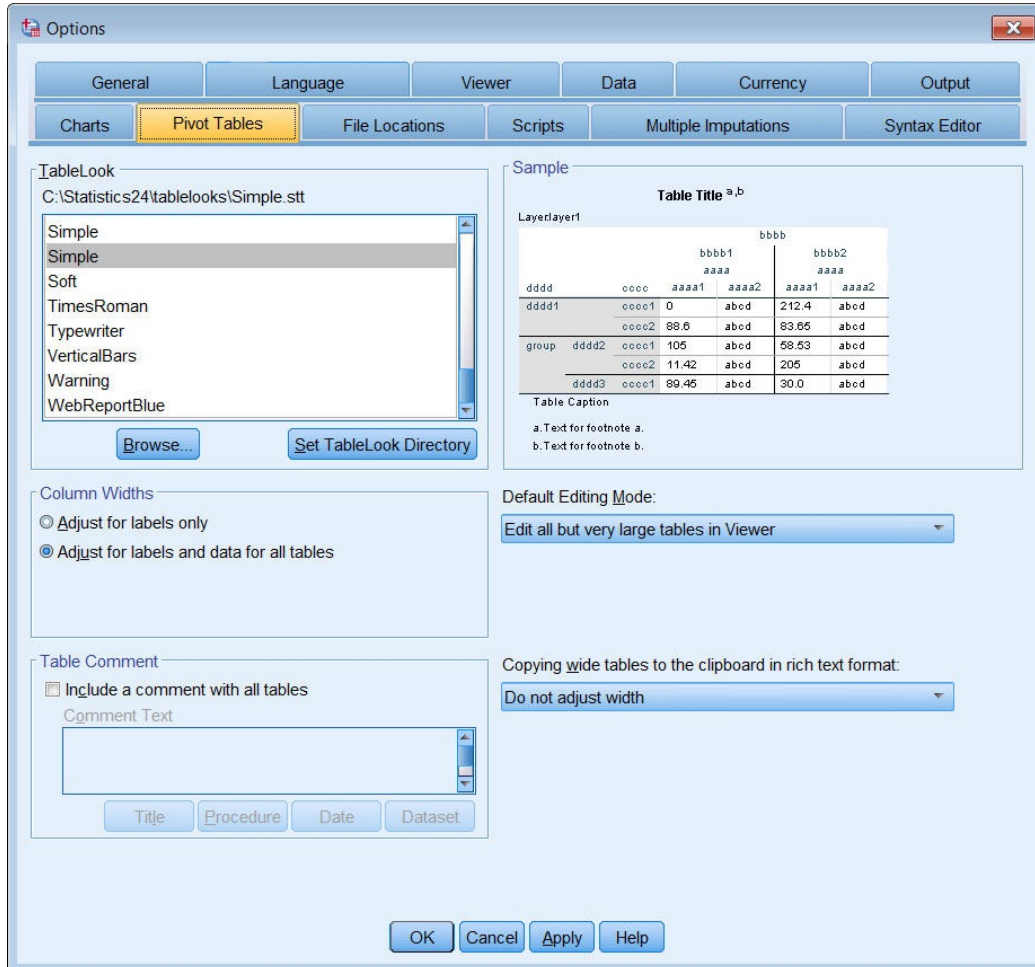
## Zmiana domyślnych formatów tabel

Mimo że można zmienić format tabeli po jej utworzeniu, bardziej efektywne może okazać się zmodyfikowanie domyślnego szablonu TableLook, aby nie trzeba było zmieniać formatu za każdym razem, gdy tworzona jest tabela.

Aby zmienić domyślny styl TableLook tabel przestawnych, z menu wybierz:

### Edycja > Opcje...

1. Kliknij kartę **Tabele przestawne** w oknie dialogowym Opcje.



Rysunek 51. Okno dialogowe Opcje

2. Wybierz styl TableLook, który ma być używany dla wszystkich nowych tabel.

W oknie Przykład po prawej stronie wyświetlany jest podgląd każdego z szablonów TableLook.

3. Kliknij przycisk **OK**, aby zapisać ustawienia i zamknąć okno dialogowe.

Wszystkie tabele tworzone po zmianie domyślnego szablonu TableLook będą zgodne z nowymi regułami formatowania.

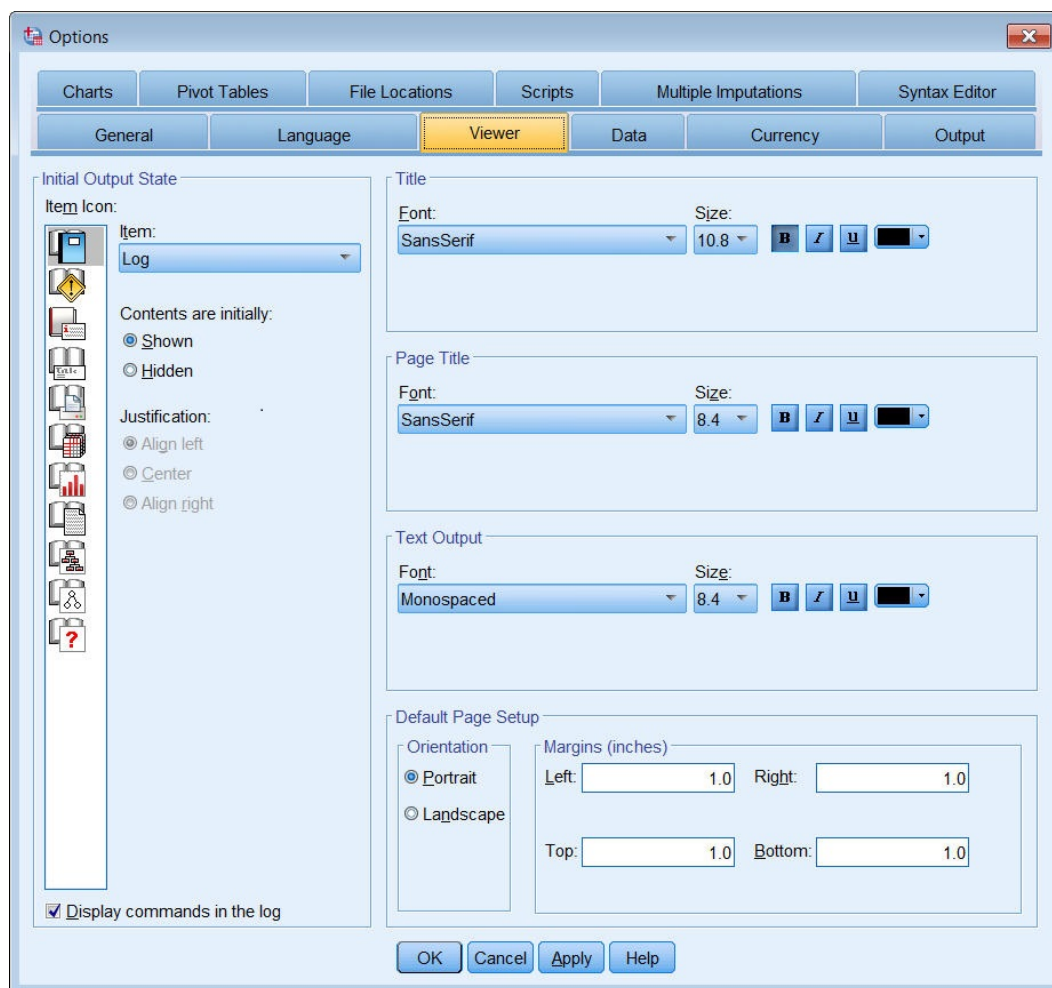
## Dostosowywanie początkowych ustawień wyświetlania

Początkowe ustawienia wyświetlania obejmują sposób wyrównania obiektów w oknie raportu, domyślną widoczność obiektów, a także szerokość okna raportu. Aby zmienić te ustawienia:

1. Z menu wybierz:

### Edycja > Opcje...

2. Kliknij kartę **Raporty**.



Rysunek 52. Opcje okna raportu

Ustawienia są stosowane do każdego obiektu z osobna. Na przykład można dostosować sposób wyświetlania wykresów bez wprowadzania zmian w sposobie wyświetlania tabel. Wystarczy wybrać obiekt, który ma zostać dostosowany, i wprowadzić zmiany.

3. Kliknij ikonę **Tytuł**, aby wyświetlić jego ustawienia.
4. Kliknij opcję **Wyśrodkuj**, aby wyśrodkować (poziomo) wszystkie tytuły w oknie raportu.

Można także ukrywać elementy, takie jak komunikaty dziennika i ostrzeżenia, które pogarszają czytelność wyników. Dwukrotne kliknięcie ikony automatycznie zmienia właściwość wyświetlania obiektu.

5. Kliknij dwukrotnie ikonę **Ostrzeżenia**, aby ukryć komunikaty ostrzegawcze w wynikach.
6. Kliknij przycisk **OK**, aby zapisać zmiany i zamknąć okno dialogowe.

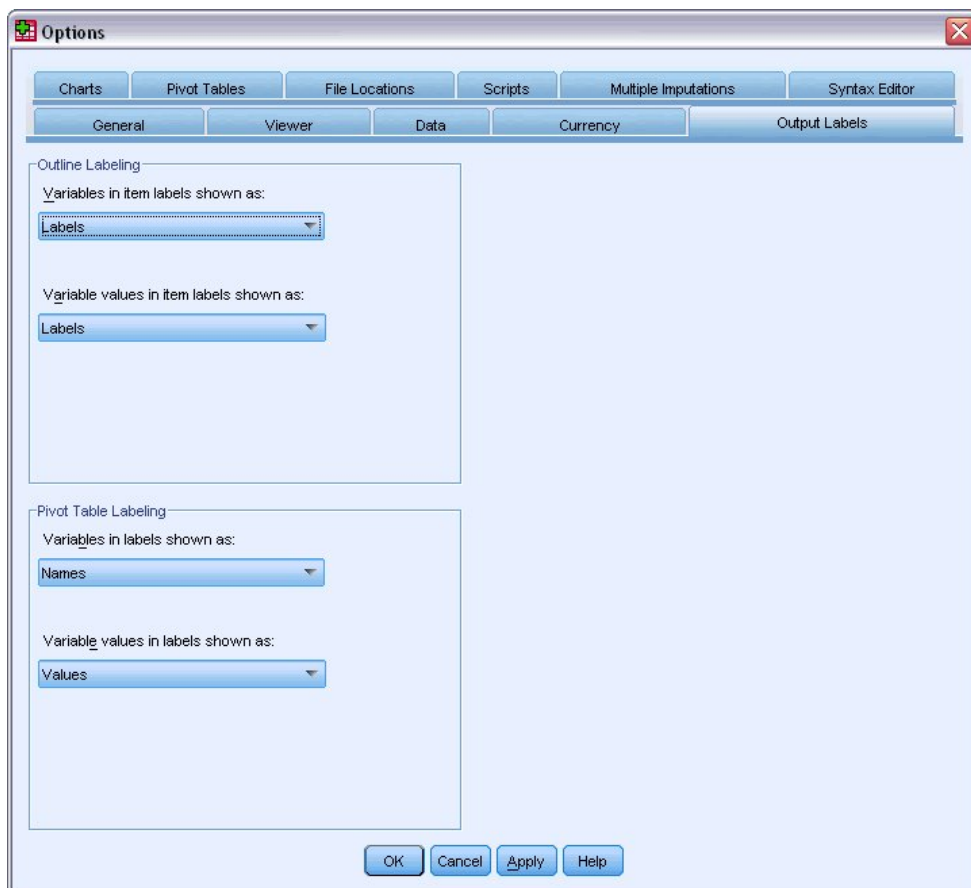
## Wyświetlanie etykiet zmiennych i wartości

W większości przypadków wyświetlanie etykiet zmiennych i wartości jest bardziej efektywne niż wyświetlanie nazw zmiennej lub rzeczywistych wartości danych. Mogą jednak wystąpić przypadki, w których pożądanym jest wyświetlanie zarówno nazwy, jak i etykiety.

1. Z menu wybierz:

**Edycja > Opcje...**

2. Kliknij kartę **Opis wyników**.



Rysunek 53. Ustawienia etykiet tabel przestawnych

Istnieje możliwość określenia różnych ustawień dla struktury i zawartości raportu. Aby na przykład wyświetlić etykiety w strukturze, a nazwy i wartości zmiennych w zawartości raportu:

3. W grupie Opisywanie tabel przestawnych wybierz opcję **Nazwy** z listy rozwijanej Zmienne w opisach, aby wyświetlać nazwy zmiennych, a nie etykiety.
4. Następnie wybierz opcję **Wartości** z listy rozwijanej Wartości zmiennych w opisach, aby wyświetlać wartości danych zamiast etykiet.

Kolejne tabele utworzone w bieżącej sesji będą odzwierciedlać te zmiany.

**marital**

|       |       | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------|-----------|---------|---------------|--------------------|
| Valid | 0     | 3224      | 50.4    | 50.4          | 50.4               |
|       | 1     | 3176      | 49.6    | 49.6          | 100.0              |
|       | Total | 6400      | 100.0   | 100.0         |                    |

Rysunek 54. Wyświetlone nazwy i wartości zmiennych

## Korzystanie z wyników w innych aplikacjach

Wyniki mogą być używane w wielu różnych aplikacjach. Na przykład można uwzględnić tabelę lub wykres w prezentacji lub raporcie.

Poniższe przykłady dotyczą programu Microsoft Word, ale mogą działać podobnie w innych aplikacjach do przetwarzania tekstu.

## Wklejanie wyników jako tabel do programu Word

Tabele przestawne można wklejać do programu Word jako rodzime tabele tego programu. Wszystkie atrybuty tabeli, takie jak wielkości i kolory czcionek, są zachowywane. Ponieważ tabela jest wklejana w formacie programu Word, można edytować ją w programie Word, tak jak każdą inną tabelę.

1. Kliknij tabelę w oknie raportu, aby ją zaznaczyć.
2. Z menu wybierz:

**Edytuj > Kopiuj**

3. Otwórz aplikację do przetwarzania tekstu.
4. Z menu procesora tekstu wybierz:

**Edycja > Wklej specjalnie...**

5. W oknie dialogowym Wklej specjalnie wybierz opcję **Tekst sformatowany (RTF)**.
6. Kliknij przycisk **OK**, aby wkleić wyniki do bieżącego dokumentu.

Tabela jest teraz wyświetlana w dokumencie. Można stosować formatowanie niestandardowe, edytować dane i zmienić rozmiar tabeli odpowiednio do potrzeb.

## Wklejanie wyników jako tekstu

Tabele przestawne można kopiować do innych aplikacji jako zwykły tekst. W przypadku tej metody style formatowania nie są zachowywane, ale można edytować dane tabeli po wklejeniu jej do aplikacji docelowej.

