

*Guida introduttiva per IBM SPSS  
Statistics 24*

**IBM**

**Nota**

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in "Informazioni particolari" a pagina 93.

**Informazioni sul prodotto**

Questa edizione si applica alla versione 24, release 0, modifica 0 di IBM SPSS Statistics e a tutte le release e modifiche successive se non diversamente indicato in nuove edizioni.

# Indice

## Capitolo 1. Introduzione . . . . . 1

|                                       |   |
|---------------------------------------|---|
| File di esempio . . . . .             | 1 |
| Apertura di un file di dati . . . . . | 1 |
| Esecuzione di un'analisi . . . . .    | 3 |
| Creazione di grafici . . . . .        | 4 |

## Capitolo 2. Lettura dei dati. . . . . 7

|                                                                     |    |
|---------------------------------------------------------------------|----|
| Struttura di base dei file di dati di IBM SPSS Statistics . . . . . | 7  |
| Letture dei file di dati di IBM SPSS Statistics . . . . .           | 7  |
| Letture dei dati Excel . . . . .                                    | 8  |
| Letture dei dati di un database . . . . .                           | 11 |
| Letture dei dati di un file di testo . . . . .                      | 14 |

## Capitolo 3. Utilizzo dell'Editor dei dati 19

|                                                        |    |
|--------------------------------------------------------|----|
| Inserimento di dati numerici. . . . .                  | 19 |
| Inserimento di dati di stringa . . . . .               | 20 |
| Definizione di dati . . . . .                          | 21 |
| Aggiunta delle etichette di variabile . . . . .        | 21 |
| Modifica del formato e del tipo di variabile . . . . . | 22 |
| Aggiunta di etichette valore . . . . .                 | 23 |
| Gestione dei dati mancanti . . . . .                   | 23 |
| Valori mancanti per una variabile numerica . . . . .   | 24 |
| Valori mancanti per una variabile stringa . . . . .    | 24 |

## Capitolo 4. Esame delle statistiche di riepilogo per singole variabili . . . . . 27

|                                                     |    |
|-----------------------------------------------------|----|
| Livello di misurazione. . . . .                     | 27 |
| Misure di riepilogo per dati categoriali . . . . .  | 27 |
| Grafici per i dati categoriali . . . . .            | 28 |
| Misure di riepilogo per variabili di scala. . . . . | 29 |
| Istogrammi per variabili di scala . . . . .         | 30 |

## Capitolo 5. Creazione e modifica di grafici . . . . . 33

|                                                           |    |
|-----------------------------------------------------------|----|
| Nozioni base per la creazione di grafici . . . . .        | 33 |
| Utilizzo di Modelli del Builder di grafico . . . . .      | 34 |
| Definizione delle variabili e delle statistiche . . . . . | 34 |
| Aggiunta di testo . . . . .                               | 36 |
| Per creare il grafico. . . . .                            | 37 |

## Capitolo 6. Utilizzo dell'output. . . . . 39

|                                                 |    |
|-------------------------------------------------|----|
| Utilizzo del Visualizzatore . . . . .           | 39 |
| Utilizzo dell'Editor di tabelle pivot . . . . . | 40 |
| Accesso alle definizioni dell'output . . . . .  | 40 |
| Tabelle pivot . . . . .                         | 41 |
| Creazione e visualizzazione di strati . . . . . | 42 |
| Modifica di tabelle . . . . .                   | 43 |
| Nascondere righe e colonne . . . . .            | 44 |
| Modifica dei formati di vista dati . . . . .    | 44 |
| Tablelook . . . . .                             | 45 |
| Utilizzo di formati predefiniti . . . . .       | 46 |

|                                                                                     |    |
|-------------------------------------------------------------------------------------|----|
| Personalizzazione di stili TableLook . . . . .                                      | 46 |
| Modifica dei formati di tabella predefiniti . . . . .                               | 49 |
| Personalizzazione delle impostazioni iniziali di visualizzazione . . . . .          | 50 |
| Visualizzazione delle etichette valore e variabili . . . . .                        | 51 |
| Utilizzo dell'output in altre applicazioni . . . . .                                | 52 |
| Incollare i risultati come tabelle Word . . . . .                                   | 53 |
| Incollare i risultati come testo . . . . .                                          | 53 |
| Esportazione dei risultati in file di Microsoft Word, PowerPoint ed Excel . . . . . | 54 |
| Esportazione dei risultati in formato PDF . . . . .                                 | 59 |
| Esportazione dei risultati in formato HTML . . . . .                                | 61 |

## Capitolo 7. Utilizzo della sintassi . . . . . 63

|                                                         |    |
|---------------------------------------------------------|----|
| Incollare la sintassi . . . . .                         | 63 |
| Modificare la sintassi . . . . .                        | 64 |
| Apertura ed esecuzione di un file di sintassi . . . . . | 65 |
| Utilizzo dei punti di interruzione . . . . .            | 65 |

## Capitolo 8. Modifica dei valori dei dati 67

|                                                                           |    |
|---------------------------------------------------------------------------|----|
| Creazione di una variabile categoriale da una variabile di scala. . . . . | 67 |
| Calcolo di nuove variabili . . . . .                                      | 69 |
| Utilizzo delle funzioni nelle espressioni . . . . .                       | 70 |
| Utilizzo delle espressioni condizionali . . . . .                         | 71 |
| Utilizzo di date e orari . . . . .                                        | 72 |
| Calcolo del periodo di tempo tra due date . . . . .                       | 73 |
| Aggiunta di una durata a una data . . . . .                               | 74 |

## Capitolo 9. Ordinamento e selezione di dati . . . . . 75

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Ordinamento di dati . . . . .                                                        | 75 |
| Elaborazione file suddiviso . . . . .                                                | 76 |
| Ordinamento dei casi per l'elaborazione file suddiviso . . . . .                     | 77 |
| Attivazione e disattivazione dell'elaborazione file suddiviso . . . . .              | 77 |
| Selezione di sottoinsiemi di casi . . . . .                                          | 77 |
| Per selezionare sottoinsiemi di casi in base a un'espressione condizionale . . . . . | 78 |
| Per selezionare un campione casuale di dati . . . . .                                | 79 |
| Selezione di un intervallo di tempo o di casi . . . . .                              | 80 |
| Trattamento dei casi non selezionati . . . . .                                       | 80 |
| Criteri di selezione dei casi . . . . .                                              | 81 |

## Capitolo 10. File di esempio . . . . . 83

## Informazioni particolari. . . . . 93

|                  |    |
|------------------|----|
| Marchi . . . . . | 95 |
|------------------|----|

## Indice analitico . . . . . 97



---

## Capitolo 1. Introduzione

Questa guida illustrerà come utilizzare molte delle funzioni disponibili. Si tratta di una guida pratica con procedure visualizzate. Tutti i file mostrati negli esempi vengono installati con l'applicazione in modo da poter seguire le istruzioni, eseguendo le stesse analisi ed ottenendo gli stessi risultati visualizzati qui.

Se si desiderano esempi dettagliati delle varie tecniche di analisi statistiche, provare con Studi di casi, disponibile dal menu Guida.

---

### File di esempio

La maggior parte degli esempi illustrati utilizzano il file di dati *demo.sav*. Si tratta di un file di dati contenente un'indagine fittizia realizzata su un campione di un migliaio di persone con informazioni di base di tipo demografico e consumistico.

Se si sta utilizzando la versione per studenti, il file *demo.sav* è un campione rappresentativo del file di dati originale, ridotto per soddisfare il limite di 1.500 casi. I risultati che si otterranno utilizzando tale file di dati saranno diversi da quelli mostrati qui.

Il file di esempio installato con il prodotto si trova nella sottocartella *Samples* della cartella di installazione. È presente una cartella separata all'interno della sottocartella *Samples* per ciascuna delle seguenti lingue: inglese, francese, tedesco, italiano, giapponese, coreano, polacco, russo, cinese semplificato, spagnolo e cinese tradizionale.

Non tutti i file di esempio sono disponibili in tutte le lingue. Se un file di esempio non è disponibile in una lingua, la cartella di tale lingua contiene una versione inglese del file.

---

### Apertura di un file di dati

Per aprire un file di dati:

1. Dai menu, scegliere:

**File > Apri > Dati...**

Viene visualizzata una finestra di dialogo per l'apertura dei file.

Per impostazione predefinita, vengono visualizzati i file di dati in formato IBM® SPSS Statistics (con estensione *.sav*).

In questo esempio viene utilizzato il file *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. |

Figura 1. File demo.sav nell'Editor dei dati

Il file di dati viene visualizzato nell'Editor dei dati. Nella Vista dati, se si posiziona il cursore del mouse su un nome variabile (le intestazioni delle colonne), viene visualizzata un'etichetta di variabile più descrittiva (se è stata definita un'etichetta per quella variabile).

Per impostazione predefinita, vengono visualizzati i valori effettivi dei dati. Per visualizzare le etichette:

2. Dai menu, scegliere:

**Visualizza > Etichette valore**



Figura 2. Pulsante Etichette valore

È anche possibile utilizzare il pulsante Etichette valore della barra degli strumenti.

|    | 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. |

Figura 3. Etichette valore visualizzate nell'Editor dei dati

A questo punto vengono visualizzate le etichette valore descrittive per semplificare l'interpretazione delle risposte.

## Esecuzione di un'analisi

Se si dispone delle opzioni aggiuntive, il menu **Analizza** contiene un elenco delle categorie di analisi statistica e di reporting.

Il primo esempio si riferisce a una tabella delle frequenze (tabella di conteggi). Questo esempio richiede il modulo **Statistics Base**.

1. Dai menu, scegliere:

**Analizza > Statistiche descrittive > Frequenze...**

Si visualizza la finestra di dialogo **Frequenze**.

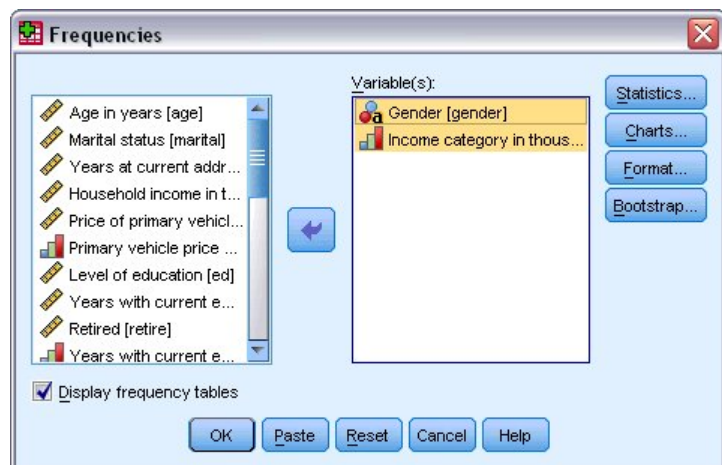


Figura 4. Finestra di dialogo *Frequenze*

L'icona accanto a ogni variabile consente di ottenere le informazioni sul tipo di dati e sul livello di misurazione.

|                  | Numerico | Stringa | Dati | Ora |
|------------------|----------|---------|------|-----|
| Scala (continuo) |          | n/d     |      |     |
| Ordinale         |          |         |      |     |
| Nominale         |          |         |      |     |

Se l'etichetta e/o il nome della variabile appare troncato nell'elenco, la dicitura completa viene visualizzata quando il cursore si sofferma sull'etichetta/sul nome. Il nome di variabile *catredd* viene visualizzato tra parentesi quadre dopo l'etichetta descrittiva di variabile. *Categoria reddito in migliaia* è l'etichetta di variabile. Se non fosse presente alcuna etichetta di variabile, nella casella di elenco verrebbe visualizzato solo il nome della variabile.

È possibile ridimensionare le finestre di dialogo proprio come per le finestre, selezionando e trascinando i bordi o gli angoli esterni. Ad esempio, se la finestra di dialogo viene allargata, anche gli elenchi delle variabili risulteranno più ampi.