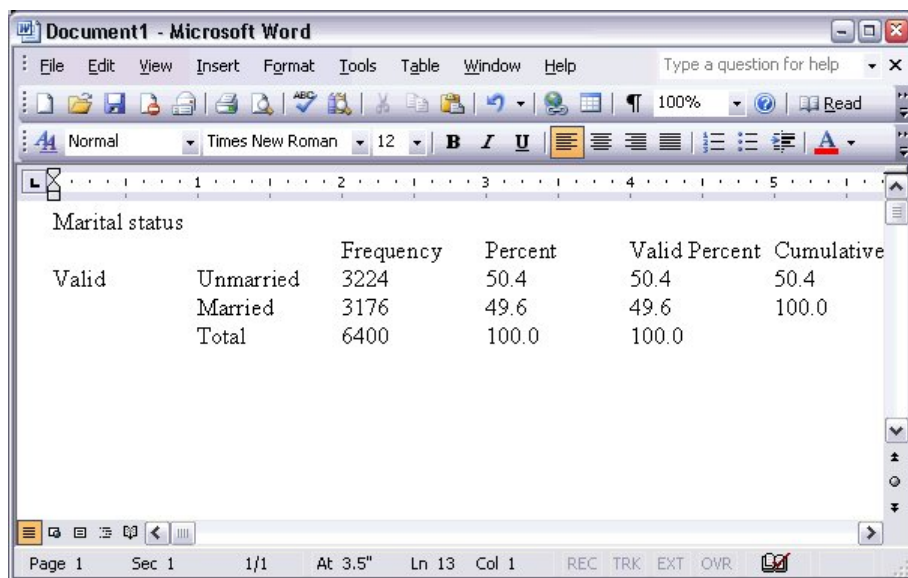
1. Kliknij tabelę w oknie raportu, aby ją zaznaczyć.
2. Z menu wybierz:

**Edytuj > Kopiuj**

3. Otwórz aplikację do przetwarzania tekstu.
4. Z menu procesora tekstu wybierz:

**Edycja > Wklej specjalnie...**

5. W oknie dialogowym Wklej specjalnie wybierz opcję **Tekst niesformatowany**.
6. Kliknij przycisk **OK**, aby wkleić wyniki do bieżącego dokumentu.



The screenshot shows a Microsoft Word window titled 'Document1 - Microsoft Word'. The menu bar includes File, Edit, View, Insert, Format, Tools, Table, Window, and Help. The ribbon shows the 'Format' tab with options for font (Normal, Times New Roman, 12), bold (B), italic (I), underline (U), and text color. The main content area displays a pivot table with the following data:

| Marital status |           | Frequency | Percent | Valid Percent | Cumulative |
|----------------|-----------|-----------|---------|---------------|------------|
| Valid          | Unmarried | 3224      | 50.4    | 50.4          | 50.4       |
|                | Married   | 3176      | 49.6    | 49.6          | 100.0      |
| Total          |           | 6400      | 100.0   | 100.0         |            |

The status bar at the bottom indicates Page 1, Sec 1, 1/1, At 3.5", Ln 13, Col 1, and includes icons for REC, TRK, EXT, and OVR.

Rysunek 55. Tabela przestawna wyświetlana w programie Word

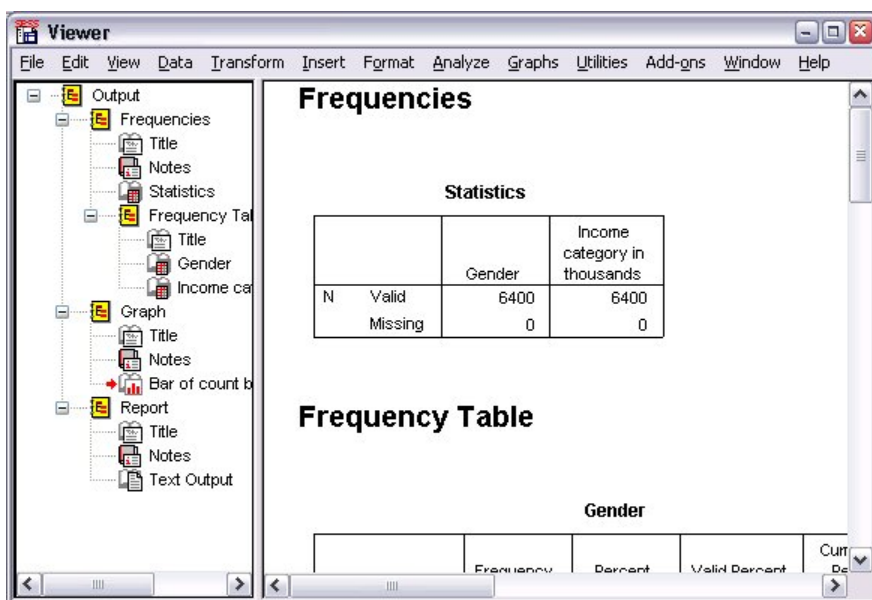
Kolumny w tabeli są oddzielone znakami tabulacji. Szerokość kolumn można zmienić, modyfikując tabulatory w aplikacji do przetwarzania tekstu.

## Eksportowanie wyników do plików programów Microsoft Word, PowerPoint i Excel

Wyniki można wyeksportować do pliku programu Microsoft Word, PowerPoint lub Excel. Można wyeksportować wybrane elementy lub wszystkie elementy z okna raportu. W tej sekcji używane są pliki *msoutut.spv* i *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

*Uwaga:* Eksport do programu PowerPoint jest dostępny tylko w systemach operacyjnych Windows i nie jest dostępny w wersji Student.

W strukturze raportu w oknie raportu można wybrać konkretne elementy, które mają zostać wyeksportowane, lub wyeksportować wszystkie elementy bądź wszystkie widoczne elementy.

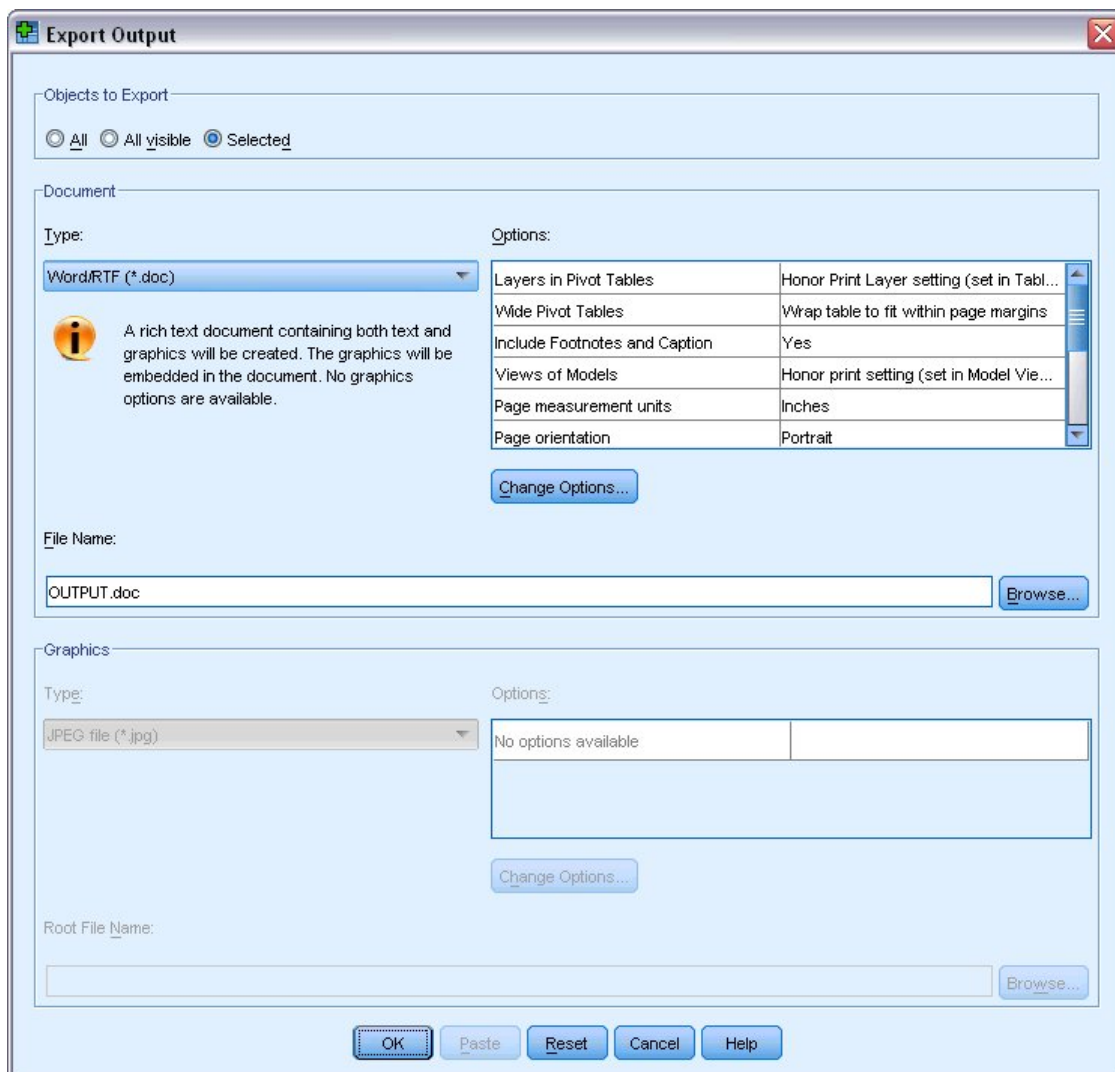


Rysunek 56. Okno raportu

1. Z menu Edytora raportów wybierz kolejno następujące pozycje:

### **Plik > Eksportuj...**

Zamiast eksportować wszystkie obiekty z okna raportu, można wybrać opcję eksportowania tylko widocznych obiektów (z otwartymi książkami w strukturze) lub tych, które zostały wybrane w strukturze raportu. Jeśli w strukturze nie wybrano żadnych elementów, nie ma możliwości wyeksportowania wybranych obiektów.



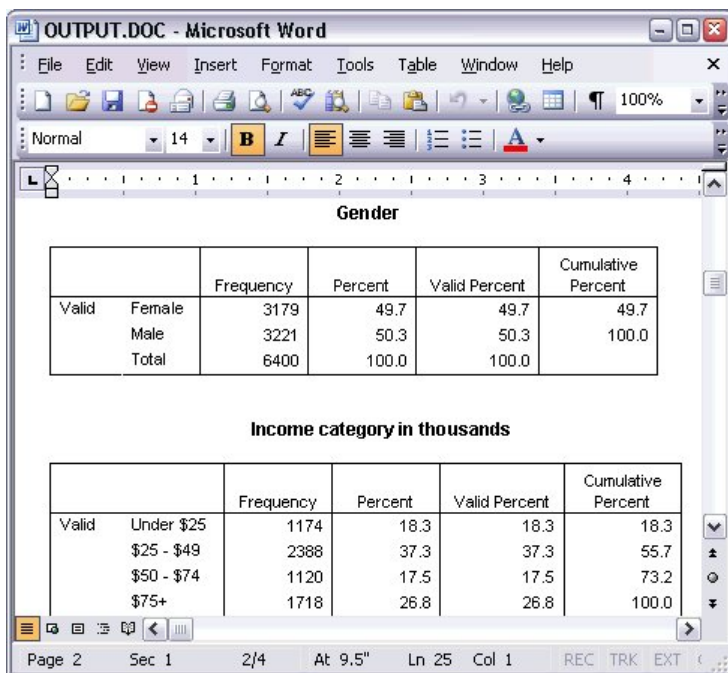
Rysunek 57. Okno dialogowe Eksportuj raport wynikowy

2. W grupie Obiekty do wyeksportowania wybierz opcję **Wszystkie**.
3. Z listy rozwijanej Typ wybierz opcję **Plik Word/RTF (\*.doc)**.
4. Kliknij przycisk **OK**, aby wygenerować plik w formacie programu Word.

Po otwarciu wynikowego pliku w programie Word można zobaczyć, w jaki sposób zostały wyeksportowane wyniki. Przypisy, które nie są obiektami widocznymi, są widoczne w programie Word, ponieważ wybrano opcję eksportu wszystkich obiektów.

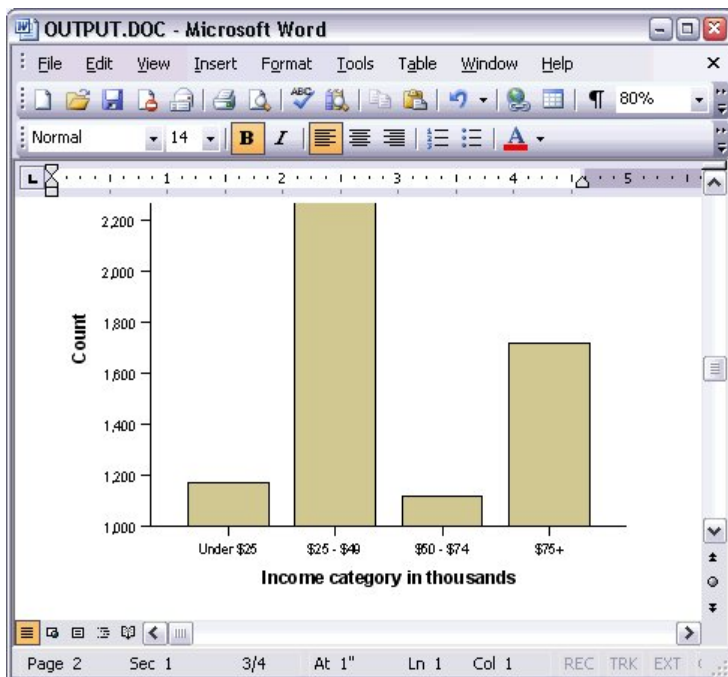
Tabele przestawne stają się tabelami programu Word, z całym formatowaniem pierwotnej tabeli przestawnej, w tym czcionkami, kolorami, obramowaniem itp.





Rysunek 58. Tabele przestawne w programie Word

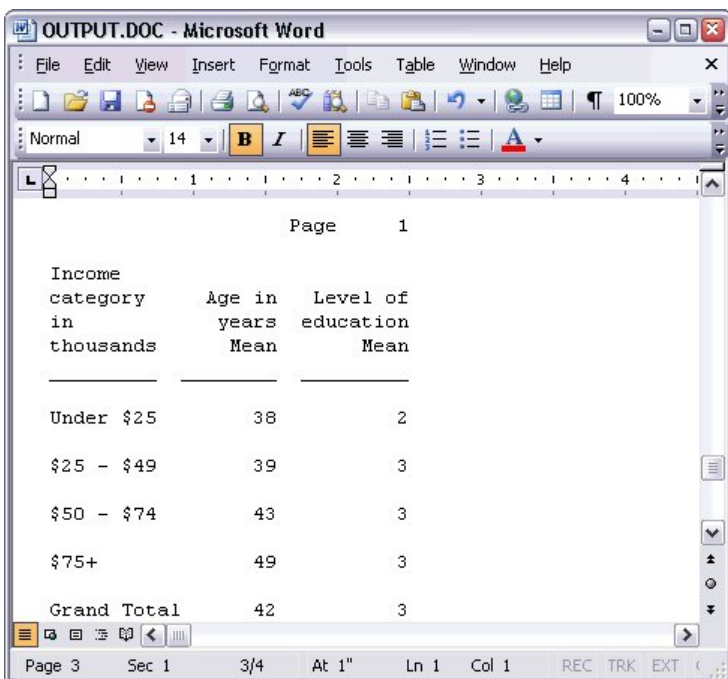
Wykresy są umieszczane w dokumencie programu Word jako obrazy graficzne.



Rysunek 59. Wykresy w programie Word

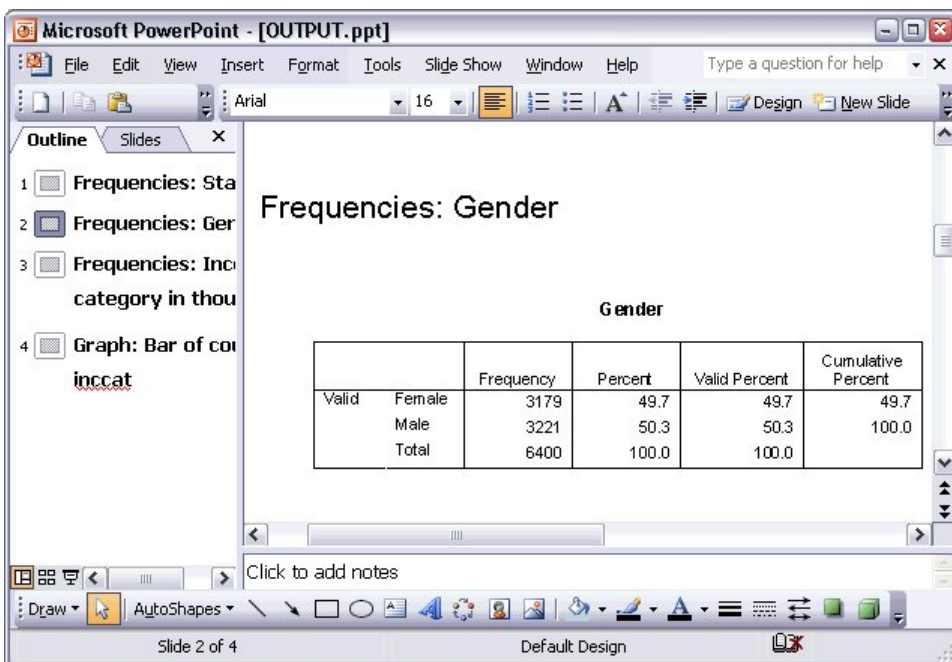
Wyniki tekstowe są wyświetlane tą samą czcionką, co używana dla obiektu tekstowego w oknie raportu. W celu zapewnienia właściwego wyrównania tekstu powinien być zapisany czcionką o stałej szerokości.





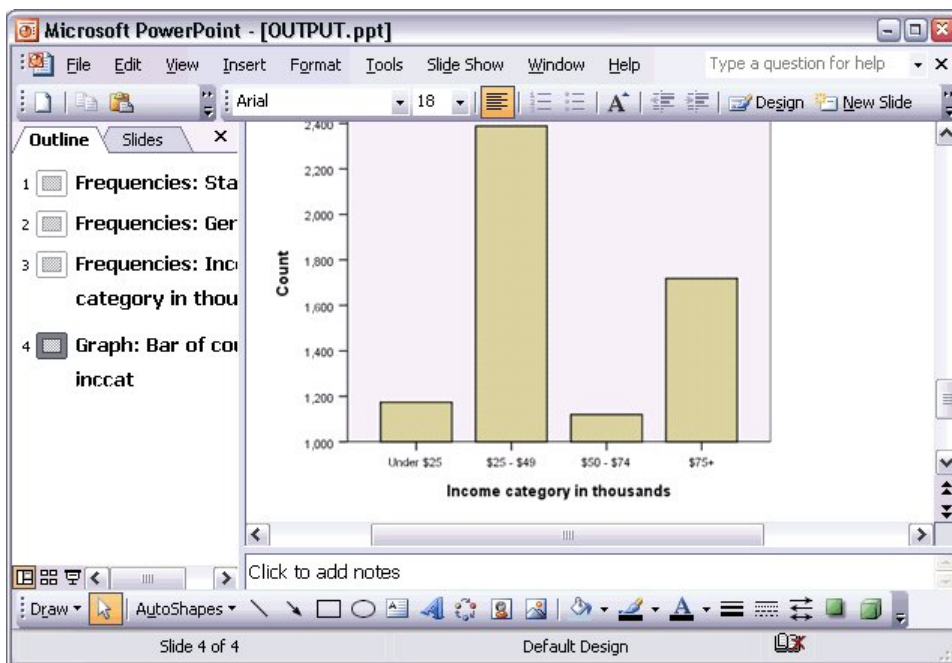
Rysunek 60. Wyniki tekstowe w programie Word

W przypadku eksportu do pliku programu PowerPoint każdy wyeksportowany element jest umieszczany na osobnym slajdzie. Tabele przestawne wyeksportowane do programu PowerPoint stają się tabelami programu Word, z całym formatowaniem pierwotnej tabeli przestawnej, w tym czcionkami, kolorami, obramowaniem itp.



Rysunek 61. Tabele przestawne w programie PowerPoint

Wykresy wybrane do eksportu do programu PowerPoint są osadzone w pliku PowerPoint.



Rysunek 62. Wykresy w programie PowerPoint

*Uwaga:* Eksport do programu PowerPoint jest dostępny tylko w systemach operacyjnych Windows i nie jest dostępny w wersji Student.

W przypadku eksportu do pliku programu Excel wyniki są eksportowane w inny sposób.

|       |        | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|--------|-----------|---------|---------------|--------------------|
| Valid | Female | 3,179     | 49.7    | 49.7          | 49.7               |
|       | Male   | 3,221     | 50.3    | 50.3          | 100.0              |
|       | Total  | 6,400     | 100.0   | 100.0         |                    |

|       |             | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|-------------|-----------|---------|---------------|--------------------|
| Valid | Under \$25  | 1,174     | 18.3    | 18.3          | 18.3               |
|       | \$25 - \$49 | 2,388     | 37.3    | 37.3          | 55.7               |
|       | \$50 - \$74 | 1,120     | 17.5    | 17.5          | 73.2               |
|       | \$75+       | 1,718     | 26.8    | 26.8          | 100.0              |
|       | Total       | 6,400     | 100.0   | 100.0         |                    |

Rysunek 63. Plik Output.xls w programie Excel

Wiersze, kolumny i komórki tabeli przestawnej stają się wierszami programu, kolumnami i komórkami arkusza programu Excel.

Każdy wiersz wyniku tekstowego stanowi wiersz w pliku Excel, przy czym zawartość całego wiersza znajduje się w pojedynczej komórce.

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - OUTPUT.xls". The spreadsheet contains the following data:

|    | A           | B      | C         | D | E |
|----|-------------|--------|-----------|---|---|
| 68 | Income      |        |           |   |   |
| 69 | category    | Age in | Level of  |   |   |
| 70 | in          | years  | education |   |   |
| 71 | thousands   | Mean   | Mean      |   |   |
| 72 |             |        |           |   |   |
| 73 |             |        |           |   |   |
| 74 | Under \$25  | 38     | 2         |   |   |
| 75 |             |        |           |   |   |
| 76 | \$25 - \$49 | 39     | 3         |   |   |
| 77 |             |        |           |   |   |
| 78 | \$50 - \$74 | 43     | 3         |   |   |
| 79 |             |        |           |   |   |
| 80 | \$75+       | 49     | 3         |   |   |
| 81 |             |        |           |   |   |
| 82 | Grand Total | 42     | 3         |   |   |

Rysunek 64. Wyniki tekstowe w programie Excel

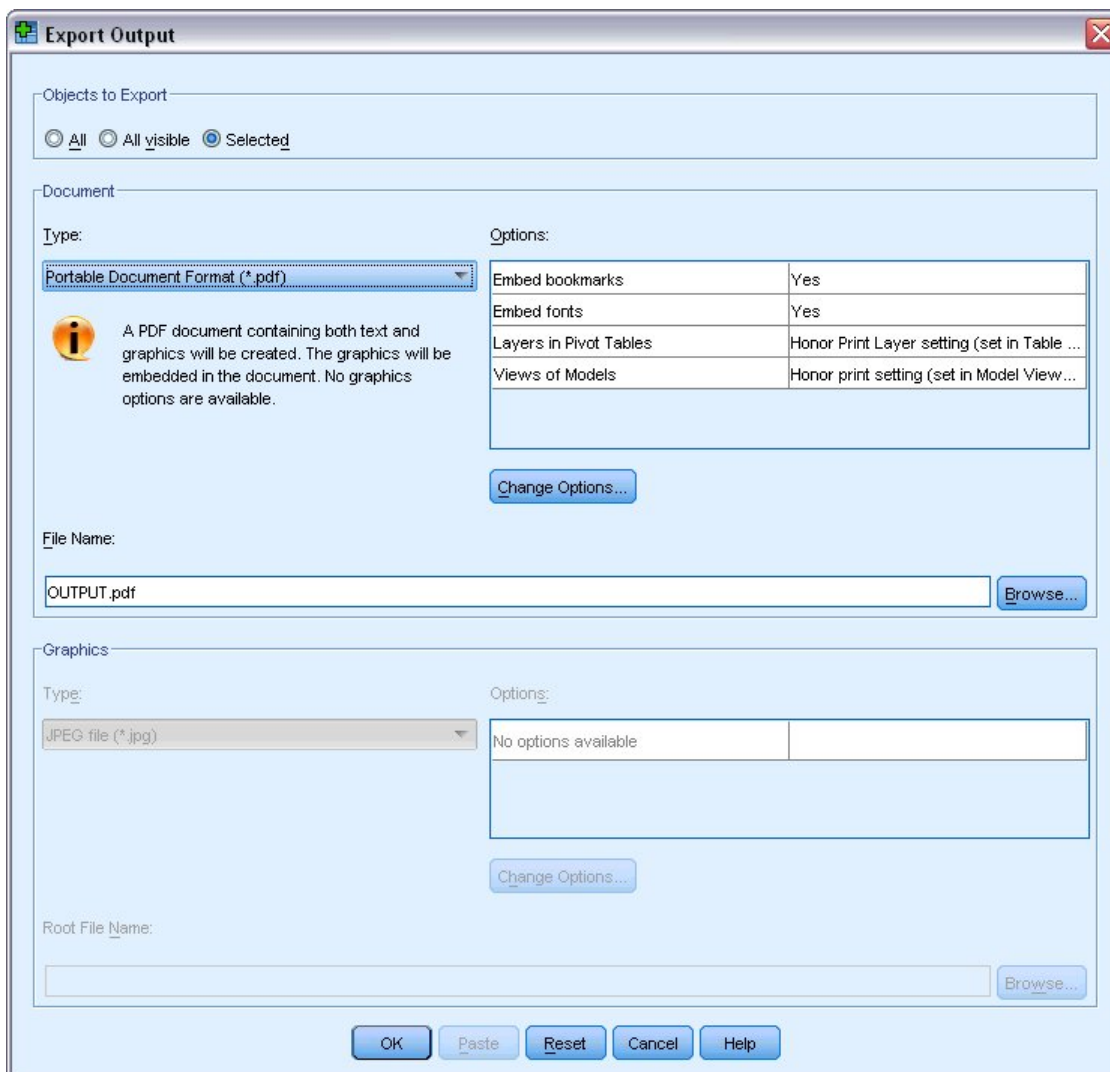
## Eksportowanie wyników do pliku PDF

Istnieje możliwość wyeksportowania wszystkich lub wybranych elementów z okna raportów do pliku PDF.

1. Z menu w oknie raportu zawierającego wynik, który chcesz wyeksportować do pliku PDF, wybierz:

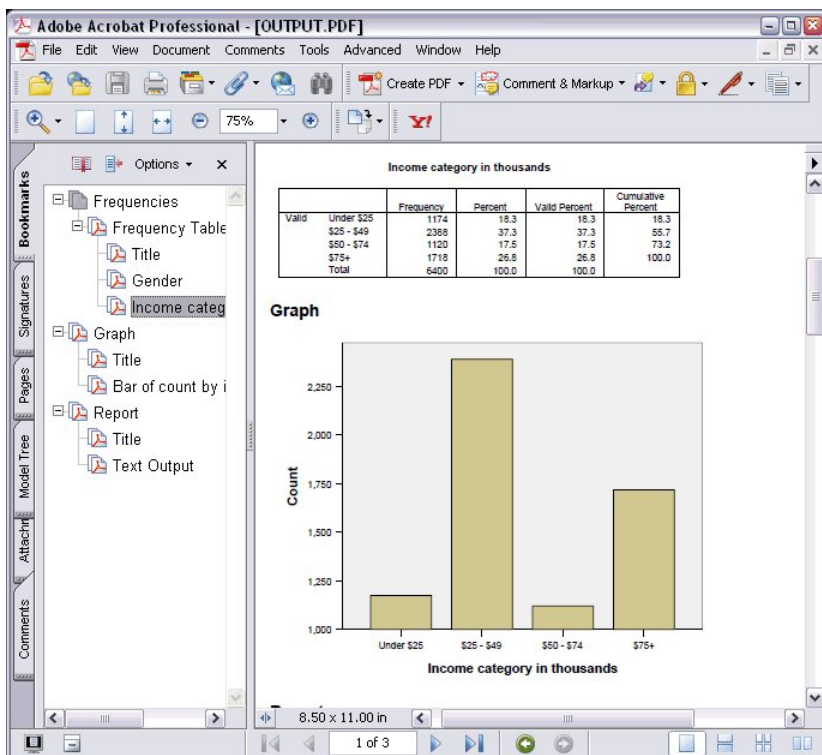
**Plik > Eksportuj...**

2. W oknie dialogowym Eksportuj raport wynikowy, na liście rozwijanej typu pliku eksportu, wybierz pozycję **Portable Document Format**.