Nella finestra di dialogo, scegliere le variabili che si desidera analizzare dall'elenco di origine a sinistra e trascinarle e rilasciarle nell'elenco Variabili a destra. Il pulsante **OK** che esegue l'analisi viene disabilitato finché almeno una variabile non viene inserita nell'elenco Variabili.

In molte finestre di dialogo è possibile ottenere ulteriori informazioni facendo clic con il tasto destro del mouse su qualsiasi nome variabile dell'elenco e selezionando **Informazioni sulla variabile** dal menu a comparsa.

2. Fare clic su *Sesso [sesso]* nell'elenco delle variabili origine e trascinare la variabile nell'elenco Variabili obiettivo.
3. Quindi fare clic su *Categoria di reddito in migliaia [catredd]* nell'elenco origine e trascinarla nell'elenco di destinazione.

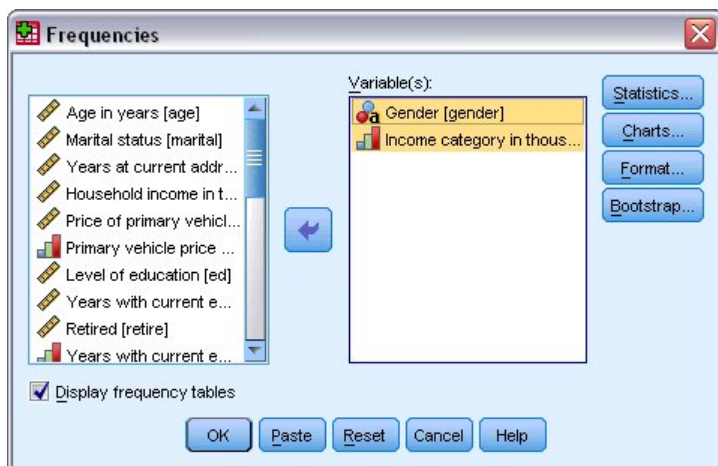


Figura 5. Variabili selezionate per l'analisi

4. Fare clic su **OK** per eseguire la procedura.

I risultati vengono visualizzati nel Visualizzatore.

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

|       |             | Frequency | Percent | Valid Percent | Cumul: Percent |
|-------|-------------|-----------|---------|---------------|----------------|
| 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         |                |

Figura 6. Tabella delle frequenze per le categorie di reddito

## Creazione di grafici

Sebbene alcune procedure statistiche possano creare grafici, è anche possibile utilizzare il menu Grafici per crearli.



Ad esempio, è possibile creare un grafico che mostra la relazione esistente tra titolarità di servizi telefonici senza fili e PDA (Personal Digital Assistant).

1. Dai menu, scegliere:

**Grafici > Builder di grafico...**

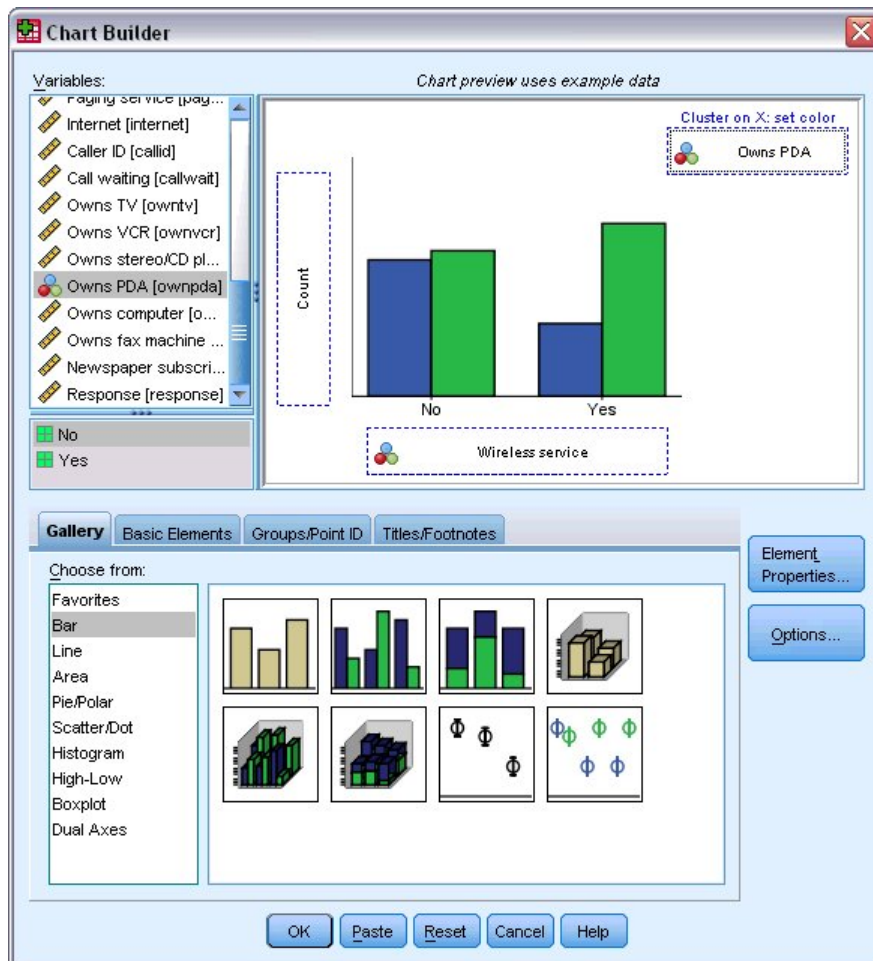


Figura 7. Finestra di dialogo Builder di grafico con zone di rilascio complete

2. Fare clic sulla scheda **Modelli** (se non è selezionata).
3. Fare clic su **A barre** (se non è selezionata).
4. Trascinare l'icona A barre raggruppate nell'area, ovvero nell'area estesa situata sopra i Modelli.
5. Scorrere verso il basso l'elenco Variabili, fare clic con il pulsante destro del mouse su *Telefono cellulare [cellul]*, quindi selezionare **Nominale** come rispettivo livello di misurazione.
6. Trascinare la variabile *Telefono cellulare [cellul]* nell'asse *x*.
7. Fare clic su *Possiede PDA [pda]* e selezionare **Nominale** come rispettivo livello di misurazione.
8. Trascinare la variabile *Possiede PDA [pda]* nella zona di rilascio di raggruppamento situata nell'angolo destro superiore dell'area.
9. E fare clic su **OK** per creare il grafico.

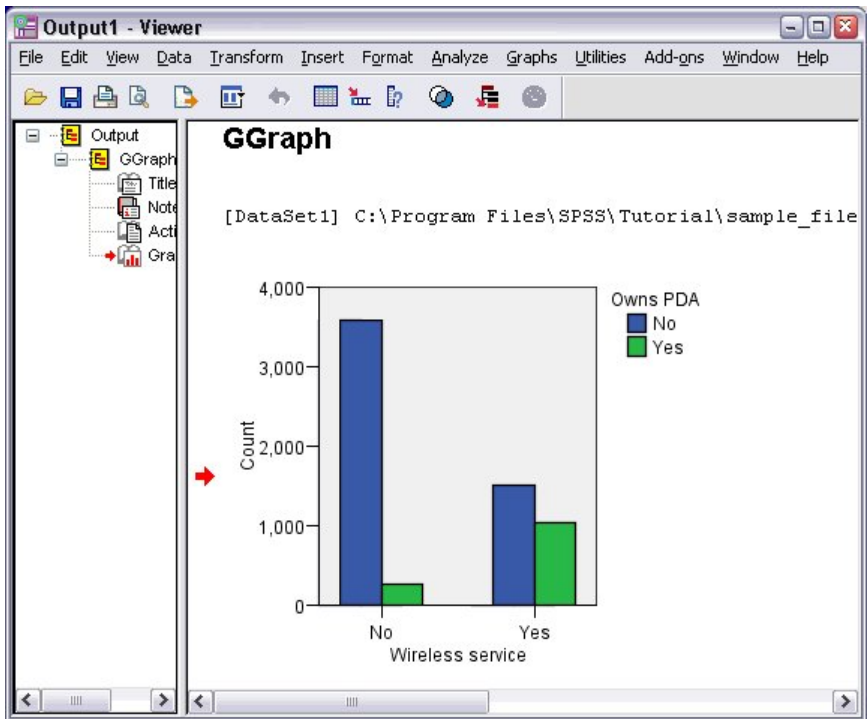


Figura 8. Grafico a barre visualizzato nella finestra del Visualizzatore

Il grafico a barre viene visualizzato nel Visualizzatore. Il grafico mostra che chi possiede un servizio di telefonia senza fili è più propenso ad avere un PDA rispetto a chi non lo possiede.

È possibile modificare grafici e tabelle facendo doppio clic su di essi nel riquadro del contenuto della finestra Visualizzatore, quindi è possibile copiare e incollare i risultati in altre applicazioni. Questi argomenti verranno trattati in seguito.

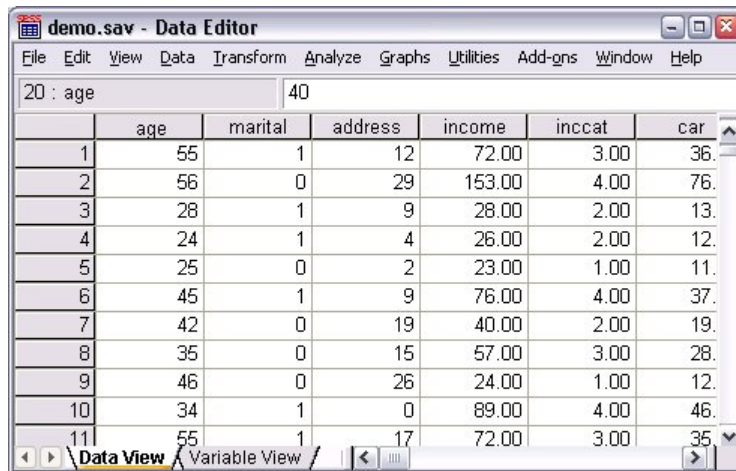
---

## Capitolo 2. Lettura dei dati

È possibile inserire i dati direttamente oppure importarli da origini diverse. In questo capitolo vengono descritte le procedure per la lettura dei dati in file di dati di IBM SPSS Statistics, in fogli di calcolo quali Microsoft Excel, in applicazioni database quali Microsoft Access e nei file di testo.

---

### Struttura di base dei file di dati di IBM SPSS Statistics



The screenshot shows the 'demo.sav - Data Editor' window. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The main area displays a table with 11 rows and 8 columns. The columns are labeled 'age', 'marital', 'address', 'income', 'inccat', and 'car'. The first row is highlighted. The status bar at the bottom indicates 'Data View' and 'Variable View'.

|    | 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. |

Figura 9. Editor dei dati

I file di dati in formato IBM SPSS Statistics sono organizzati per casi (righe) e per variabili (colonne). In questo file di dati, i casi rappresentano i singoli rispondenti di un'indagine, mentre le variabili rappresentano le risposte fornite a ogni quesito dell'indagine.

---

### Letture dei file di dati di IBM SPSS Statistics

I file di dati di IBM SPSS Statistics hanno estensione *.sav* e contengono i dati salvati.

1. Dai menu, scegliere:  
**File > Apri > Dati...**
2. Spostarsi e aprire il file *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

I dati vengono visualizzati nell'Editor dei dati.

|    | 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. |

Figura 10. File di dati aperto

## Letture dei dati Excel

Anziché digitare direttamente i dati nell'Editor dei dati, è possibile leggere i dati da applicazioni quali Microsoft Excel. È inoltre possibile leggere le intestazioni delle colonne come nomi di variabili.

1. Dai menu, scegliere:

**File > Importa dati > Excel**

2. Andare alla cartella Samples\English e selezionare demo.xlsx.

La finestra di dialogo Leggi file Excel visualizza un'anteprima del file di dati. Viene visualizzato il contenuto del primo foglio nel file. Se il file ha più fogli, è possibile selezionare il foglio dall'elenco.