Rysunek 65. Okno dialogowe Eksportuj raport wynikowy

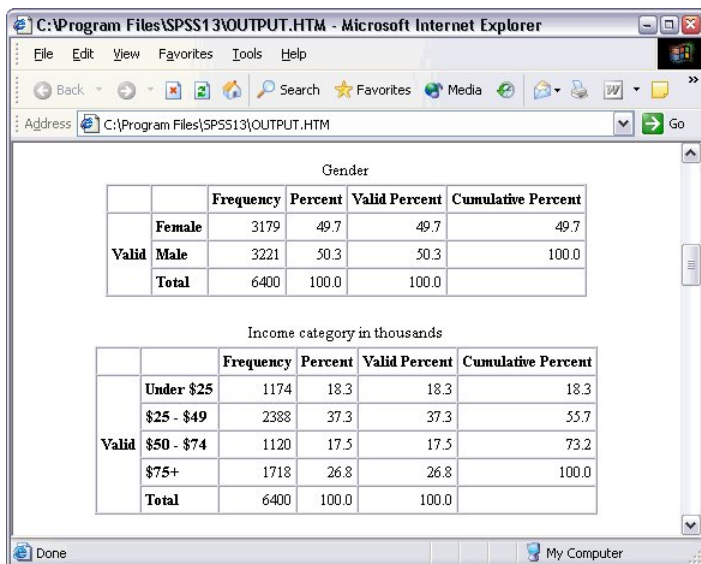
- Panel struktury raportu jest przekształcany w zakładki w pliku PDF, aby ułatwić nawigację.
- Rozmiar strony, jej orientacja, marginesy, zawartość i wyświetlanie nagłówek i stopek oraz rozmiar wykresów w dokumentach PDF są uzależnione od opcji ustawień strony (menu Plik, Ustawienia strony w oknie raportu).
- Rozdzielczość (DPI) dokumentu PDF to bieżące ustawienie rozdzielczości dla domyślnej lub bieżącej drukarki (co można zmienić za pomocą ustawień strony). Maksymalna rozdzielczość wynosi 1200 DPI. Jeśli ustawienie drukarki jest wyższe, rozdzielczość dokumentu PDF będzie równa 1200 DPI. *Uwaga:* Jakość dokumentów w wysokiej rozdzielczości drukowanych na drukarkach o niższej rozdzielczości może być niska.



Rysunek 66. Plik PDF z zakładkami

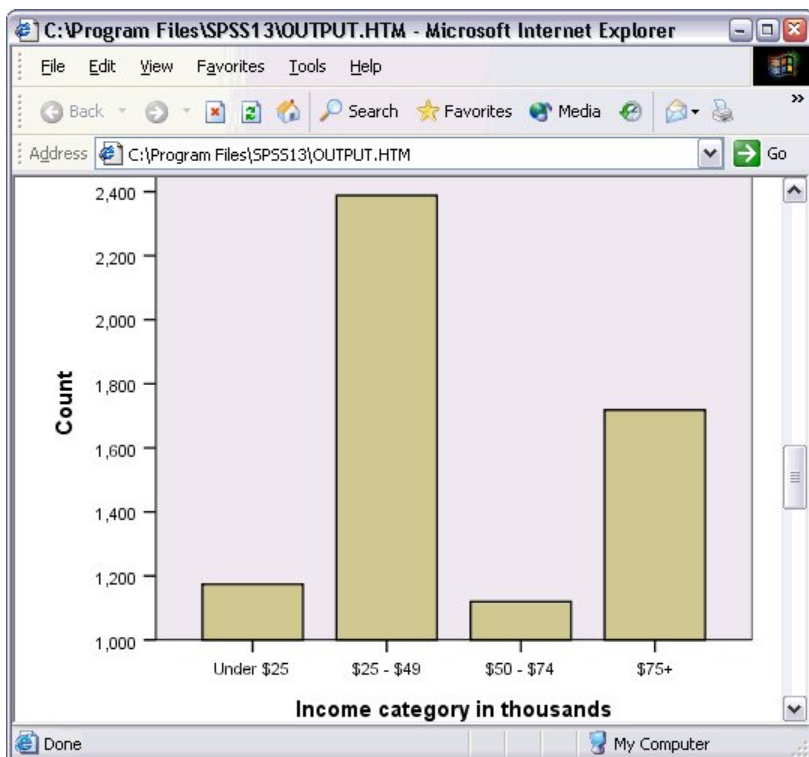
## Eksportowanie wyników do pliku HTML

Wyniki można również eksportować do pliku HTML. Przy zapisywaniu w formacie HTML wszystkie wyniki niegraficzne są eksportowane do jednego pliku HTML.



Rysunek 67. Output.htm w przeglądarce WWW

Przy eksportowaniu do formatu HTML można również wyeksportować wykresy, ale nie do jednego pliku.



Rysunek 68. Wykres w formacie HTML

Każdy wykres zostanie zapisany jako plik w formacie określonym przez użytkownika, a odwołania do tych plików graficznych zostaną umieszczone w kodzie HTML. Dostępna jest także opcja eksportu wszystkich wykresów (lub wybranych wykresów) do oddzielnych plików graficznych.

## Rozdział 7. Praca z komendami

Korzystając z rozbudowanego języka komend, można zapisywać i automatyzować wiele typowych zadań. Zawiera również funkcje, które nie są dostępne w menu ani oknach dialogowych. Większość komend dostępnych jest w menu i oknach dialogowych. Jednakże niektóre komendy i opcje dostępne są jedynie za pośrednictwem języka komend. Język komend umożliwia również zapisanie zadania w pliku komend, dzięki czemu można powtórzyć tę samą analizę w późniejszym terminie.

Plik komend to po prostu plik tekstowy, który zawiera komendy programu IBM SPSS Statistics. Można otworzyć okno komend i wpisywać je bezpośrednio, jednak często łatwiej jest skorzystać z okien dialogowych, które generują część komend automatycznie.

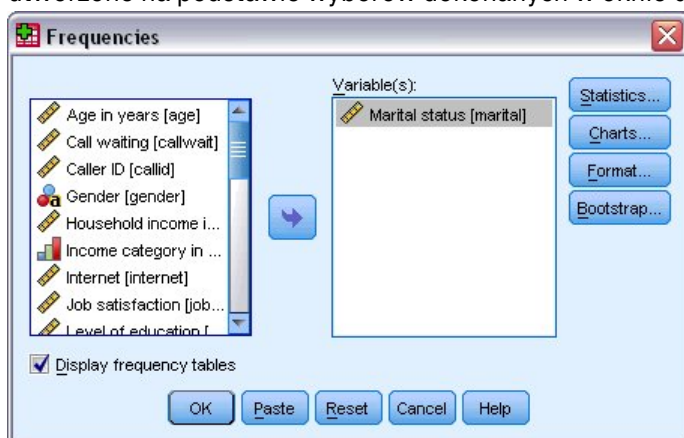
W przykładach podanych w niniejszym rozdziale wykorzystano plik danych *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

*Uwaga:* Komendy nie są dostępne w wersji Student.

### Wklejanie komend

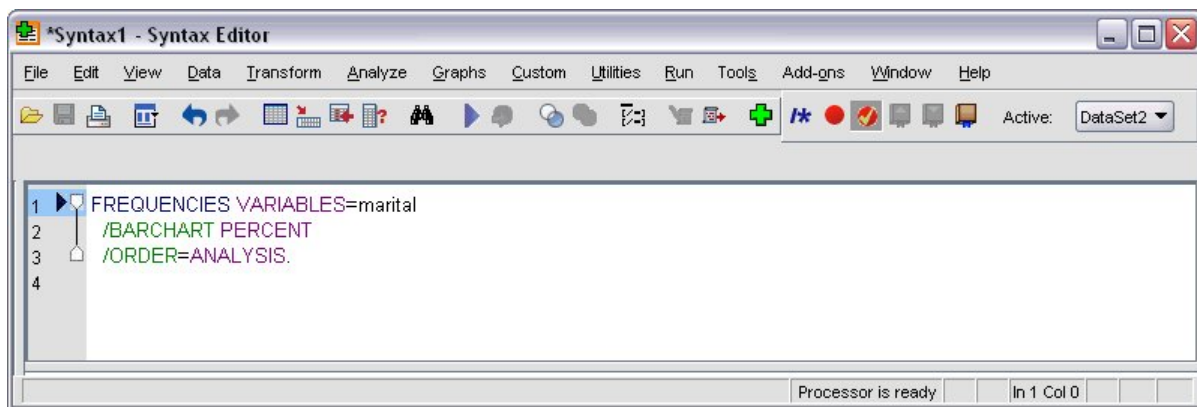
Najprostszym sposobem utworzenia komend jest użycie przycisku Wklej znajdującego się w większości okien dialogowych.

1. Otwórz plik danych *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.
2. Z menu wybierz:  
**Analiza > Statystyki opisowe > Częstości...**
3. Wybierz zmienną *Marital status [marital]* i przenieś ją na listę Zmienne.
4. Kliknij opcję **Wykresy**.
5. W oknie dialogowym Wykresy wybierz opcję **Wykresy słupkowe**.
6. W grupie Wartości na wykresie wybierz opcję **Procenty**.
7. Kliknij przycisk **Kontynuuj**. Kliknij opcję **Wklej**, aby skopiować do Edytora komend komendy utworzone na podstawie wyborów dokonanych w oknie dialogowym.



Rysunek 69. Okno dialogowe Częstości





Rysunek 70. Komenda częstości

8. Aby uruchomić aktualnie wyświetlane komendy, z menu wybierz:

**Uruchom > Wybór**

## Edytowanie komend

W oknie edytora komend można edytować komendy. Na przykład można zmienić opcję komendy / BARCHART, aby wyświetlić częstości zamiast odsetków. (Opcja komendy jest oznaczona ukośnikiem). Jeśli znasz słowo kluczowe służące do wyświetlania częstości, możesz wprowadzić je bezpośrednio. Jeśli nie znasz słowa kluczowego, możesz uzyskać listę dostępnych słów kluczowych dla tej opcji, umieszczając kursor w dowolnym miejscu po nazwie opcji i naciskając klawisze Ctrl+spacja. Spowoduje to wyświetlenie elementu sterującego automatycznego uzupełniania opcji komendy.

Usuń słowo kluczowe PERCENT z opcji komendy BARCHART.

Naciśnij klawisze Ctrl-spacja.

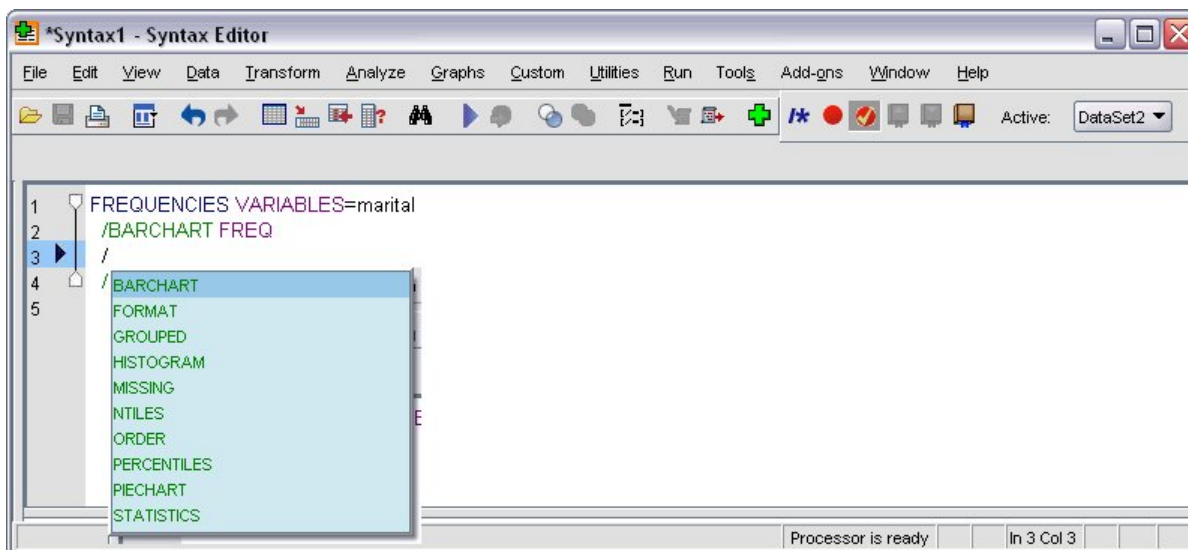
1. Kliknij pozycję z etykietą **FREQ** dla częstotliwości. Kliknięcie wpisu w elemencie sterującym automatycznego uzupełniania spowoduje wstawienie tego wpisu w bieżącej pozycji kursora.

Domyślnie element sterujący automatycznego uzupełniania proponuje listę dostępnych terminów podczas pisania. Załóżmy, że chcemy dołożyć wykres kołowy do wykresu słupkowego. Wykres kołowy definiuje się za pomocą osobnej opcji komendy.

2. Naciśnij klawisz Enter po słowie kluczowym **FREQ** i wpisz ukośnik, aby wskazać początek podkomendy.

W Edytorze komend zostanie wyświetlona lista opcji bieżącej komendy.





Rysunek 71. Element sterujący automatycznego uzupełniania z widocznymi opcjami komendy

Aby uzyskać bardziej szczegółową pomoc do bieżącej komendy, naciśnij klawisz F1. Spowoduje to bezpośrednie przejście do informacji o składni bieżącej komendy.

Jak nietrudno zauważyć, tekst wyświetlany w oknie edytora komend jest kolorowy. Kolorowanie kolorami pozwala na szybkie zauważenie nierozpoznanych terminów, ponieważ tylko rozpoznane terminy są kolorowane. Załóżmy, że błędnie wpisaliśmy opcję FORMAT jako FRMAT. Opcje komend są domyślnie oznaczone kolorem zielonym, ale tekst FRMAT będzie nie będzie wyróżniony żadnym kolorem, ponieważ nie został rozpoznany.

## Otwieranie i uruchamianie pliku komend

1. Aby otworzyć zapisany plik komend, z menu wybierz:

**Plik > Otwórz > Komendy...**

Zostanie wyświetlone standardowe okno dialogowe służące do otwierania plików.

2. Wybierz plik komend. Jeśli nie są wyświetlane żadne pliki komend, upewnij się, że jako typ plików, które mają być wyświetlane, wybrano **Komendy (\*.sps)**.
3. Kliknij przycisk **Otwórz**.
4. Aby uruchomić komendy, należy użyć menu Uruchom w oknie edytora komend.

Jeśli komendy odnoszą się do określonego pliku danych, to przed uruchomieniem komend należy otworzyć plik danych lub dodać komendę, która otworzy plik danych. Można wkleić taką komendę z okna dialogowego, które otwiera pliki danych.

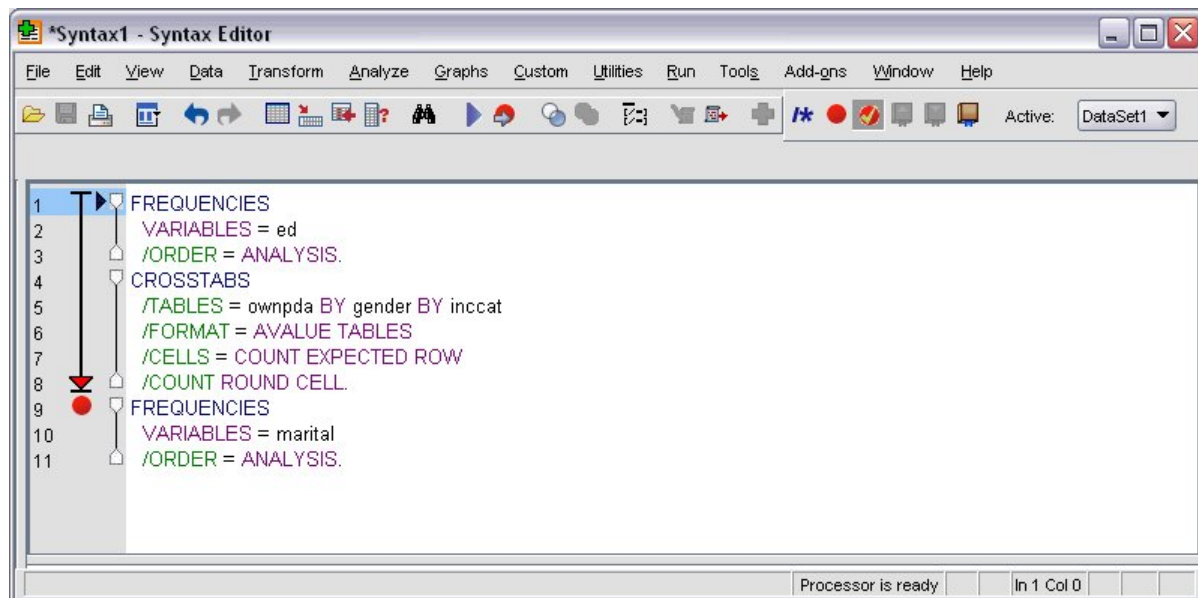
## Korzystanie z punktów zatrzymania

Punkty zatrzymania umożliwiają zatrzymywanie wykonywania komend w określonych punktach i późniejsze wznowianie wykonywania, gdy użytkownik będzie na to gotowy. Pozwalają w ten sposób na wyświetlanie wyników lub danych w punkcie pośrednim, a także w celu wykonywania komend, które wyświetlają informacje o bieżącym stanie danych, takich jak FREQUENCIES. Punkty zatrzymania można ustawiać tylko na poziomie komendy, a nie na konkretnych wierszach w komendzie.

Aby wstawić punkt zatrzymania na komendzie:

1. Kliknij w dowolnym miejscu w obszarze po lewej stronie tekstu powiązanej z komendą.

Punkt zatrzymania jest widoczny jako czerwone kółko w obszarze po lewej stronie tekstu komendy, w tym samym wierszu, co nazwa komendy, niezależnie od miejsca, w którym kliknięto.



Rysunek 72. Wykonywanie zatrzymane w punkcie zatrzymania

Wykonywanie komend zawierających punkty zatrzymania zatrzymuje się przed każdą komendą zawierającą taki punkt.

Strzałka skierowana w dół po lewej stronie tekstu komendy odzwierciedla postęp wykonania komend. Obejmuje odcinek od pierwszej do ostatniej wykonanej komendy i jest szczególnie przydatna przy wykonywaniu komend zawierających punkty zatrzymania.

Aby wznowić zatrzymanie po punkcie zatrzymania:

2. Z menu w Edytorze komend wybierz kolejno opcje:

**Uruchom > Kontynuuj**

---

## Rozdział 8. Modyfikowanie wartości danych

Początkowo dane nie zawsze są zorganizowane w sposób najbardziej odpowiedni do analizy lub raportowania. Można więc wykonać na nich m.in. następujące czynności:

- Utworzyć zmienną kategorialną ze zmiennej ilościowej.
- Połączyć kilka kategorii odpowiedzi w pojedynczą kategorię.
- Utworzyć nową zmienną, która stanowi obliczoną różnicę między dwiema istniejącymi zmiennymi.
- Obliczyć długość okresu między dwiema datami.

W tym rozdziale używany jest plik danych *demo.sav*. Więcej informacji zawiera temat [Rozdział 10, “Pliki przykładowe”](#), na stronie 77.

---

### Tworzenie zmiennej kategorialnej na podstawie zmiennej ilościowej

Kilka zmiennych kategorialnych w pliku danych *demo.sav* to w rzeczywistości zmienne wywiedzione ze zmiennych ilościowych zawartych w tym samym pliku danych. Na przykład zmienna *inccat* jest po prostu zmienną *income* pogrupowaną w cztery kategorie. Ta zmienna kategorialna przyjmuje wartości całkowite od 1 do 4 reprezentujące następujące kategorie dochodu (w tysiącach): poniżej 25, 25-49, 50-74 i 75 lub więcej.

Aby utworzyć zmienną kategorialną *inccat*:

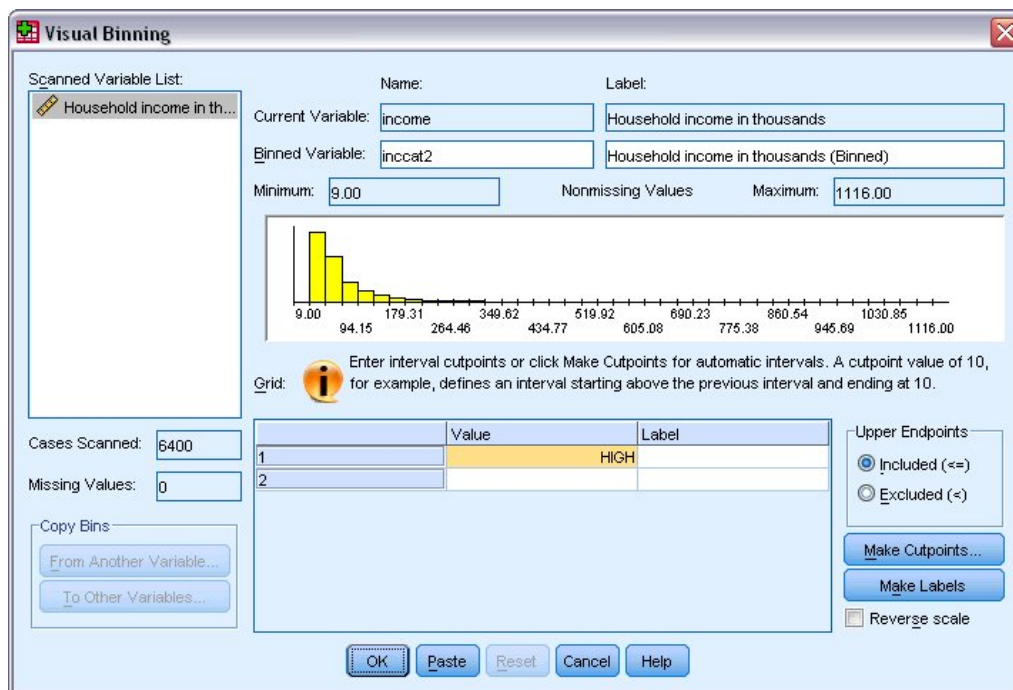
1. Z menu okna Edytora danych wybierz kolejno następujące pozycje:

#### **Przekształcenia > Kategoryzacja wizualna...**

W początkowym oknie dialogowym Kategoryzacja wizualna wybierasz zmienne ilościowe i/lub porządkowe, z których chcesz utworzyć nowe zmienne kategoryzowane. **Kategoryzacja** oznacza grupowanie w jedną kategorię co najmniej dwóch sąsiadujących wartości.

Ponieważ funkcja kategoryzacji wizualnej opiera się na rzeczywistych wartościach z pliku danych, musi najpierw odczytać plik danych. Ponieważ może to zająć dłuższy czas, jeśli plik danych zawiera dużą liczbę obserwacji, to początkowe okno dialogowe umożliwia również ograniczenie liczby odczytywanych obserwacji. Nie jest to konieczne w przypadku naszego przykładowego pliku danych. Mimo że zawiera on ponad 6 000 obserwacji, odczytanie jego obserwacji nie trwa zbyt długo.

2. Przeciągnij i upuść zmienną *Household income in thousands [income]* z listy zmiennych do kategoryzacji, a następnie kliknij przycisk **Kontynuuj**.



Rysunek 73. Główne okno dialogowe Kategoryzacja wizualna

3. W głównym oknie dialogowym Kategoryzacja wizualna, na liście Zeskanowane zmienne, wybierz pozycję *Household income in thousands* [*income*].

Pojawi się histogram rozkładu wybranej zmiennej (w tym przypadku jest bardzo skośny).

4. Wpisz *inccat2* jako nową nazwę zmiennej kategoryzowanej i *Kategoria dochodu [w tysiącach]* jako etykietę tej zmiennej.
5. Kliknij przycisk **Punkty podziału**.
6. Wybierz opcję **Równe szerokości przedziałów**.
7. Wprowadź wartość 25 jako położenie pierwszego punktu podziału, 3 jako liczbę punktów podziału i 25 jako szerokość.

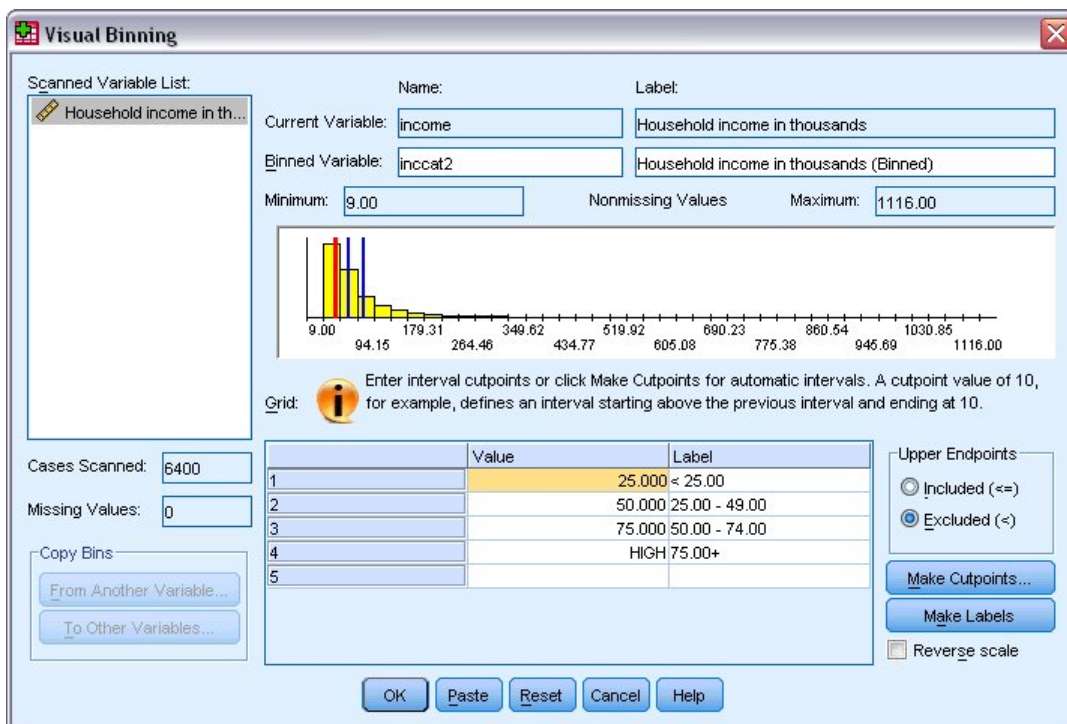
Liczba kategorii jest o jeden większa niż liczba punktów podziału. W tym przykładzie nowa zmienna kategoryzowana będzie zawierać cztery kategorie, przy czym trzy pierwsze będą obejmować po 25 (tysięcy), a ostatnia – wszystkie wartości powyżej najwyższego punktu podziału, czyli 75 (tysięcy).

8. Kliknij przycisk **Zastosuj**.

Wartości wyświetlane teraz w siatce reprezentują zdefiniowane punkty podziału, które są górnymi punktami końcowymi każdej kategorii. Pionowe linie na histogramie również wskazują punkty podziału.

Domyślnie wartości punktów podziału są uwzględniane w odpowiednich kategoriach. Na przykład pierwsza kategoria 25 będzie zawierać wszystkie wartości mniejsze lub równe 25. Ale w tym przykładzie chcemy, aby kategorie odpowiadały wartościom mniejszym niż 25, następnie 25-49, 50-74 oraz 75 i wyższym.

9. W grupie Górne granice przedziałów wybierz opcję **Wykluczone (<)**.
10. Następnie kliknij przycisk **Utwórz etykiety**.



Rysunek 74. Automatycznie wygenerowane etykiety wartości

Spowoduje to automatyczne wygenerowanie opisowych etykiet wartości dla każdej kategorii. Ponieważ rzeczywiste wartości przypisane do nowej zmiennej kategoryzowanej są po prostu kolejnymi liczbami całkowitymi, począwszy od 1, etykiety wartości mogą być bardzo użyteczne.

Można także ręcznie wprowadzać lub zmieniać punkty podziału i etykiety w siatce, zmieniać położenia punktów podziału poprzez przeciąganie i upuszczanie ich linii na histogramie, a także usuwać punkty podziału poprzez przeciąganie ich linii poza histogram.

11. Kliknij przycisk **OK**, aby utworzyć nową zmienną kategoryzowaną.

Nowa zmienna zostanie wyświetlona w Edytorze danych. Ponieważ zmienna jest dodawana na końcu pliku, jest ona wyświetlana w prawej kolumnie w widoku danych i w ostatnim wierszu w widoku zmiennych.

## Obliczanie zmiennych wynikowych

Korzystając z szerokiej gamy funkcji matematycznych, można obliczać nowe zmienne w oparciu o złożone równania. W tym przykładzie jednak po prostu obliczymy nową zmienną, która jest różnicą między wartościami dwóch istniejących zmiennych.