Si noterà che alcuni dei valori stringa per *Sesso* hanno spazi iniziali. Alcuni dei valori per *StatoCivile* sono visualizzati come punti (.).

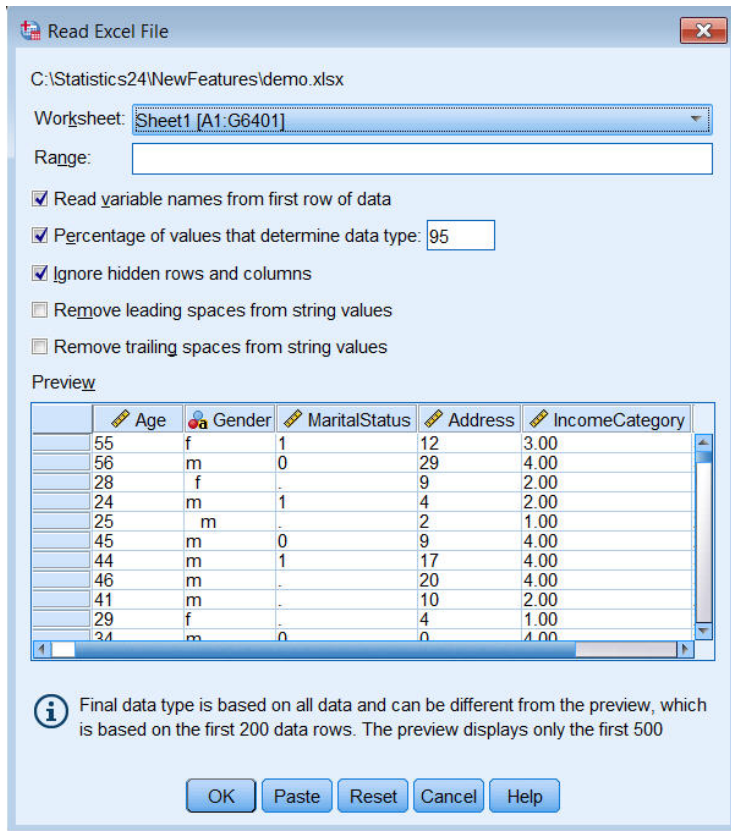
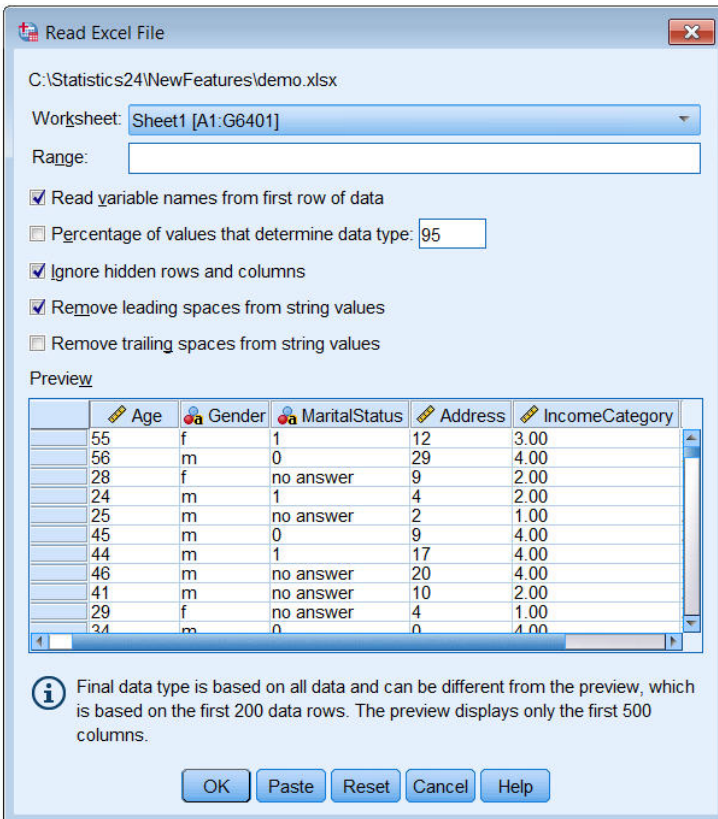


Figura 11. Finestra di dialogo Leggi file Excel

3. Assicurarsi che l'opzione **Leggi i nomi delle variabili dalla prima riga di dati** sia selezionata. Se le intestazioni delle colonne non sono conformi alle regole per la denominazione delle variabili, vengono convertite in nomi di variabili validi. Le intestazioni di colonna originali vengono salvate come etichette di variabile.
4. Selezionare **Rimuovi spazi iniziali dai valori stringa**.
5. Deselezionare **Percentuale di valori che determinano il tipo di dati**.



Il valore stringa "nessuna risposta" ora viene visualizzato nelle celle che erano mancanti di sistema. Se non è presente il parametro per la percentuale di valori e la colonna contiene una combinazione di tipi di dati, la variabile viene letta come tipo di dati stringa. Tutti i valori vengono mantenuti, ma i valori numerici sono considerati stringhe.

6. Selezionare **Percentuale di valori che determinano il tipo di dati** per considerare *StatoCivile* come una variabile numerica.
7. Fare clic su **OK** per leggere il file di Excel.

I dati vengono visualizzati nell'Editor dei dati, con le intestazioni delle colonne utilizzate come nomi di variabili. Poiché i nomi delle variabili non possono contenere degli spazi, sono rimossi gli spazi dalle intestazioni delle colonne originali. Ad esempio, l'intestazione di colonna "Stato civile" viene convertita nella variabile *StatoCivile*. L'intestazione di colonna originale viene mantenuta come etichetta di variabile.

The screenshot shows the IBM SPSS Statistics Data Editor window with a dataset containing 10 rows of data. The variables are Age, Gender, MaritalStatus, Address, IncomeCategory, and JobCategory. The data is displayed in a grid format with row numbers 1 through 10 on the left. The bottom of the window shows 'Data View' selected.

|    | 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           |

Figura 12. Dati di Excel importati

#### Informazioni correlate:

Capitolo 10, "File di esempio", a pagina 83

## Lettura dei dati di un database

La Procedura guidata di database semplifica l'importazione dei dati da un database. I database che utilizzano driver ODBC (Open Database Connectivity) possono essere letti direttamente dopo l'installazione dei driver. Nel CD di installazione sono disponibili i driver ODBC per numerosi formati di database. Per ottenere driver aggiuntivi, è possibile rivolgersi a fornitori di terze parti. In questo esempio viene utilizzata una delle applicazioni di database più comuni, ovvero Microsoft Access.

*Nota:* questo esempio è specifico per Microsoft Windows e richiede un driver ODBC per Access. Il driver ODBC Microsoft Access funziona solo con la versione a 32 bit di IBM SPSS Statistics. Le fasi sono simili a quelle di altre piattaforme, ma potrebbero richiedere un driver ODBC di terze parti per Access.

1. Dai menu, scegliere:  
**File > Importa dati > Database > Nuova query...**

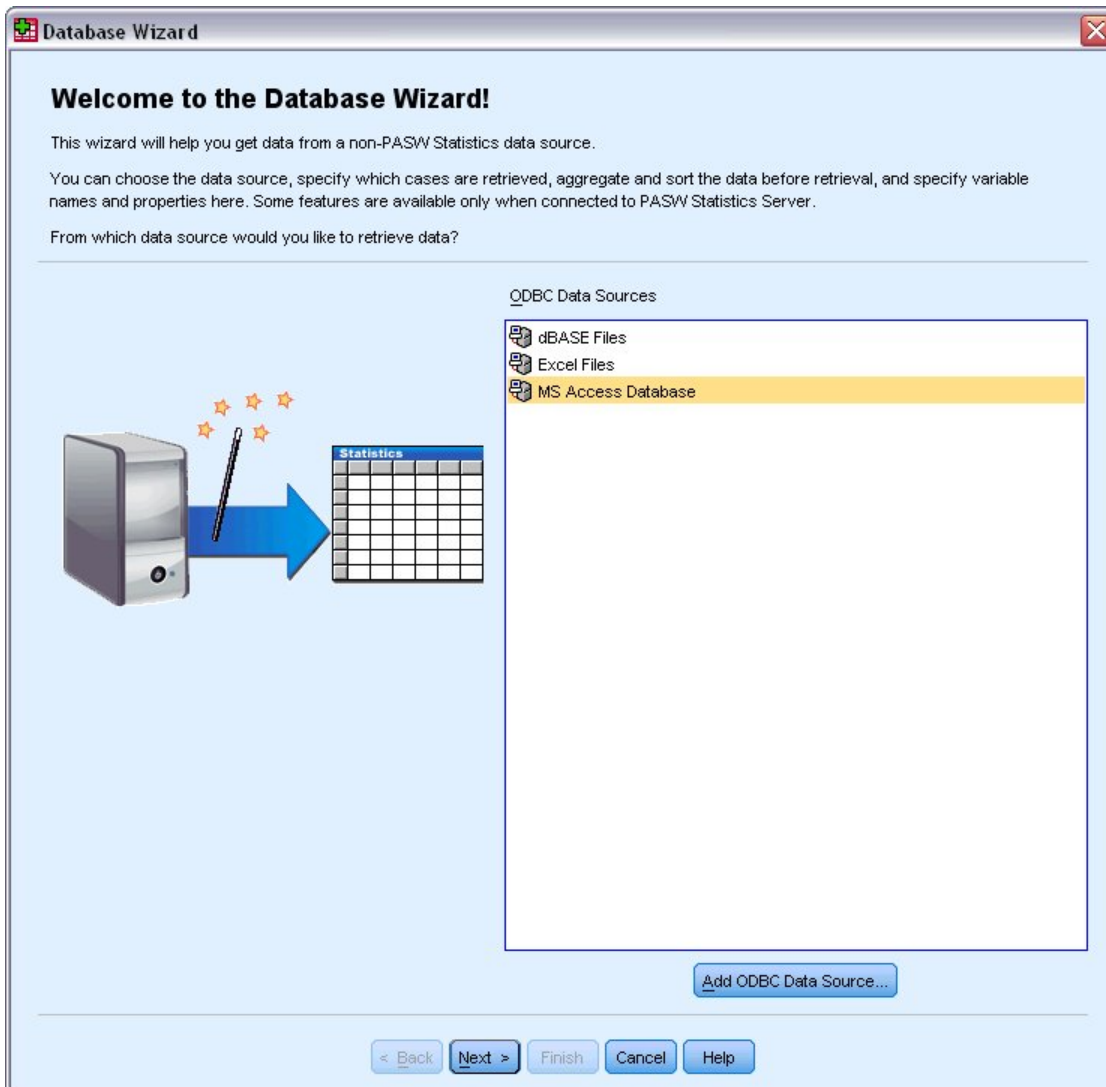


Figura 13. Finestra di dialogo di benvenuto di Procedura guidata di database

2. Selezionare **Database di Microsoft Access** nell'elenco delle origini dati e fare clic su **Avanti**.  
*Nota:* in base all'installazione, potrebbe essere visualizzato anche un elenco di origini dati OLEDB sul lato sinistro della procedura guidata (solo per sistemi operativi Windows), ma questo esempio utilizza l'elenco di origini dati ODBC visualizzato sul lato destro.
3. Fare clic su **Sfogliare** per spostarsi sul file del database di Access che si desidera aprire.
4. Aprire *demo.mdb*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.
5. Fare clic su **OK** nella finestra di dialogo di login.  
 Nella fase successiva è possibile specificare le tabelle e le variabili che si desidera importare.