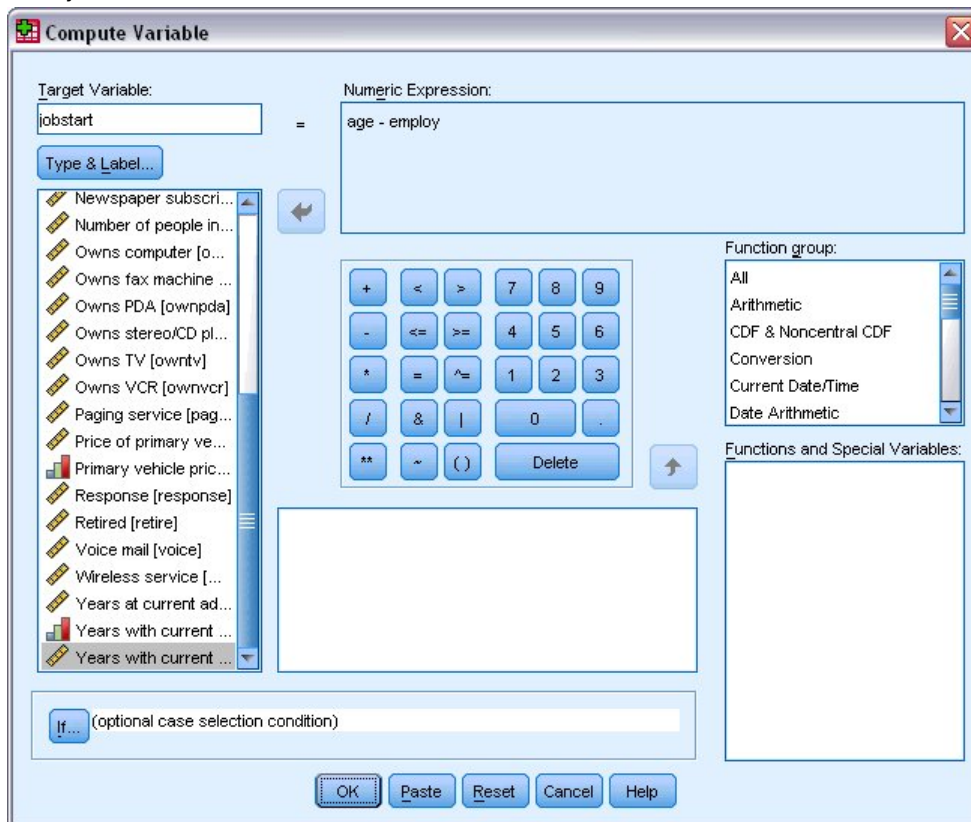
Plik danych *demo.sav* zawiera zmienną bieżącego wieku respondenta i zmienną liczby lat w obecnej pracy. Nie zawiera jednak zmiennej wieku respondenta w momencie, gdy podejmował tę pracę. Możemy utworzyć nową zmienną, która stanowi obliczoną różnicę między bieżącym wiekiem a liczbą lat w obecnej pracy, co powinno dać nam przybliżony wiek, w którym respondent podjął obecną pracę.

1. Z menu okna Edytora danych wybierz kolejno następujące pozycje:

**Przekształcenie > Oblicz wartości zmiennej...**

2. W polu Zmienna przewidywana wpisz `jobstart`.
3. Wybierz zmienną *Age in years [age]* na liście zmiennych źródłowych, a następnie kliknij przycisk ze strzałką, aby skopiować ją do pola tekstowego Wyrażenie liczbowe.
4. Kliknij przycisk minus (-) w polu kalkulatora w oknie dialogowym (lub naciśnij klawisz minus na klawiaturze).

5. Wybierz zmienną *Years with current employer [employ]* i kliknij przycisk ze strzałką, aby skopiować ją do wyrażenia.



Rysunek 75. Okno dialogowe Oblicz wartość zmiennej

*Uwaga:* Należy zwrócić uwagę, by wybrać prawidłową zmienną zatrudnienia. Istnieje również przekodowana jakościowa wersja zmiennej, która *nie* jest tą, która nas obecnie interesuje. Wyrażenie liczbowe powinno mieć postać *age-employ*, a nie *age-empcat*.

6. Kliknij przycisk **OK**, aby obliczyć nową zmienną.

Nowa zmienna zostanie wyświetlona w Edytorze danych. Ponieważ zmienna jest dodawana na końcu pliku, jest ona wyświetlana w prawej kolumnie w widoku danych i w ostatnim wierszu w widoku zmiennych.

## Używanie funkcji w wyrażeniach

W wyrażeniach można również używać predefiniowanych funkcji. Dostępnych jest ponad 70 wbudowanych funkcji, w tym:

- Funkcje arytmetyczne
- Funkcje statystyczne
- Funkcje rozkładu
- Funkcje logiczne
- Funkcje agregacji daty i czasu oraz funkcje wyodrębniania
- Funkcje dotyczące braków danych
- Funkcje wieloobserwacyjne
- Funkcje łańcuchowe

Funkcje są zebrane w logiczne grupy, np. operacji arytmetycznych i obliczania miar statystycznych. Dla wygody, szereg często używanych zmiennych systemowych, takich jak *\$TIME* (bieżąca data i godzina), również uwzględniono w odpowiednich grupach funkcji.

Wklejanie funkcji do wyrażenia

Aby wkleić funkcję do wyrażenia:

1. Umieść kursor w wyrażeniu w punkcie, w którym chcesz wkleić funkcję.
2. Wybierz odpowiednią grupę z listy Grupa funkcji. Grupa **Wszystkie** zawiera listę wszystkich dostępnych funkcji i zmiennych systemowych.
3. Kliknij dwukrotnie funkcję na liście Funkcje i zmienne specjalne (lub wybierz funkcję i kliknij przycisk ze strzałką obok listy Grupa funkcji).

Funkcja zostanie wstawiona do wyrażenia. Jeśli zaznaczysz część wyrażenia, a następnie wstawisz funkcję, zaznaczona część wyrażenia będzie używana jako pierwszy argument funkcji.

Edytowanie funkcji w wyrażeniu

Funkcja nie jest kompletna, dopóki nie zostaną wprowadzone argumenty, reprezentowane przez znaki zapytania we wklejonej funkcji. Liczba znaków zapytania wskazuje na minimalną liczbę argumentów wymaganych przez funkcję.

1. Zaznacz znak(i) zapytania we wklejonej funkcji.
2. Wprowadź argumenty. Jeśli argumenty są nazwami zmiennych, można wkleić je z listy zmiennych.

## Korzystanie z wyrażeń warunkowych

Aby stosować przekształcenia do wybranych podzbiorów obserwacji, można używać wyrażeń warunkowych (zwanymi również wyrażeniami logicznymi). Wyrażenie warunkowe dla każdej obserwacji zwraca wartości true (prawda), false (fałsz) albo missing (brak danych). Jeśli wynikiem wyrażenia warunkowego jest prawda, przekształcenie zostanie zastosowane do danej obserwacji. Jeśli wynik ma wartość fałsz lub brak danych, przekształcenie nie zostanie zastosowane do obserwacji.

Aby określić wyrażenie warunkowe:

1. Kliknij przycisk **Jeżeli** w oknie dialogowym Oblicz wartość zmiennej. Spowoduje to otwarcie okna dialogowego Wybierz obserwacje: Warunkowo.
2. Wybierz opcję **Uwzględnij, jeśli obserwacja spełnia warunek**.
3. Wprowadź wyrażenie warunkowe.

Większość wyrażeń warunkowych zawiera co najmniej jeden operator relacyjny, np.:

`age>=21`

lub

`income*3<100`

W pierwszym przykładzie wybierane są tylko obserwacje, w których zmienna wieku *Age [age]* ma wartość 21 lub większą. W drugim przykładzie wartość *Household income in thousands [income]* pomnożona przez 3 musi być mniejsza niż 100, aby obserwacja została wybrana.

Korzystając z operatorów logicznych, można również połączyć dwa lub więcej wyrażeń warunkowych, np.:

`age>=21 | ed>=4`

lub

`income*3<100 & ed=5`

W pierwszym przykładzie wybierane są obserwacje, które spełniają warunek *Age [age]* lub warunek *Level of education [ed]*. W drugim przykładzie muszą być spełnione oba warunki, *Household income in thousands [income]* i *Level of education [ed]*, aby warunek został wybrany.



## Praca z datami i godzinami

Szereg zadań często wykonywanych w odniesieniu do dat i godzin można łatwo realizować za pomocą Kreatora obliczeń daty i czasu. Za pomocą tego kreatora można:

- Utworzyć zmienną daty/czasu ze zmiennej łańcuchowej zawierającej datę lub godzinę.
- Skonstruować zmienną daty/czasu, łącząc zmienne zawierające różne części daty lub godziny.
- Dodawać lub odejmować wartości ze zmiennych daty/czasu, w tym dodawać lub odejmować dwie zmienne daty/czasu.
- Wyodrębnić część zmiennej daty lub czasu, na przykład dzień miesiąca ze zmiennej daty/czasu, która ma postać mm/dd/yyyy.

W przykładach w tej sekcji używamy pliku danych *upgrade.sav*. Więcej informacji zawiera temat [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

Aby skorzystać z Kreatora obliczeń daty i czasu:

1. Z menu wybierz:

**Przekształcenia > Definiuj datę i czas...**



Rysunek 76. Ekran wprowadzający Kreatora obliczeń daty i czasu

Na ekranie wprowadzającym Kreatora obliczeń daty i czasu widoczny jest zestaw ogólnych zadań. Zadania, które nie mają zastosowania do bieżących danych, są wyłączone. Na przykład plik danych *upgrade.sav* nie zawiera żadnych zmiennych łańcuchowych, a więc zadanie tworzenia zmiennej daty z łańcucha jest wyłączone.

Jeśli nie masz jeszcze doświadczenia w pracy z datami i godzinami w programie IBM SPSS Statistics, możesz wybrać opcję **Dowiedz się, jak są są przedstawiane data i czas** i kliknij przycisk **Dalej**. Nastąpi przejście do ekranu, który zawiera krótkie omówienie zmiennych daty/czasu oraz odsyłacz (przycisk Pomoc) do bardziej szczegółowych informacji.

### Obliczanie długości okresu między dwiema datami

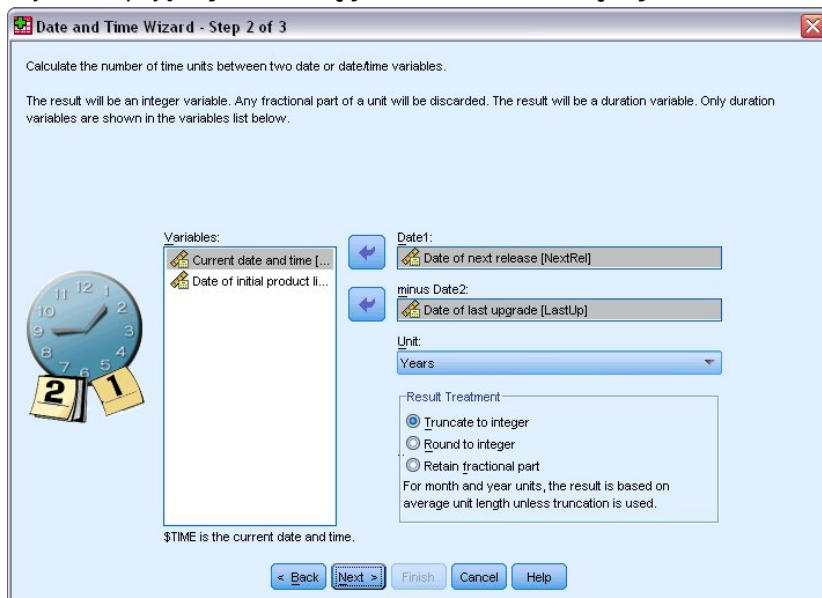
Jednym z najczęstszych zadań dotyczących dat jest obliczanie długości okresu między dwiema datami. Rozważmy na przykład firmę, która jest zainteresowana analizowaniem zakupów licencji na aktualizacje poprzez określenie liczby lat od momentu zakupu aktualizacji przez każdego z klientów. Plik danych



*upgrade.sav* zawiera zmienną daty ostatniego zakupu aktualizacji przez każdego z klientów, ale nie zawiera liczby lat od tego zakupu. Odpowiednią miarą będzie nowa zmienna, która jest długością okresu w latach między datą ostatniej aktualizacji a datą wydania następnej wersji produktu.

Aby obliczyć długość okresu między dwiema datami:

1. Na ekranie wprowadzającym Kreatora obliczeń daty i czasu wybierz opcję **Wykonać obliczenia w oparciu o istniejące wartości w formacie daty i czasu**, a następnie kliknij przycisk **Dalej**.
2. Wybierz opcję **Wylicz liczbę jednostek czasu między dwiema datami** i kliknij przycisk **Dalej**.



Rysunek 77. Obliczanie długości okresu między dwiema datami: krok 2

3. W kroku 2 wybierz zmienną *Date of next release* jako datę 1.
4. Wybierz zmienną *Date of last upgrade* jako datę 2.
5. Wybierz **Lata** w polu Jednostka i **Skróć do liczby całkowitej** jako metodę Obsługa wyników. (Są to domyślne wybory).
6. Kliknij przycisk **Dalej**.
7. W kroku 3 wpisz *LataOdOstAkt* jako nazwę zmiennej wynikowej. Zmienna wynikowa nie może mieć takiej samej nazwy, jak istniejąca zmienna.
8. Wpisz *Lata od ostatniej aktualizacji* jako etykiętę zmiennej wynikowej. Etykiety zmiennych wynikowych są opcjonalne.
9. Pozostaw domyślny wybór opcji **Utwórz zmienną**, a następnie kliknij przycisk **Zakończ**, aby utworzyć nową zmienną.

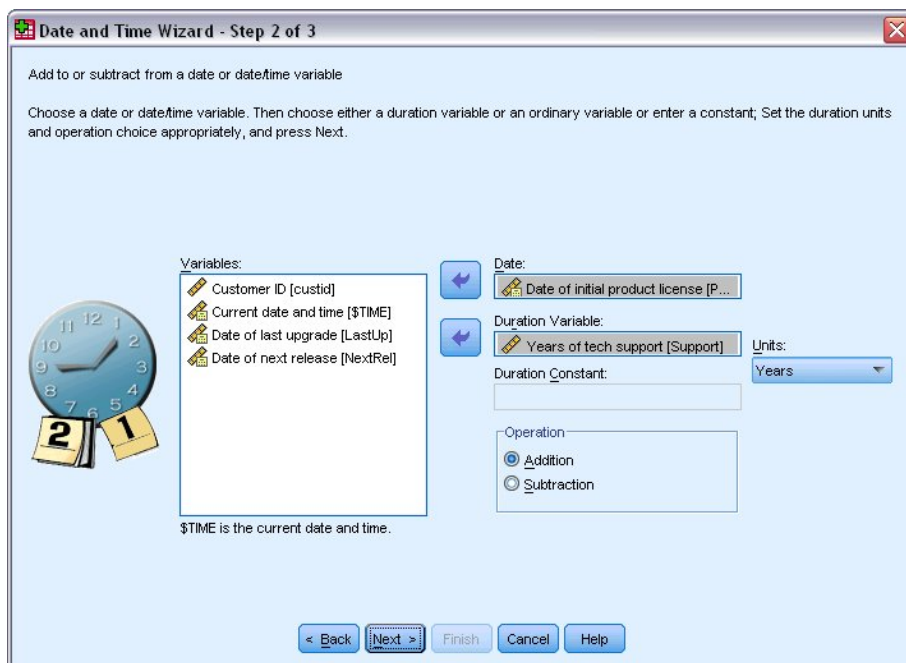
Nowa zmienna, *LataOdOstAkt*, wyświetlana w Edytorze danych, jest całkowitą liczbą lat między dwiema datami. Ułamkowa część roku została obcięta.