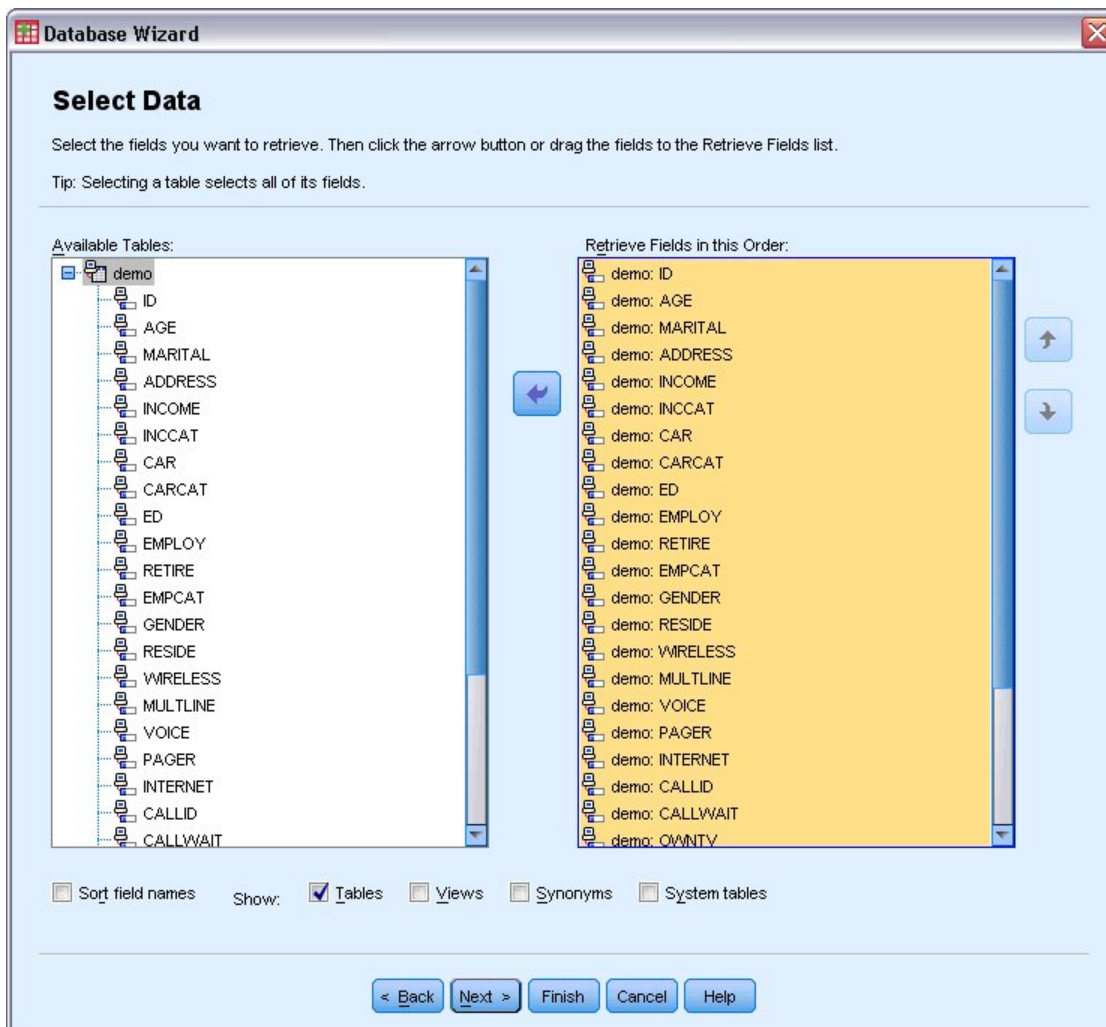


Figura 14. Fase Selezione dati

6. Trascinare l'intera tabella della **demo** nell'elenco Importa i campi in questo ordine.
7. Fare clic su **Avanti**.

Nella fase successiva è possibile selezionare quali record (casi) importare.

Se non si desidera importare tutti i casi, è possibile importare un sottoinsieme di casi, ad esempio i maschi con età superiore a 30 anni, oppure un campione casuale di casi dall'origine dati. Per le origini dati di grandi dimensioni, è possibile ridurre il tempo di elaborazione limitando il numero dei casi a un campione piccolo e rappresentativo.

8. Fare clic su **Avanti** per continuare.

I nomi dei campi vengono utilizzati per creare nomi di variabili. Se necessario, i nomi vengono convertiti in nomi di variabili validi. I nomi dei campi originali vengono conservati come etichette di variabile. È possibile, inoltre, modificare i nomi delle variabili prima di importare il database.

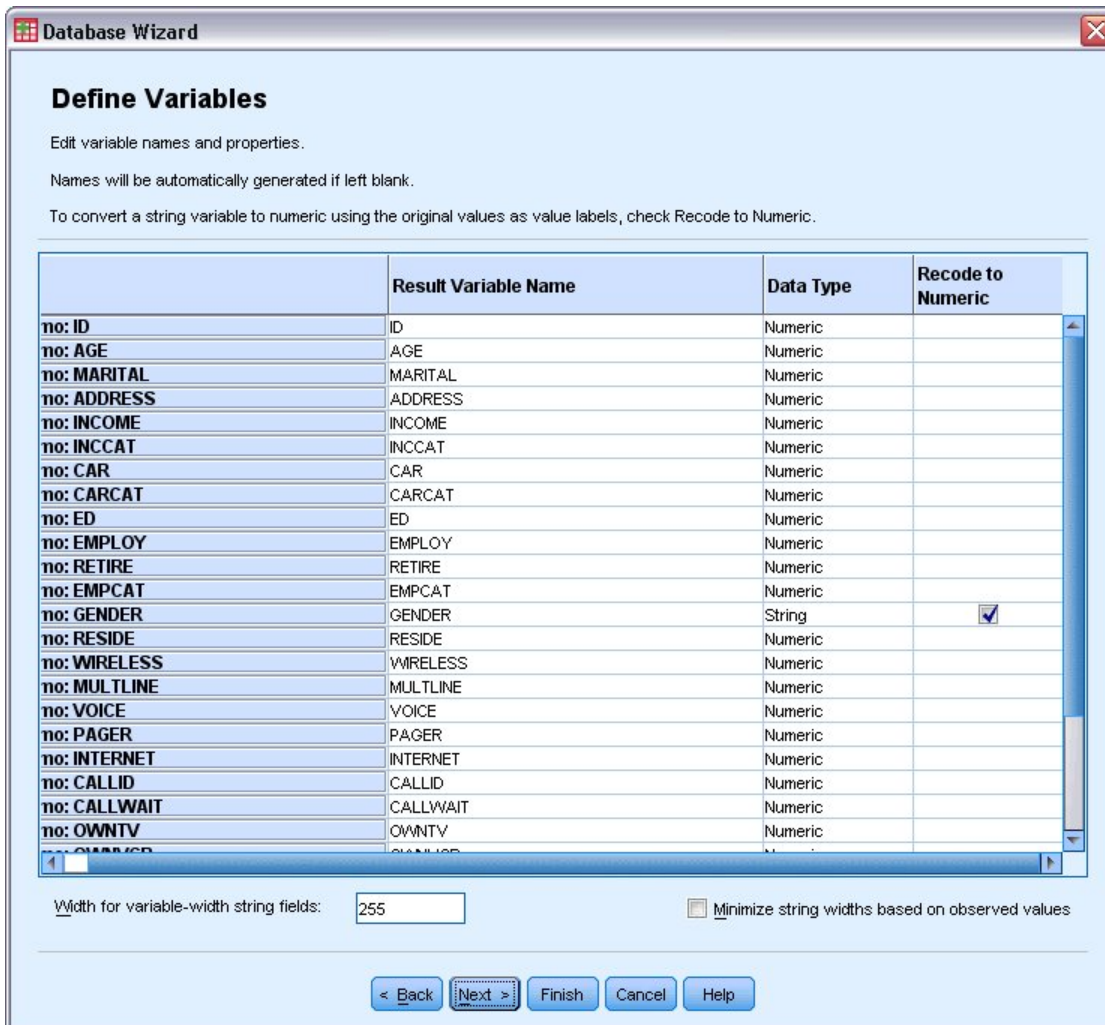


Figura 15. Fase Definisci variabili

9. Fare clic sulla cella **Ricodifica in numerica** nel campo Sesso. Questa opzione converte le variabili stringa in variabili intero e mantiene il valore originale dell'etichetta per la nuova variabile.

10. Fare clic su **Avanti** per continuare.

L'istruzione SQL creata in base alle selezioni in Procedura guidata di database viene visualizzata nella fase Risultati e può essere eseguita immediatamente oppure può essere salvata in un file per utilizzarla in seguito.

11. Fare clic su **Fine** per importare i dati.

Tutti i dati del database di Access selezionati per l'importazione sono ora disponibili nell'Editor dei dati.

## Letture dei dati di un file di testo

I file di testo rappresentano un'altra origine dati di uso comune. Molti fogli di calcolo e database sono in grado di salvare i contenuti in numerosi formati di file di testo. Nei file delimitati da virgole o da tabulazioni, viene fatto riferimento a righe di dati che utilizzano le virgole o le tabulazioni per indicare ciascuna variabile. In questo esempio, i dati sono delimitati da tabulazioni.

1. Dai menu, scegliere:

**File > Importa dati > Dati di testo**

2. Andare alla cartella Samples\English e selezionare demo.txt.

Procedura guidata di importazione di testo consente di definire la modalità di interpretazione del file di testo specificato.

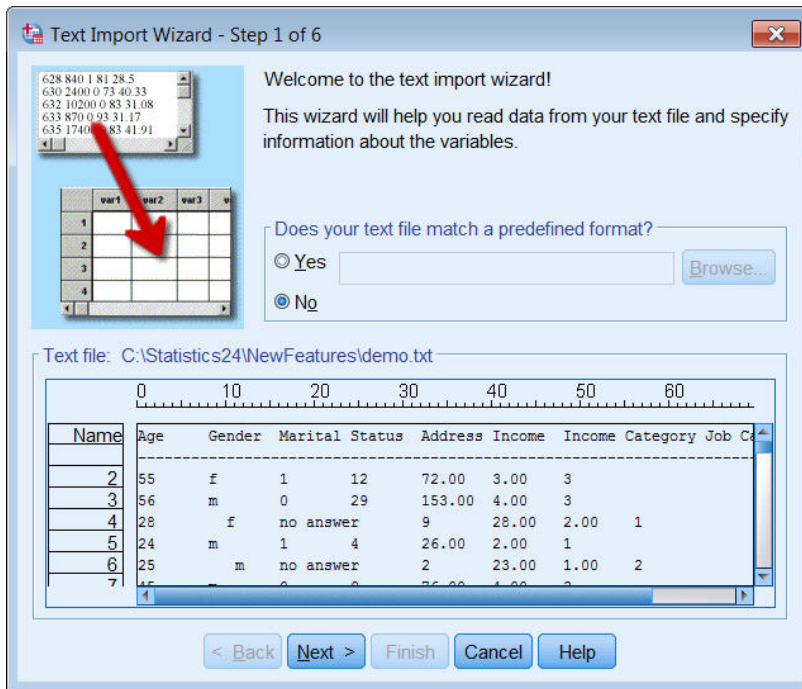


Figura 16. Procedura guidata Importazione testo: Fase 1 di 6

3. Nella Fase 1, è possibile scegliere un formato predefinito oppure crearne uno nuovo. Selezionare **No**.
4. Fare clic su **Avanti** per continuare.  
Come indicato in precedenza, in questo file viene utilizzato un formato delimitato da tabulazioni. I nomi delle variabili vengono inoltre definiti nella riga iniziale del file.
5. Nella fase 2 della procedura guidata selezionare **Delimitato** per indicare che il file utilizza una struttura di formattazione delimitata.
6. Selezionare **Sì** per specificare che il file include i nomi delle variabili nella parte iniziale del file.
7. Fare clic su **Avanti** per continuare.
8. Nella fase 3, immettere 2 per il numero riga in cui inizia il primo caso di dati (in quanto i nomi variabile sono nella prima riga).
9. Mantenere i valori predefiniti per il resto di questa fase e far clic su **Avanti** per continuare.  
In Anteprima dati (fase 4) è possibile verificare in modo rapido che il file venga letto correttamente.
10. Selezionare **Tabulatore** e deselegionare le altre opzioni per i delimitatori. **Spazio** è selezionato per impostazione predefinita perché il file contiene spazi. Per questo file, gli spazi fanno parte dei valori di dati, non sono delimitatori. È necessario deselegionare **Spazio** per leggere il file correttamente.
11. Selezionare **Rimuovi spazi iniziali dai valori stringa**. Gli spazi all'inizio dei valori stringa influenzano la modalità di calcolo nelle espressioni dei valori stringa. In questo file, alcuni valori per *Sesso* hanno spazi iniziali che non fanno parte del valore. Se non si rimuovono tali spazi, un valore " f" viene considerato come un valore diverso da "f".

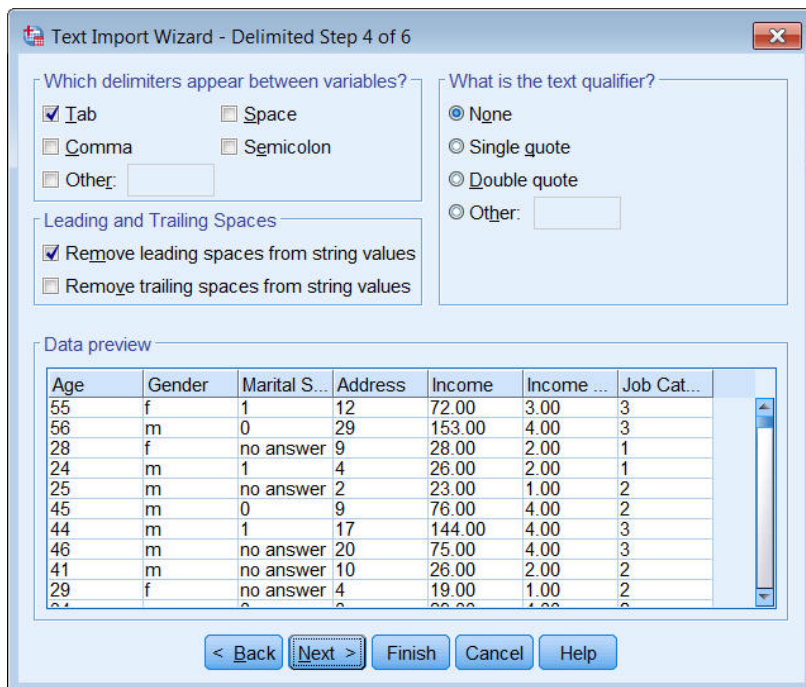


Figura 17. Procedura guidata Importazione testo: Fase 4 di 6

12. Fare clic su **Avanti** per continuare.

Poiché i nomi variabile vengono modificati per conformità alle regole di denominazione, la fase 5 fornisce l'opportunità di modificare gli eventuali nomi non graditi.

In questa finestra è inoltre possibile definire i tipi di dati. Ad esempio, è possibile modificare *Reddito* nel formato valuta dollaro.

Per modificare un tipo di dati:

13. In **Anteprima dati**, selezionare *Reddito*.

14. Selezionare **Euro** nell'elenco a discesa Formato dei dati.

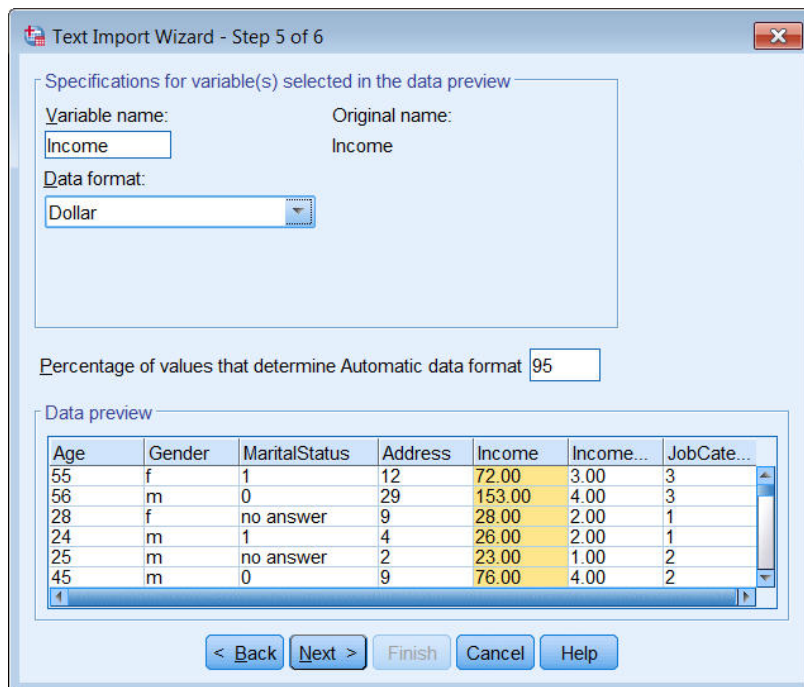


Figura 18. Modificare il tipo di dati

La variabile *StatoCivile* contiene sia valori stringa che numerici. Meno del cinque per cento dei valori è di tipo stringa. Con l'impostazione predefinita del 95% per la **Percentuale di valori che determinano il formato dati Automatic**, la variabile viene considerata come numerica e i valori stringa sono impostati su mancanti di sistema. Se nessun formato dati rispetta il valore di percentuale, la variabile viene considerata come variabile stringa. Se si modifica l'impostazione in 100, vengono mantenuti tutti i valori, ma i valori numerici sono considerati stringhe.

15. Fare clic su **Avanti** per continuare.
16. Lasciare le selezioni predefinite nell'ultima fase e fare clic su **Fine** per importare i dati.



---

## Capitolo 3. Utilizzo dell'Editor dei dati

Nell'Editor dei dati viene visualizzato il contenuto del file di dati attivo e le informazioni sono rappresentate da variabili e casi.

- Nella Vista dati, le colonne rappresentano le variabili e le righe i casi (osservazioni).
- Nella Vista variabile, ogni riga rappresenta una variabile e ogni colonna rappresenta un attributo associato alla variabile specifica.

Le variabili consentono di rappresentare i diversi tipi di dati che sono stati compilati. Un'analogia frequente è quella dell'indagine, in cui la risposta a ognuna delle domande equivale a una variabile. Esistono molti tipi diversi di variabili, ad esempio le variabili numeriche, le variabili stringa, le variabili di valuta e di dati.

---

### Inserimento di dati numerici

È possibile digitare dati nell'Editor dei dati e ciò può risultare utile per i file di dati di piccole dimensioni o per apportare modifiche di lieve entità ai file di dati più grandi.

1. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.

È necessario definire le variabili che si desidera utilizzare. In questo caso, sono necessarie solo tre variabili: *età*, *stato civile* e *reddito*.

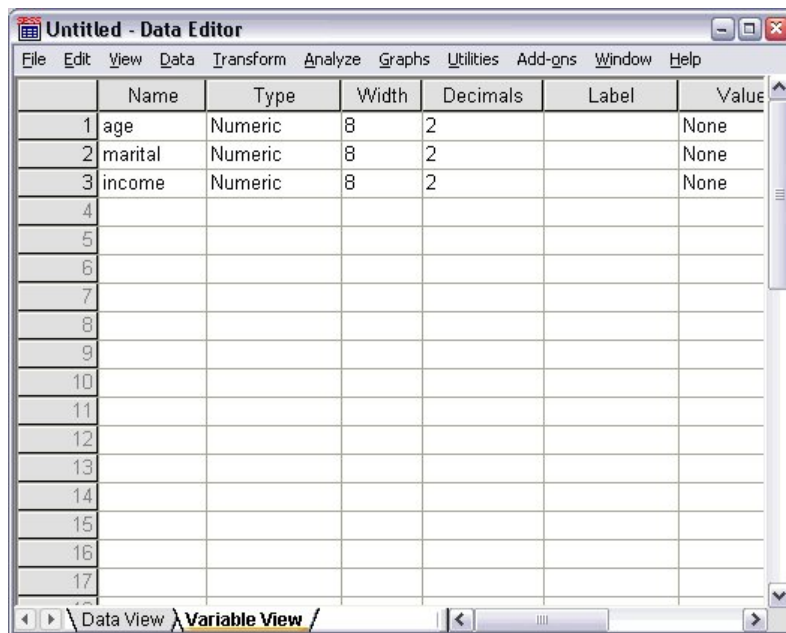


Figura 19. Nomi di variabili in Vista variabile

2. Nella prima riga della prima colonna digitare età.
3. Nella seconda riga digitare statociv.
4. Nella terza riga digitare reddito.

Alle nuove variabili viene automaticamente assegnato un tipo di dati numerico.

Se non si specificano i nomi delle variabili, vengono creati automaticamente nomi univoci, che non sono tuttavia descrittivi e non sono comunque consigliabili per i file di dati di grandi dimensioni.

5. Fare clic sulla scheda **Vista dati** per continuare a inserire dati.

I nomi inseriti nella Vista variabile rappresentano le intestazioni delle prime tre colonne della Vista dati.

Iniziare a inserire i dati nella prima riga, a partire dalla prima colonna.

|    | age   | marital | income   | var | var | var |
|----|-------|---------|----------|-----|-----|-----|
| 1  | 55.00 | 1.00    | 72000.00 |     |     |     |
| 2  | 53.00 | .00     | 153000.0 |     |     |     |
| 3  |       |         |          |     |     |     |
| 4  |       |         |          |     |     |     |
| 5  |       |         |          |     |     |     |
| 6  |       |         |          |     |     |     |
| 7  |       |         |          |     |     |     |
| 8  |       |         |          |     |     |     |
| 9  |       |         |          |     |     |     |
| 10 |       |         |          |     |     |     |
| 11 |       |         |          |     |     |     |
| 12 |       |         |          |     |     |     |
| 13 |       |         |          |     |     |     |
| 14 |       |         |          |     |     |     |
| 15 |       |         |          |     |     |     |
| 16 |       |         |          |     |     |     |

Figura 20. Valori inseriti in Vista dati

6. Nella colonna *età* digitare 55.
7. Nella colonna *statociv* digitare 1.
8. Nella colonna *reddito* digitare 72000.
9. Per inserire i dati relativi al soggetto successivo, spostare il cursore sulla seconda riga della prima colonna.
10. Nella colonna *età*, immettere 53.
11. Nella colonna *statociv*, immettere 0.
12. Nella colonna *reddito*, immettere 153000.  
Nelle colonne *età* e *statociv* sono visualizzati i segni decimali, anche se i valori vengono considerati come interi. Per nascondere i segni decimali delle variabili:
13. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.
14. Nella colonna *Decimali* della riga *età* digitare 0 per nascondere il valore decimale.
15. Nella colonna *Decimali* della riga *statociv* digitare 0 per nascondere il valore decimale.

---

## Inserimento di dati di stringa

Nell'Editor dei dati è inoltre possibile inserire dati non numerici, ad esempio stringhe di testo.

1. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.
2. Nella prima cella della prima riga vuota digitare sesso come nome di variabile.
3. Fare clic sulla cella *Tipo* accanto alla voce desiderata.
4. Fare clic sul pulsante a destra della cella *Tipo* per aprire la finestra di dialogo Tipo di variabile.
5. Selezionare **Stringa** per specificare il tipo di variabile.
6. Fare clic su **OK** per salvare la selezione e tornare all'Editor dei dati.