## Dodawanie czasu trwania do daty

Istnieje możliwość dodania lub odejmowania czasów trwania, na przykład 10 dni lub 12 miesięcy, do/od daty. Kontynuując pracę z przykładem producenta oprogramowania, rozważmy problem wyznaczania daty końca pierwszej umowy o wsparcie techniczne każdego z klientów. Plik danych *upgrade.sav* zawiera zmienną liczby lat zakontraktowanego wsparcia i zmienną daty zakupu. Można określić datę końcową okresu wsparcia, dodając liczbę lat wsparcia do daty zakupu.

Aby dodać czas trwania do daty:

1. Na ekranie wprowadzającym Kreatora obliczeń daty i czasu wybierz opcję **Wykonać obliczenia w oparciu o istniejące wartości w formacie daty i czasu**, a następnie kliknij przycisk **Dalej**.
2. W kroku 1 wybierz **Dodaj lub odejmij czas trwania od wartości daty** i kliknij przycisk **Dalej**.



Rysunek 78. Dodawanie czasu trwania do daty: krok 2

3. Wybierz pozycję *Date of initial product license* jako datę.

4. W kroku 2 wybierz *Years of tech support* jako zmienną czasu trwania.

Ponieważ *Years of tech support* to po prostu zmienna liczbowa, należy wskazać jednostki, które mają być używane podczas dodawania tej zmiennej jako czasu trwania.

5. Z listy rozwijanej Jednostki wybierz pozycję **Lata**.

6. Kliknij przycisk **Dalej**.

7. W kroku 3 wprowadź wartość *DataKońWsp* jako nazwę zmiennej wynikowej. Zmienna wynikowa nie może mieć takiej samej nazwy, jak istniejąca zmienna.

8. Jako etykietę zmiennej wynikowej wpisz *Data końca wsparcia*. Etykiety zmiennych wynikowych są opcjonalne.

9. Kliknij przycisk **Zakończ**, aby utworzyć nową zmienną.

Nowa zmienna zostanie wyświetlona w Edytorze danych.

---

## Rozdział 9. Sortowanie i wybieranie danych

Struktura plików danych nie zawsze idealnie odpowiada określonym potrzebom użytkownika. Aby przygotować dane do analizy, można skorzystać z szerokiej gamy przekształceń plików, takich jak:

- **Sortowanie danych.** Można sortować obserwacje według wartości jednej lub kilku zmiennych.
- **Wybór podzbiorów obserwacji.** Można ograniczyć analizę do podzbioru obserwacji lub przeprowadzić jednocześnie analizy na różnych podzbiórach.

W przykładach podanych w niniejszym rozdziale wykorzystano plik danych *demo.sav*. Więcej informacji można znaleźć w temacie [Rozdział 10, "Pliki przykładowe"](#), na stronie 77.

---

### Sortowanie danych

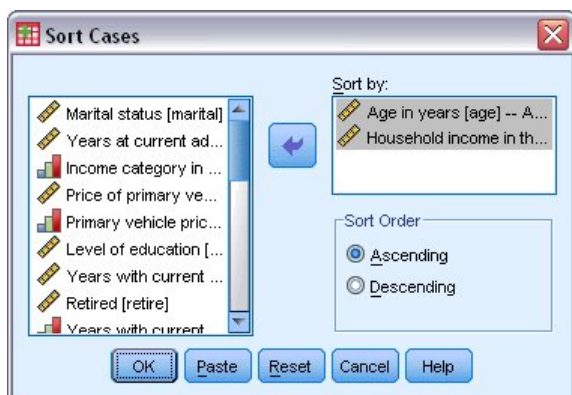
Sortowanie obserwacji (sortowanie wierszy pliku danych) jest często przydatne i czasami niezbędne w niektórych typach analiz.

Aby zmienić kolejność obserwacji w pliku danych na podstawie wartości jednej lub większej liczby zmiennych sortujących:

1. Z menu wybierz:

**Dane > Sortuj obserwacje...**

Zostanie wyświetlone okno dialogowe Sortuj obserwacje.



Rysunek 79. Okno dialogowe Sortuj obserwacje

2. Dodaj zmienne *Age in years [age]* i *Household income in thousands [income]* do listy Sortuj wg.

W przypadku wybrania więcej niż jednej zmiennej sortującej kolejność, w jakiej są one wymienione na liście Sortuj wg, wpływa na kolejność sortowania obserwacji. W tym przykładzie, na podstawie pozycji na liście Sortuj wg, obserwacje będą sortowane według wartości *Household income in thousands [income]* w obrębie kategorii *Age in years [age]*. W przypadku zmiennych łańcuchowych wielkie litery poprzedzają małe odpowiedniki w porządku sortowania (na przykład wartość łańcuchowa *Tak* występuje przed *tak*).

---

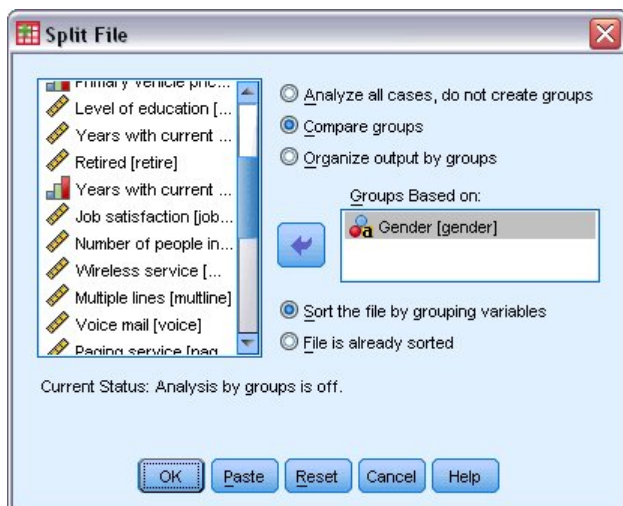
### Przetwarzanie pliku dzielonego

Aby podzielić plik danych na osobne grupy do analizy:

1. Z menu wybierz:

**Dane > Podziel na podzbiory...**

Zostanie wyświetlone okno dialogowe Podział na podzbiory.



Rysunek 80. Okno dialogowe Podział na podzbiory

- Wybierz opcję **Porównaj grupy** lub **Przedstaw wyniki w podziale na grupy**. (W następnych przykładach objaśnimy różnice między tymi dwoma opcjami).
- Wybierz zmienną *Gender [gender]*, aby podzielić plik na osobne grupy według tej zmiennej.

Jako zmiennych grupujących można użyć zmiennych liczbowych, krótkich łańcuchów i długich łańcuchów. Dla każdej podgrupy, która jest zdefiniowana przez zmienne grupujące, wykonywana jest odrębna analiza. W przypadku wybrania więcej niż jednej zmiennej grupującej kolejność, w jakiej są one wymienione na liście Grupy na podstawie, wpływa na sposób grupowania obserwacji.

Jeśli zostanie wybrana opcja **Porównaj grupy**, wyniki ze wszystkich grup podzielonego pliku zostaną uwzględnione w tej samej tabeli, co ilustruje poniższa tabela statystyk podsumowujących wygenerowana przez procedurę Częstości.

| Statistics                    |                |         |          |
|-------------------------------|----------------|---------|----------|
| Household income in thousands |                |         |          |
| Female                        | N              | Valid   | 3179     |
|                               |                | Missing | 0        |
|                               | Mean           |         | 68.7798  |
|                               | Median         |         | 44.0000  |
|                               | Std. Deviation |         | 75.73510 |
| Male                          | N              | Valid   | 3221     |
|                               |                | Missing | 0        |
|                               | Mean           |         | 70.1608  |
|                               | Median         |         | 45.0000  |
|                               | Std. Deviation |         | 81.56216 |

Rysunek 81. Wynik analizy pliku dzielonego z jedną tabelą przestawną

Jeśli wybierzesz opcję **Przedstaw wyniki w podziale na grupy** i uruchomisz procedurę Częstości, zostaną utworzone dwie tabele przestawne: jedna dla kobiet i jedna dla mężczyzn.

| Statistics <sup>a</sup>       |         |          |
|-------------------------------|---------|----------|
| Household income in thousands |         |          |
| N                             | Valid   | 3179     |
|                               | Missing | 0        |
| Mean                          |         | 68.7798  |
| Median                        |         | 44.0000  |
| Std. Deviation                |         | 75.73510 |

a. Gender = Female

Rysunek 82. Wyniki analizy pliku dzielonego z tabelą przestawną dla kobiet

**Statistics<sup>a</sup>**

Household income in thousands

|                |         |          |
|----------------|---------|----------|
| N              | Valid   | 3221     |
|                | Missing | 0        |
| Mean           |         | 70.1608  |
| Median         |         | 45.0000  |
| Std. Deviation |         | 81.56216 |

a. Gender = Male

Rysunek 83. Wyniki analizy pliku dzielonego z tabelą przestawną dla mężczyzn

## Sortowanie obserwacji przy przetwarzaniu pliku dzielonego

Procedura dzielenia na podzbiory tworzy nową podgrupę za każdym razem, gdy napotka inną wartość dla jednej z zmiennych grupujących. Dlatego ważne jest, aby przed rozpoczęciem przetwarzania pliku dzielonego posortować obserwacje według wartości zmiennych grupujących.

Domyślnie podział na podzbiory automatycznie sortuje plik danych w oparciu o wartości zmiennych grupujących. Jeśli plik został już posortowany w odpowiedniej kolejności, można skrócić czas przetwarzania, wybierając opcję **Plik jest już posortowany**.

## Włączanie i wyłączanie przetwarzania pliku dzielonego

Po włączeniu przetwarzania pliku dzielonego funkcja jest aktywna do końca sesji, chyba że zostanie wyłączona.

- **Analizuj wszystkie obserwacje.** Ta opcja powoduje wyłączenie przetwarzania pliku dzielonego.
- **Porównaj grupy i Przedstaw wyniki w podziale na grupy.** Ta opcja powoduje włączenie przetwarzania pliku dzielonego.

Jeśli przetwarzanie pliku dzielonego jest aktywne, na pasku statusu w dolnej części okna aplikacji widoczny komunikat **Dzielenie pliku włączone**.

## Wybieranie podzbiorów obserwacji

Analizę można ograniczyć do konkretnej podgrupy w oparciu o kryteria obejmujące zmienne i wyrażenia złożone. Można także wybrać próbę losową obserwacji. Kryteria używane do definiowania podgrupy mogą obejmować:

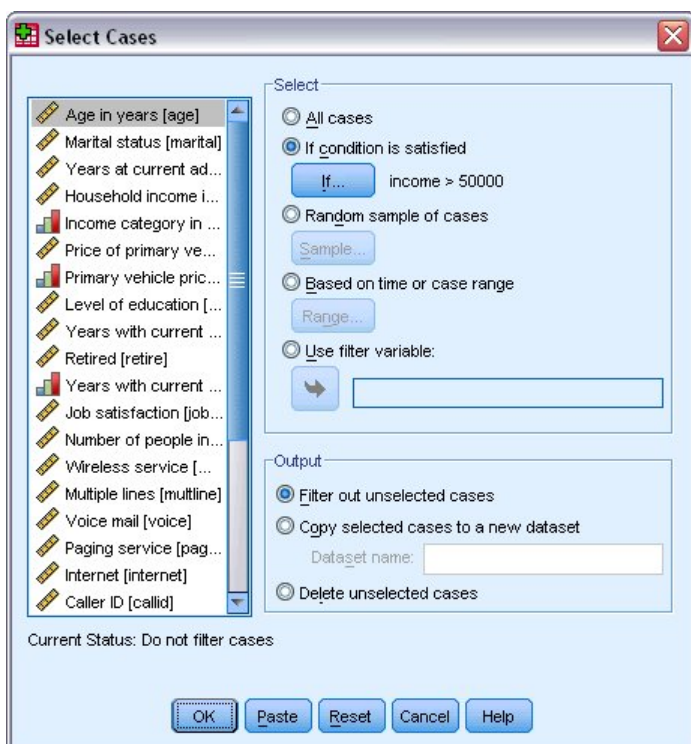
- wartości i zakresy zmiennych,
- zakresy dat i czasów,
- numery obserwacji (wierszy),
- wyrażenia arytmetyczne,
- wyrażenia logiczne,
- funkcje.

Aby wybrać podzbiór obserwacji do analizy:

1. Z menu wybierz:

**Dane > Wybierz obserwacje...**

Spowoduje to otwarcie okna dialogowego Wybierz obserwacje.



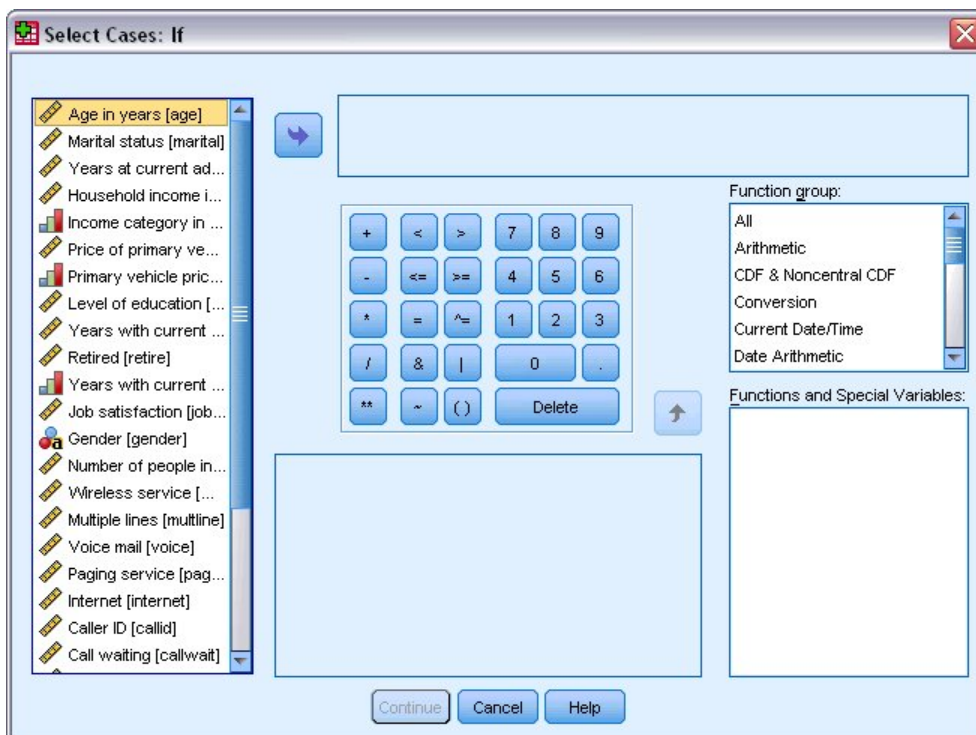
Rysunek 84. Okno dialogowe Wybierz obserwacje

## Wybieranie obserwacji na podstawie wyrażenia warunkowego

Aby wybrać obserwacje na podstawie wyrażenia warunkowego:

1. Wybierz opcję **Jeśli spełniony jest warunek** i kliknij opcję **Jeśli** w oknie dialogowym Wybieranie obserwacji.

Spowoduje to otwarcie okna dialogowego Wybierz obserwacje: Warunkowo.