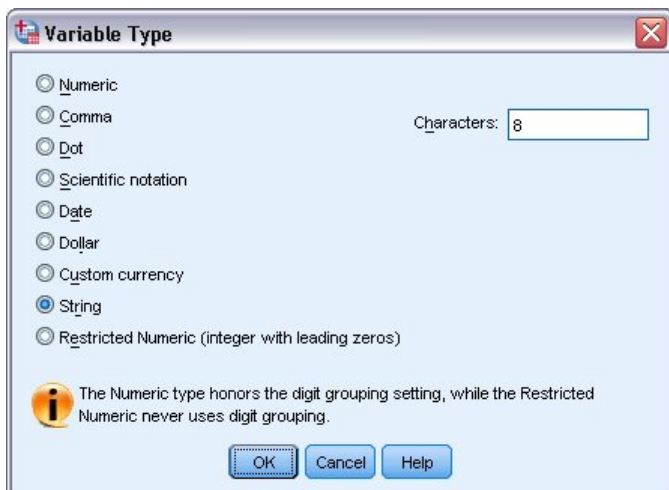


Figura 21. Finestra di dialogo Tipo di variabile

---

## Definizione di dati

Oltre a definire i tipi di dati, è possibile definire anche le etichette valore e variabile descrittive per i nomi delle variabili e i valori dei dati, che vengono utilizzate nei report statistici e nei grafici.

### Aggiunta delle etichette di variabile

Le etichette forniscono una descrizione delle variabili, che spesso è rappresentata da una versione estesa del nome della variabile. La dimensione massima delle etichette è 255 byte. Tali etichette consentono di identificare le diverse variabili nell'output.

1. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.
2. Nella colonna *Etichetta* della riga *età* digitare Età del rispondente.
3. Nella colonna *Etichetta* della riga *statociv* digitare Stato civile.
4. Nella colonna *Etichetta* della riga *reddito* digitare Reddito familiare.
5. Nella colonna *Etichetta* della riga  *Sesso* digitare Sesso.

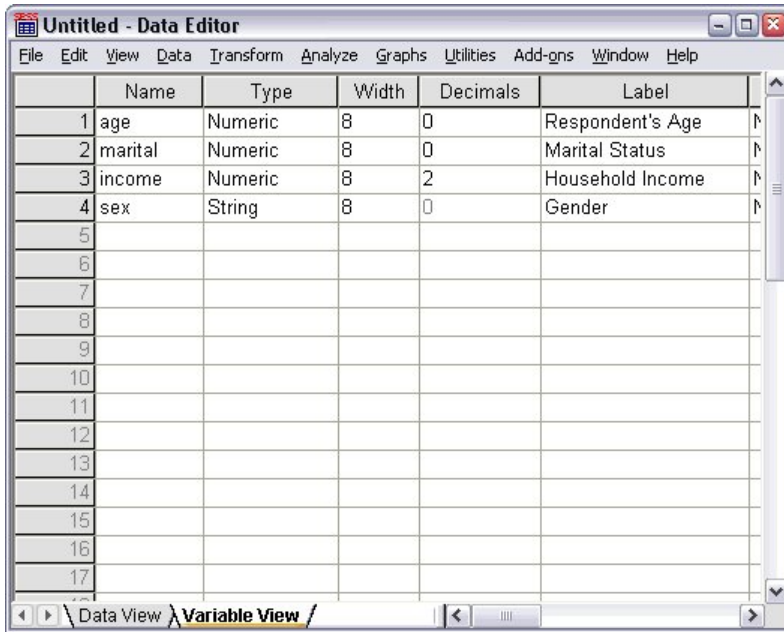


Figura 22. Etichette di variabile inserite in Vista variabile

## Modifica del formato e del tipo di variabile

Nella colonna *Tipo* viene visualizzato il tipo di dati corrente per ogni variabile. I tipi di dati più frequenti sono i dati numerici e le stringhe, ma sono supportati molti altri formati. Nel file di dati corrente, la variabile *reddito* è definita come un tipo numerico.

1. Fare clic sulla cella *Tipo* della riga *reddito*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Tipo di variabile.
2. Selezionare **Dollaro**.

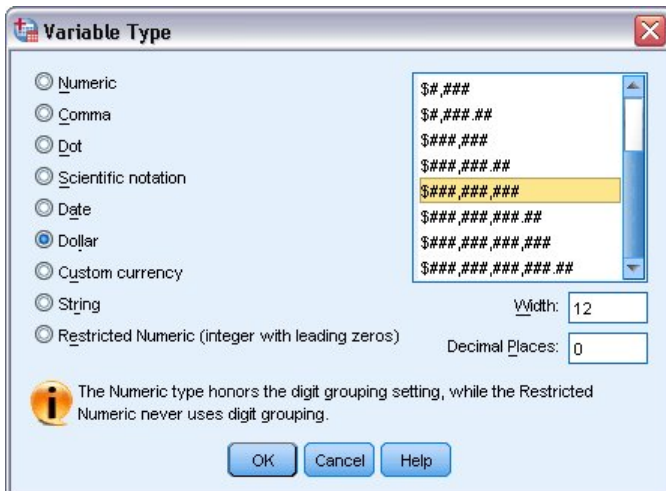


Figura 23. Finestra di dialogo Tipo di variabile

Vengono visualizzate le opzioni di formattazione per il tipo di dati correntemente selezionato.

3. Per il formato della valuta relativo a questo esempio, selezionare **\$###,###,###**.
4. Fare clic su **OK** per salvare le modifiche.

## Aggiunta di etichette valore

Le etichette valore consentono di eseguire il mapping dei valori delle variabili all'etichetta di una stringa. In questo esempio, sono disponibili due valori validi per la variabile *statociv*. Il valore 0 significa che il soggetto è celibe o nubile, mentre il valore 1 significa che è coniugato o coniugata.

1. Fare clic sulla cella *Valori* della riga *statociv*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Etichette valore.

Il **valore** corrisponde al valore numerico effettivo.

L'**etichetta valore** rappresenta l'etichetta di stringa applicata al valore numerico specificato.

2. Digitare 0 nel campo Valore.
3. Digitare Celibe/Nubile nel campo Etichetta.
4. Fare clic su **Aggiungi** per aggiungere l'etichetta all'elenco.

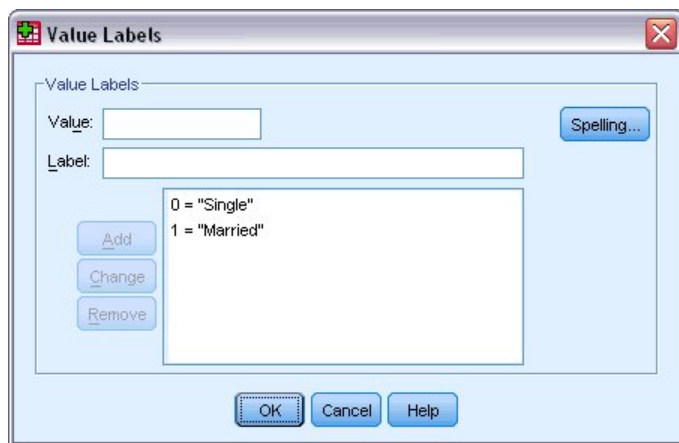


Figura 24. Finestra di dialogo Etichette valore.

5. Digitare 1 nel campo Valore, quindi digitare Coniugato/a nel campo Etichetta.
6. Fare clic su **Aggiungi** e quindi su **OK** per salvare le modifiche e tornare all'Editor dei dati.  
Le etichette possono essere visualizzate anche nella Vista dati, in modo da rendere più leggibili i dati.
7. Fare clic sulla scheda **Vista dati** nella parte inferiore della finestra dell'Editor dei dati.
8. Dai menu, scegliere:

### Visualizza > Etichette valore

Adesso le etichette vengono visualizzate in un elenco quando si inseriscono valori nell'Editor dei dati. Con questa impostazione le opzioni valide hanno una forma più descrittiva.

Se la voce del menu Etichette valore è già attiva (la rispettiva casella di controllo è contrassegnata), selezionando di nuovo **Etichette valore** si *disattiva* la visualizzazione delle etichette valore.

## Gestione dei dati mancanti

I dati mancanti o non validi sono in genere molto frequenti e pertanto non è possibile ignorarli. È possibile che i rispondenti si rifiutino di rispondere ad alcune domande dell'indagine, non sappiano cosa rispondere oppure rispondano in una forma non prevista. Se non si filtrano o si identificano questi dati, i risultati dell'analisi potrebbero rivelarsi inesatti.

Per quanto riguarda i dati numerici, i campi vuoti o i campi che contengono voci non valide vengono convertiti in dati mancanti di sistema, identificati da un punto.

Il motivo per cui un valore risulta mancante può rivelarsi importante ai fini dell'analisi. È possibile, ad esempio, che sia importante distinguere i rispondenti che hanno rifiutato di rispondere a una domanda da quelli che invece non hanno fornito una risposta perché hanno ritenuto che la domanda non li riguardasse.

## Valori mancanti per una variabile numerica

1. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.
2. Fare clic sulla cella *Mancanti* della riga *età*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Valori mancanti.

In questa finestra di dialogo è possibile specificare fino a tre valori mancanti distinti oppure un intervallo di valori e un valore discreto aggiuntivo.



Figura 25. Finestra di dialogo Valori mancanti

3. Selezionare **Valori mancanti discreti**.
4. Digitare 999 nella prima casella di testo e lasciare vuote le altre due.
5. Fare clic su **OK** per confermare le modifiche e tornare all'Editor dei dati.  
Dopo aver aggiunto il valore dei dati mancante, è possibile applicarvi un'etichetta.
6. Fare clic sulla cella *Valori* della riga *età*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Etichette valore.
7. Immettere 999 nel campo Valore.
8. Digitare Nessuna risposta nel campo Etichetta.
9. Fare clic su **Aggiungi** per aggiungere l'etichetta al file di dati.
10. Fare clic su **OK** per confermare le modifiche e tornare all'Editor dei dati.

## Valori mancanti per una variabile stringa

I valori mancanti per le variabili stringa vengono gestiti allo stesso modo di quelli per le variabili numeriche. Tuttavia, a differenza delle variabili numeriche, i campi vuoti nelle variabili stringa non vengono identificati come valori mancanti di sistema, ma vengono interpretati come stringhe vuote.

1. Fare clic sulla scheda **Vista variabile** nella parte inferiore della finestra dell'Editor dei dati.
2. Fare clic sulla cella *Mancanti* della riga  *sesso*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Valori mancanti.

3. Selezionare **Valori mancanti discreti**.
4. Digitare NR nella prima casella di testo.

Per i valori mancanti delle variabili stringa è attiva la distinzione tra maiuscole e minuscole e pertanto il valore *nr* non viene identificato come un valore mancante.

5. Fare clic su **OK** per confermare le modifiche e tornare all'Editor dei dati.

È ora possibile aggiungere un'etichetta per il valore mancante.

6. Fare clic sulla cella *Valori* della riga  *Sesso*, quindi fare clic sul pulsante a destra della cella per aprire la finestra di dialogo Etichette valore.
7. Digitare NR nel campo Valore.
8. Digitare Nessuna risposta nel campo Etichetta.
9. Fare clic su **Aggiungi** per aggiungere l'etichetta al progetto.
10. Fare clic su **OK** per confermare le modifiche e tornare all'Editor dei dati.



---

## Capitolo 4. Esame delle statistiche di riepilogo per singole variabili

Questa sezione tratta semplici misure di riepilogo e in che modo il livello di misurazione di una variabile influenza i tipi di statistiche da utilizzare. Verrà utilizzato il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

---

### Livello di misurazione

In base al livello di misurazione, sono disponibili diverse misure di riepilogo per diversi tipi di dati:

**Categoriale.** Dati con un numero limitato di valori o categorie diversi (ad esempio, stato civile o sesso). Vengono denominati anche **dati qualitativi**. Le variabili categoriali possono essere variabili di dati stringa (alfanumerici) o variabili numeriche che utilizzano un codice numerico per rappresentare le categorie, ad esempio 0 = *Celibe/Nubile* e 1 = *Coniugato/a*. Esistono due tipi fondamentali di dati categoriali:

- **Nominale.** Dati categoriali in cui non esiste un ordine inerente alle categorie. Ad esempio, la categoria lavorativa *vendite* non è di un ordine più alto o più basso rispetto alla categoria lavorativa *marketing* o *ricerca*.
- **Ordinale.** Dati categoriali caratterizzati da un ordine significativo di categorie, ma senza una distanza misurabile tra categorie. Ad esempio, i valori possono essere di ordine *alto*, *medio* o *basso* ma non è possibile calcolare la "distanza" effettiva esistente tra i valori.

**Scala.** Dati misurati in base a una scala di intervallo o di rapporto, in cui i valori dei dati indicano sia l'ordine dei valori sia la distanza tra valori. Ad esempio, uno stipendio di \$72.195 è superiore ad uno stipendio di \$52.398 e la distanza tra i due valori è \$19.797. Indicati anche come **quantitativi** o **dati continui**.

---

### Misure di riepilogo per dati categoriali

Per i dati categoriali, la misura di riepilogo più tipica è il numero o la percentuale di casi presenti in ogni categoria. **Moda** è la categoria con il maggior numero di casi. Per i dati ordinali, la **mediana** (il valore sopra e sotto il quale ricade metà dei casi) potrebbe essere una misura di riepilogo utile in presenza di un elevato numero di categorie.

La procedura Frequenze genera tabelle delle frequenze che visualizzano sia il numero sia la percentuale di casi per ogni valore di variabile osservato.

1. Dai menu, scegliere:

**Analizza > Statistiche descrittive > Frequenze...**

Nota: questa funzione richiede il modulo Statistics Base.

2. Selezionare *Possiede PDA [pda]* e *Possiede TV [tv]* e spostarli nell'elenco Variabili.

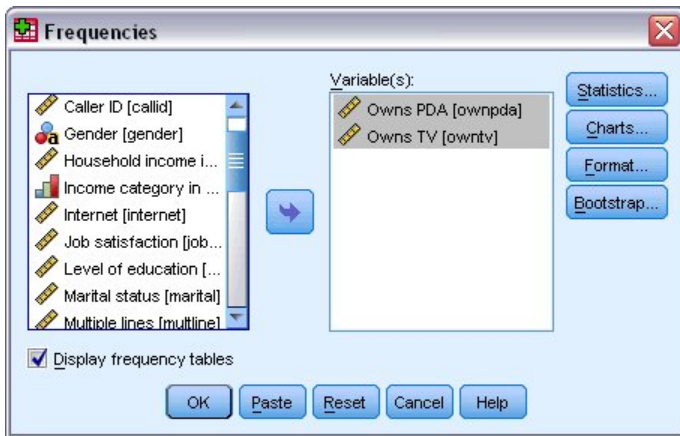


Figura 26. Variabili categoriali selezionate per l'analisi

3. Fare clic su **OK** per eseguire la procedura.

**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         |                    |

Figura 27. Tabelle delle frequenze

Le tabelle delle frequenze vengono visualizzate nel Visualizzatore. Le tabelle delle frequenze rivelano che solo il 20,4% delle persone possiede un PDA ma quasi tutti possiedono una TV (99,0%). Tali dati potrebbero non essere molto rivelatori, ma potrebbe essere interessante scoprire qualcos'altro sul piccolo gruppo di persone che non possiede una TV.

## Grafici per i dati categoriali

È possibile visualizzare le informazioni in una tabella delle frequenze con un grafico a barre o a torta.

1. Verrà di nuovo aperta la finestra di dialogo Frequenze. Ad ogni modo le due variabili dovrebbero essere selezionate.

È possibile utilizzare il pulsante Richiama finestra nella barra degli strumenti per ritornare immediatamente alle procedure utilizzate recentemente.





Figura 28. Pulsante Richiama finestra

2. Fare clic su **Grafici**.
3. Selezionare **Grafici a barre** e quindi fare clic su **Continua**.
4. Fare clic su **OK** nella finestra di dialogo principale per eseguire la procedura.

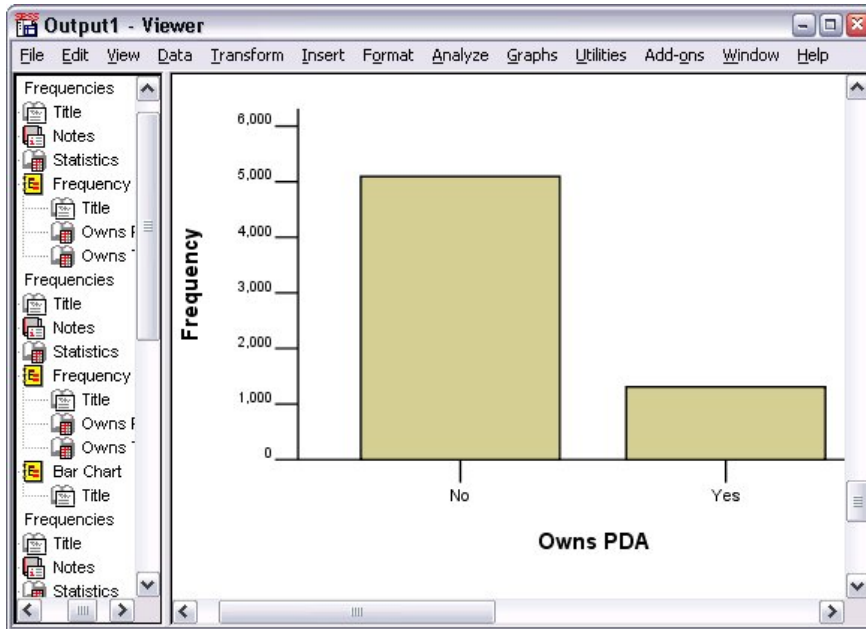


Figura 29. Grafico a barre

Oltre alle tabelle delle frequenze, le stesse informazioni vengono adesso visualizzate in formato di grafici a barre, in modo che risulti semplice constatare che la maggior parte delle persone non possiedono un PDA, ma quasi tutti possiedono una TV.

## Misure di riepilogo per variabili di scala

Sono disponibili molte misure di riepilogo per variabili di scala, comprese le seguenti:

- **Misure di tendenza centrale.** Le misure di tendenza centrale più comuni sono la **media** (media aritmetica) e la **mediana** (valore sopra e sotto il quale ricade la metà dei casi).
- **Misure di dispersione.** Le statistiche che misurano l'entità della variazione o della diffusione dei dati includono la deviazione standard, il valore minimo e massimo.

1. Verrà di nuovo aperta la finestra di dialogo Frequenze.
2. Fare clic su **Ripristina** per annullare qualsiasi impostazione precedente.
3. Selezionare *Reddito familiare in migliaia [reddito]* e spostarlo nell'elenco Variabili.
4. Fare clic su **Statistiche**.
5. Selezionare **Media, Mediana, Deviazione stand., Minimo e Massimo**.
6. Fare clic su **Continua**.

7. Deselezionare **Visualizza tabelle delle frequenze** nella finestra di dialogo principale. In genere le tabelle delle frequenze non sono utili per le variabili di scala, poiché potrebbero esistere tanti valori distinti quanti sono i casi presenti nel file di dati.
8. Fare clic su **OK** per eseguire la procedura.

La tabella Frequenze: Statistiche viene visualizzata nel Visualizzatore.

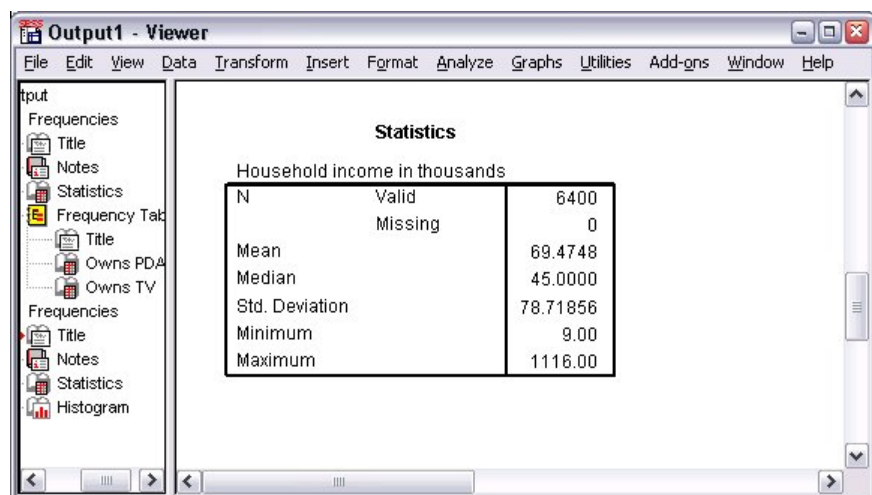


Figura 30. Tabella Frequenze: Statistiche

Nell'esempio vi è una grande differenza tra la media e la mediana. La media è quasi 25.000 volte più grande della mediana e ciò indica che i valori non sono distribuiti normalmente. È possibile verificare la distribuzione mediante un istogramma.

## Istogrammi per variabili di scala

1. Verrà di nuovo aperta la finestra di dialogo Frequenze.
2. Fare clic su **Grafici**.
3. Selezionare **Istogramma** e **Con curva normale**.
4. Fare clic su **Continua**, quindi fare clic su **OK** nella finestra di dialogo principale per eseguire la procedura.

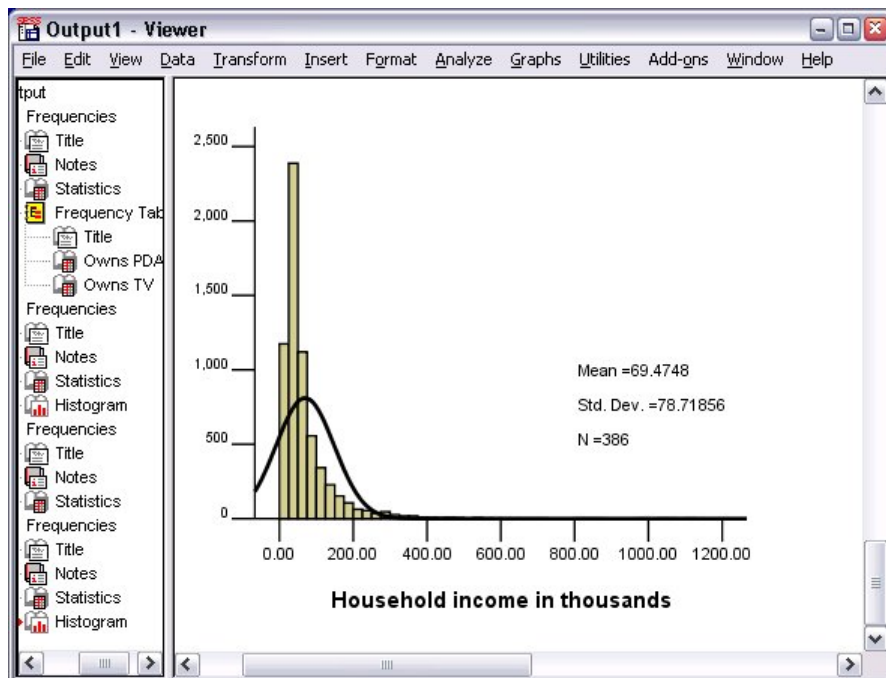


Figura 31. Istogramma

La maggior parte dei casi è raccolta in cluster sull'estremità più bassa della scala, per lo più al di sotto di 100.000. Vi sono, tuttavia, alcuni casi nell'intervallo 500.000 e oltre (troppo pochi per essere persino visibili senza modificare l'istogramma). Tali valori alti per pochi casi soltanto hanno un effetto significativo sulla media, mentre hanno un effetto scarso o nullo sulla mediana, rendendo quest'ultima un indicatore di tendenza centrale migliore nell'esempio riportato.



---

## Capitolo 5. Creazione e modifica di grafici

È possibile creare e modificare un'ampia varietà di tipi di grafici. In questo capitolo vengono creati e modificati grafici a barre. È possibile applicare i principi a qualsiasi tipo di grafico.