Rysunek 85. Okno dialogowe Wybierz obserwacje: Warunkowo



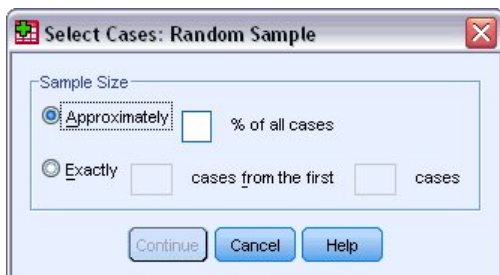
W wyrażeniu warunkowym można korzystać z istniejących nazw zmiennych, stałych, operatorów arytmetycznych, operatorów logicznych, operatorów relacyjnych i funkcji. Można wpisać i edytować wyrażenie w polu tekstowym, tak jak tekst w oknie wyników. Aby wklejać elementy do wyrażenia, można także używać pola kalkulatora, listy zmiennych i listy funkcji. Więcej informacji można znaleźć w temacie “Korzystanie z wyrażen warunkowych” na stronie 67.

## Wybieranie próby losowej

Aby uzyskać próbę losową:

1. W oknie dialogowym Wybierz obserwacje wybierz opcję **Próba losowa obserwacji**.
2. Kliknij przycisk **Próba**.

Spowoduje to otwarcie okna dialogowego Wybierz obserwacje: Próba losowa.



Rysunek 86. Okno dialogowe Wybierz obserwacje: Próba losowa.

Można wybrać jedną z następujących opcji wielkości próby:

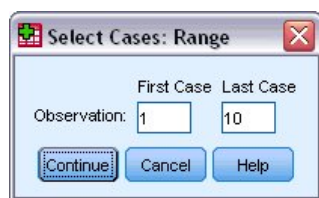
- **W przybliżeniu.** Odsetek określony przez użytkownika. Ta opcja generuje próbę losową zawierającą w przybliżeniu określony odsetek obserwacji.
- **Dokładnie.** Liczba obserwacji określona przez użytkownika. Należy również określić liczbę obserwacji, z których generowana jest próba. Ta druga liczba powinna być mniejsza lub równa całkowitej liczbie obserwacji w pliku danych. Jeśli ta liczba przekracza całkowitą liczbę obserwacji w pliku danych, próba będzie zawierać proporcjonalnie mniej obserwacji niż żądano.

## Wybieranie zakresu czasu lub zakresu obserwacji

Aby wybrać zakres obserwacji w oparciu o daty, godziny lub numery obserwacji (wierszy):

1. Wybierz opcję **Na podstawie czasu lub zakresu obserwacji**, a następnie kliknij opcję **Zakres** w oknie dialogowym Wybierz obserwacje.

Spowoduje to otwarcie okna dialogowego Wybierz obserwacje: Zakres, w którym można wybrać zakres numerów obserwacji (wierszy).



Rysunek 87. Okno dialogowe Wybierz obserwacje: Zakres

- **Pierwsza.** Wprowadź początkową datę i/lub godzinę zakresu. Jeśli nie są zdefiniowane żadne zmienne dat, należy wprowadzić końcowy numer obserwacji (numer wiersza w Edytorze danych, o ile nie jest włączony Podział na podzbiory). Jeśli ostatnia obserwacja nie zostanie określona, zostaną wybrane wszystkie obserwacje od początkowej daty/godziny do końca szeregu czasowego.
- **Ostatnia.** Wprowadź końcową datę i/lub godzinę zakresu. Jeśli nie są zdefiniowane żadne zmienne dat, należy wprowadzić końcowy numer obserwacji (numer wiersza w Edytorze danych, o ile nie jest

włączony Podział na podzbiory). Jeśli pierwsza obserwacja nie zostanie określona, zostaną wybrane wszystkie obserwacje od początku szeregu czasowego do końcowej daty/godziny.

W przypadku danych szeregów czasowych ze zdefiniowanymi zmiennymi daty można wybrać zakres dat i/lub godzin w oparciu o zdefiniowane zmienne daty. Każda obserwacja reprezentuje obserwacje w innym momencie, a plik jest posortowany w kolejności chronologicznej.



Rysunek 88. Okno dialogowe Wybierz obserwacje: Zakres (szeregi czasowe)

Aby wygenerować zmienne daty dla danych szeregu czasowego:

2. Z menu wybierz:

**Dane > Definiuj datę i czas...**

## Postępowanie z niewybranymi obserwacjami

Jako sposób postępowania z niewybranymi obserwacjami można wybrać jedną z poniższych alternatyw:

- **Odfiltruj niewybrane obserwacje.** Niewybrane obserwacje nie są uwzględniane w analizie, lecz pozostają w zbiorze danych. Jeśli odfiltrowanie zostanie wyłączone, istnieje możliwość późniejszego użycia niewybranych obserwacji w sesji. Jeśli wybrana zostanie próba losowa lub jeśli wybrane zostaną obserwacje oparte o wyrażenie warunkowe, generowana jest zmienna o nazwie *filter\_\$* z wartością 1 dla wybranych obserwacji a dla niewybranych obserwacji wybrane jest 0.
- **Skopiuj wybrane obserwacje do nowego zbioru danych.** Wybrane obserwacje są kopiowane do nowego zbioru danych, bez wprowadzania zmian w oryginalnym zbiorze danych. Niewybrane obserwacje nie są zawarte w nowym zbiorze danych i są pozostawione w niezmienionej postaci w oryginalnym zbiorze danych.
- **Usuń niewybrane obserwacje ze zbioru danych.** Niewybrane obserwacje są usuwane ze zbioru danych. Usunięte obserwacje można odzyskać wychodząc z pliku bez zachowywania jakichkolwiek zmian a następnie ponownie otwierając plik. Usunięcie obserwacji jest trwałe, jeśli zmiany zostały zapisane na pliku danych.

*Uwaga:* Jeśli niewybrane obserwacje są usunięte a plik zapisany, obserwacje nie mogą być odzyskane.

## Status wyboru obserwacji

Jeśli wybrano podzbiór obserwacji, ale nie odrzucono niewybranych obserwacji, niewybrane obserwacje są oznaczane w Edytorze danych ukośnym przekreśleniem numeru wiersza.

|    |
|----|
| 1  |
| 2  |
| 3  |
| 4  |
| 5  |
| 6  |
| 7  |
| 8  |
| 9  |
| 10 |
| 11 |

Rysunek 89. Status wyboru obserwacji



---

## Rozdział 10. 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,<sup>1</sup> 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.
- **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.
- **behavior.sav.** W klasycznym przykładzie<sup>2</sup> 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

---

<sup>1</sup> 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.

<sup>2</sup> 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.

pomiarów średnicy 16 tarcz hamulcowych pochodzących z każdej z 8 maszyn produkcyjnych. Docelowa średnica tarcz wynosi 322 milimetry.

- **breakfast.sav.** W klasycznym badaniu <sup>3</sup> 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.
- **car\_insurance\_claims.sav.** Zbiór danych przedstawiony i przeanalizowany w innej pracy <sup>4</sup> 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.
- **carpet.sav.** W popularnym przykładzie <sup>5</sup> 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 poziomy czynniki 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.

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

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

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

- **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 mrożonej<sup>6</sup>. 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.
- **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.
- **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.

<sup>6</sup> 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.

- **dietstudy.sav.** Ten plik danych hipotetycznych zawiera wyniki badania nad tzw. dietą Stillmana <sup>7</sup>. 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 podzbiorku 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.
- **german\_credit.sav.** Ten plik danych pochodzi ze zbioru danych „German credit”, znajdującego się w Repository of Machine Learning Databases <sup>8</sup> na Uniwersytecie Kalifornijskim w Irvine.
- **grocery\_1month.sav.** Ten plik danych hipotetyczny to plik danych *grocery\_coupons.sav*, w którym tygodniowe wielkości zakupów zostały zregulowane 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.
- **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 <sup>9</sup> przedstawił tabelę ilustrującą możliwe grupy społeczne. Guttman <sup>10</sup> 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 tłumu, 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.

<sup>7</sup> 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.

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

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

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

- **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 <sup>11</sup> 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.
- **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/](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. <sup>12, 13</sup>, 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 przeciwbólowego 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 („ataku 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 („ataku 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

<sup>11</sup> 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.

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

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

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. Obserwacje odpowiadają 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).
- **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 <sup>14</sup>.

---

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

- **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 pracy<sup>15</sup>, 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 opcji Przygotowanie danych.
- **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, stojących przed różnymi wyzwaniami. 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.
- **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.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.

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



- **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 <sup>16</sup>.
- **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 pracy <sup>17</sup>.
- **verd1985.sav.** Ten plik danych dotyczy badania <sup>18</sup>. 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. Prześledzono (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 dzieci <sup>19</sup>. Dane zawierają binarne pomiary sapania dzieci w wieku 7, 8, 9 i 10 lat z

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

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

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

<sup>19</sup> 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.



miejsowości 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. Prześledzono 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.



## 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 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 JAKIKOLWIEK GWARANCJI (RĘKOJMIĘ RÓWNIEŻ WYŁĄCZA SIĘ), WYRAŻNYCH 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 podmiotów 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) 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 produktami firm innych niż IBM. Pytania dotyczące możliwości produktów podmiotów 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 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 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. 2020. Fragmenty niniejszego kodu pochodzą z programów przykładowych IBM Corp.

© Copyright IBM Corp. 1989 - 2020. Wszelkie prawa zastrzeżone.

## Znaki towarowe

---

IBM, logo IBM oraz [ibm.com](http://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](http://www.ibm.com/legal/copytrade.shtml).

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

Intel, logo Intel, Intel Inside, logo Intel Inside, Intel Centrino, logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium 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

## A

Access (Microsoft) [11](#)

## B

braki danych  
systemowy brak danych [23](#)  
w przypadku zmiennych liczbowych [23](#)  
w przypadku zmiennych nieliczbowych [24](#)

## D

dane ciągłe [25](#)  
dane ilorazowe [25](#)  
dane ilościowe [25](#)  
dane jakościowe [25](#)  
dane kategoryjne  
miary podsumowujące [25](#)  
dane liczbowe [19](#)  
dane łańcuchowe  
wprowadzanie danych [20](#)  
dane nominalne [25](#)  
dane porządkowe [25](#)  
dane przedziału [25](#)

## E

Edytor danych  
wprowadzanie danych innych niż numeryczne [20](#)  
wprowadzanie danych numerycznych [19](#)  
edytowanie tabel przestawnych [40](#)  
eksportowanie wyników  
do programu Excel [50](#)  
do programu PowerPoint [50](#)  
do programu Word [50](#)  
HTML [57](#)  
etykiety wartości  
przypisywanie [22](#)  
sterowanie wyświetlaniem w oknie raportu [22](#)  
zmiennie liczbowe [22](#)  
etykiety zmiennych  
tworzenie [21](#)  
Excel (Microsoft)  
eksportowanie wyników do [50](#)

## F

funkcje w wyrażeniach [66](#)

## H

histogramy [28](#)  
HTML  
eksportowanie wyników [57](#)

## K

kreator importu tekstu [14](#)  
Kreator obliczeń daty i czasu [68](#)  
Kreator przejmowania baz danych [11](#)

## L

liczebności  
tabele liczebności [25](#)

## M

miary podsumowujące  
dane kategoryjne [25](#)  
zmiennie ilościowe [27](#)

## O

obliczanie nowych zmiennych [65](#)  
obserwacji  
sortowanie [71](#), [73](#)  
wybór [73](#)  
okna komend  
automatyczne uzupełnianie [60](#)  
edytowanie komend [60](#)  
punkty przerywające [61](#)  
system kolorów [60](#)  
uruchamianie komend [59](#), [61](#)  
wklejanie komend [59](#)  
okno raportu  
przenoszenie wyników [37](#)  
ukrywanie i wyświetlanie wyników [37](#)

## P

pliki arkuszy kalkulacyjnych  
odczytywanie [8](#)  
odczytywanie nazw zmiennych [8](#)  
pliki bazy danych  
odczytywanie [11](#)  
pliki danych tekstowych  
odczytywanie [14](#)  
pliki programu Excel  
odczytywanie [8](#)  
pliki przykładowe  
lokalizacja [77](#)  
pliki składni  
otwieranie [61](#)  
podzbiory obserwacji  
filtrowanie niewybranych obserwacji [76](#)  
jeśli spełniony jest warunek [74](#)  
na podstawie dat i godzin [75](#)  
próba losowa [75](#)  
usuwanie niewybranych obserwacji [76](#)  
wybór [73](#)

podzbiory obserwacji (*kontynuacja*)  
wyrażenia warunkowe [74](#)  
pomoc języka komend [60](#)  
PowerPoint (Microsoft)  
eksportowanie wyników do [50](#)  
poziom pomiaru [25](#)  
przenoszenie  
elementy w oknie raportu [37](#)  
elementy w tabelach przestawnych [39](#)  
przetwarzanie podzielonych danych [71](#)

## R

rekodowanie wartości [63](#)

## S

składnia [59](#)  
sortowanie obserwacji [71](#)  
systemowe braki danych [23](#)

## T

tabele częstości [25](#)  
tabele przestawne  
dostęp do definicji [38](#)  
edycja [40](#)  
formatowanie [40](#)  
formaty komórek [42](#)  
panel przestawiania [39](#)  
transponowanie wierszy i kolumn [39](#)  
typy danych w komórkach [42](#)  
ukrywanie separatorów dziesiętnych [42](#)  
ukrywanie wierszy i kolumn [41](#)  
warstwy [40](#)  
transponowanie (zamiana miejscami) wierszy i kolumn w tabelach przestawnych [39](#)  
tworzenie etykiet zmiennych [21](#)  
typy danych  
dla zmiennych [22](#)

## U

ukrywanie wierszy i kolumn w tabelach przestawnych [41](#)

## W

warstwy  
tworzenie w tabelach przestawnych [40](#)  
wklejanie komend  
z okna dialogowego [59](#)  
Word (Microsoft)  
eksportowanie wyników do [50](#)  
wprowadzanie danych  
nienumerycznych [20](#)  
numerycznych [19](#)  
wybór obserwacji [73](#)  
wykresy  
histogramy [28](#)  
stupkowe [26](#), [31](#)  
tworzenie diagramów [31](#)  
tworzenie wykresów [31](#)  
wykresy stupkowe [26](#)

wyrażenia warunkowe [67](#)

## Z

zmiennie  
etykiety [21](#)  
typy danych [22](#)  
zmiennie daty i czasu [68](#)  
zmiennie ilościowe  
miary podsumowujące [27](#)