---

### Nozioni base per la creazione di grafici

Per illustrare le nozioni base relative alla creazione di grafici, si procederà alla creazione di un grafico a barre relativo al reddito medio per diversi gradi di soddisfazione sul lavoro. In questo esempio viene utilizzato il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

1. Dai menu, scegliere:

**Grafici > Builder di grafico...**

Builder di grafico è una finestra di dialogo interattiva che consente di prevedere l'aspetto futuro di un grafico durante la creazione dello stesso.

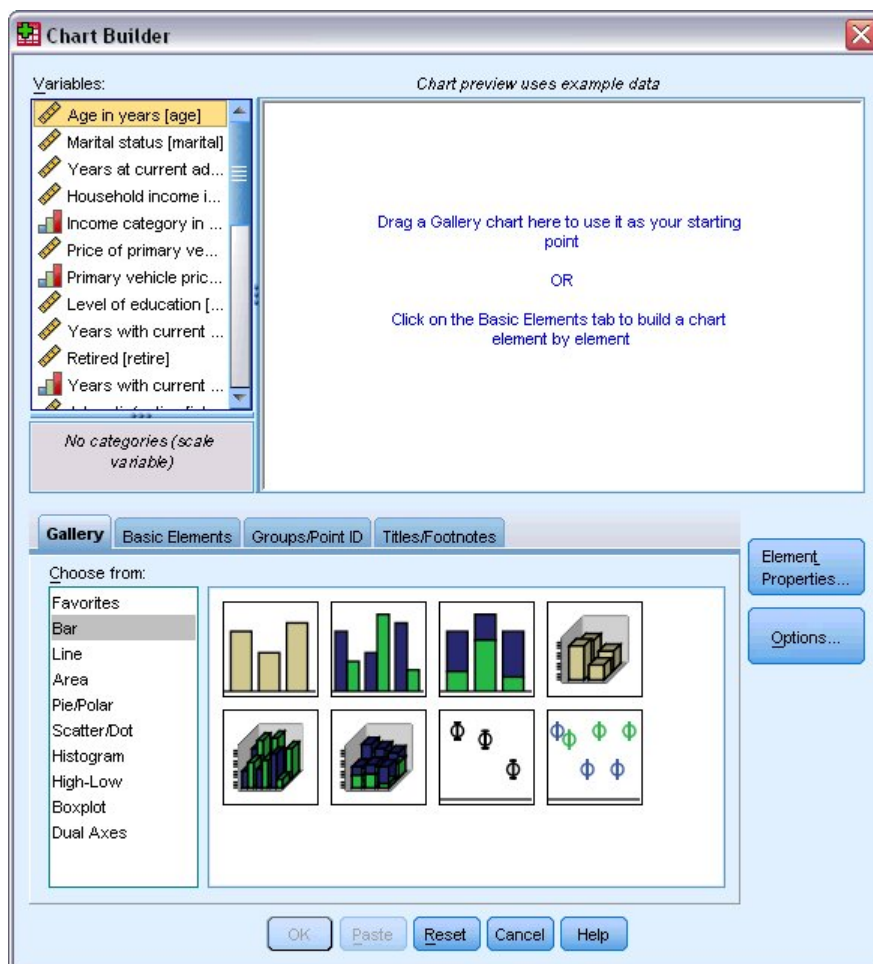


Figura 32. Finestra di dialogo Builder di grafico

## Utilizzo di Modelli del Builder di grafico

1. Fare clic sulla scheda **Modelli** se non è selezionata.

La scheda Modelli contiene vari grafici predefiniti, organizzati per tipo di grafico. Anche la scheda Elementi di base contiene gli elementi di base (quali gli assi e gli elementi del grafico) per la creazione di un grafico vuoto, ma rispetto alla scheda Modelli è più semplice da utilizzare.

2. Fare clic su **A barre** se non è selezionata.

Le icone che rappresentano i grafici a barre disponibili in Modelli sono visualizzate nella finestra di dialogo. Le immagini devono fornire le informazioni sufficienti per identificare il tipo di grafico specifico. Se sono necessarie ulteriori informazioni è possibile visualizzare anche una descrizione Suggerimento del grafico collocando il cursore su un'icona.

3. Trascinare l'icona del grafico a barre semplice nell'area, ovvero nell'ampia area situata sopra i Modelli. Builder di grafico visualizza sull'area un'anteprima del grafico. Si noti che i dati utilizzati per tracciare il grafico non sono i dati effettivi, bensì dei dati esemplificativi.

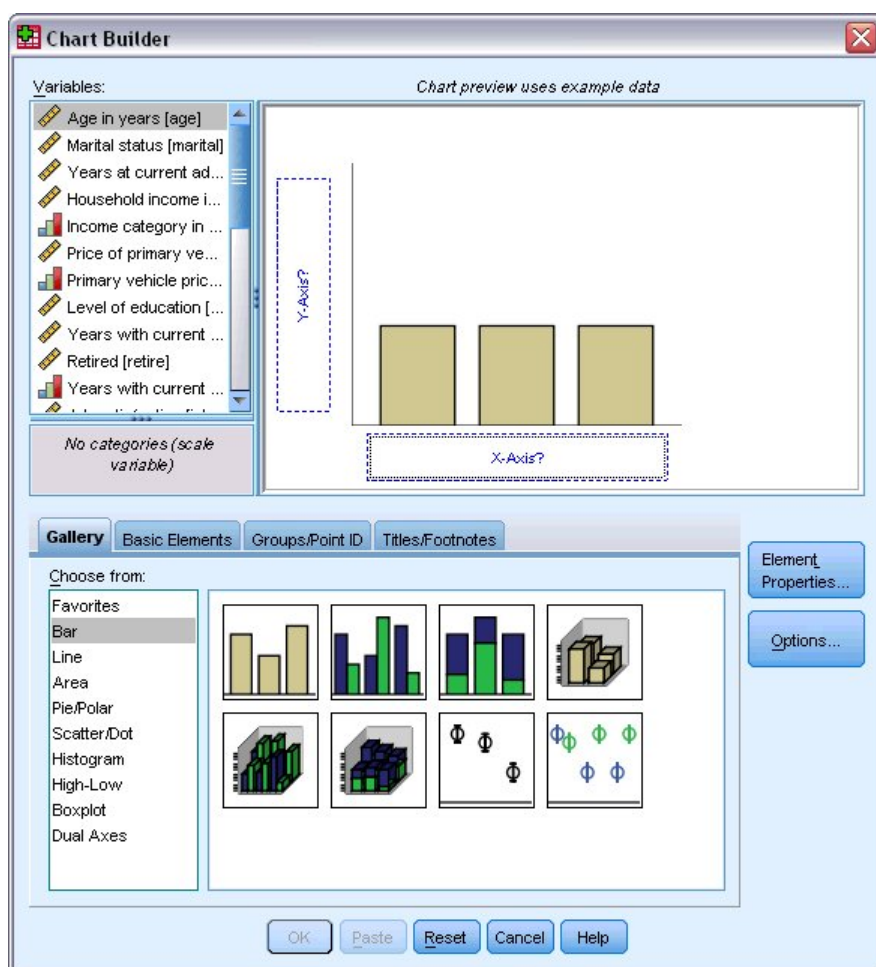


Figura 33. Grafico a barre sull'area di Builder di grafico

## Definizione delle variabili e delle statistiche

Sebbene vi sia un grafico nell'area, esso non è completo perché è privo di variabili o di statistiche, necessarie per controllare l'altezza delle barre e per specificare quale categoria di variabili corrisponde a ogni barra. Non è possibile avere un grafico senza variabili e statistiche. Per aggiungere le variabili, trascinarle dall'elenco Variabili situato sul lato sinistro dell'area.

Nel Builder di grafico è fondamentale il livello di misurazione delle variabili. Se si sta per utilizzare la variabile *Soddisfazione sul lavoro* sull'asse  $x$ , l'icona (che assomiglia a un righello) situata accanto alla variabile, indica che il livello di misurazione della variabile viene definito di scala. Per creare il grafico esatto, è necessario utilizzare un livello di misurazione categoriale. È possibile modificare il livello di misurazione temporaneamente nel Builder di grafico, invece di tornare indietro e modificarlo nella scheda Vista variabile.

1. Fare clic con il pulsante destro del mouse su *Soddisfazione sul lavoro* nell'elenco Variabili e selezionare **Ordinale**. Il livello di misurazione Ordinale è appropriato perché le categorie in *Soddisfazione sul lavoro* possono essere classificate per livello di soddisfazione. Si noti che dopo aver modificato il livello di misurazione, l'icona cambia aspetto.
2. Ora trascinare *Soddisfazione sul lavoro* dall'elenco Variabili nella zona di rilascio dell'asse  $x$ . La zona di rilascio dell'asse  $y$  è il valore predefinito per la statistica *Conteggio*. Se si desidera utilizzare un'altra statistica, quale percentuale o media, è possibile farlo facilmente. Nell'esempio non si utilizzeranno queste statistiche, ma si ripasserà il processo qualora fosse necessario cambiare la statistica in un altro momento.
3. Fare clic su **Proprietà elemento** per visualizzare la finestra corrispondente.

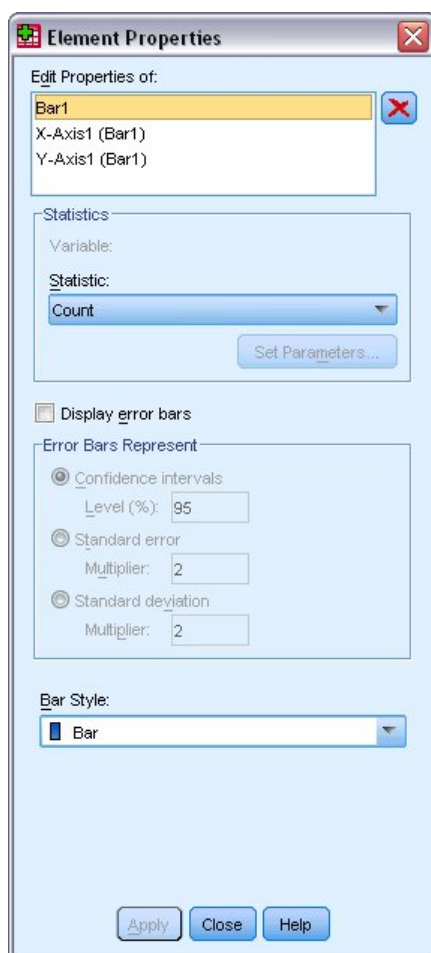


Figura 34. Finestra Proprietà elemento

La finestra Proprietà elemento consente di modificare le proprietà dei vari elementi del grafico. Tali elementi includono gli elementi del grafico (ad esempio le barre del grafico a barre) e gli assi del grafico. Selezionare uno degli elementi nell'elenco 'Modifica proprietà di' per modificare le proprietà associate a quell'elemento. Si noti inoltre la X rossa situata a destra dell'elenco. Questo pulsante

elimina un elemento del grafico dall'area. Poiché **A barre1** è selezionato, le proprietà illustrate vengono applicate agli elementi del grafico, in particolare alle barre.

L'elenco a discesa Statistiche mostra le statistiche specifiche disponibili. Di norma, per ogni tipo di grafico sono disponibili le stesse statistiche. Tenere presente che alcune statistiche richiedono che la zona di rilascio dell'asse  $y$  contenga una variabile.

4. Tornare alla finestra di dialogo Builder di grafico e trascinare *Reddito familiare in migliaia* dall'elenco Variabili alla zona di rilascio dell'asse  $y$ . Poiché la variabile nell'asse  $y$  è scalare e la variabile nell'asse  $x$  è categoriale (ordinale è un tipo di livello di misurazione categoriale), la zona di rilascio dell'asse  $y$  è il valore predefinito per la statistica *Media*. Tali variabili e statistiche sono quelle desiderate, quindi non è necessario modificare le proprietà degli elementi.

## Aggiunta di testo

È inoltre possibile aggiungere titoli e piè di pagina al grafico.

1. Fare clic sulla scheda **Titoli/Piè di pagina**.
2. Selezionare **Titolo 1**.

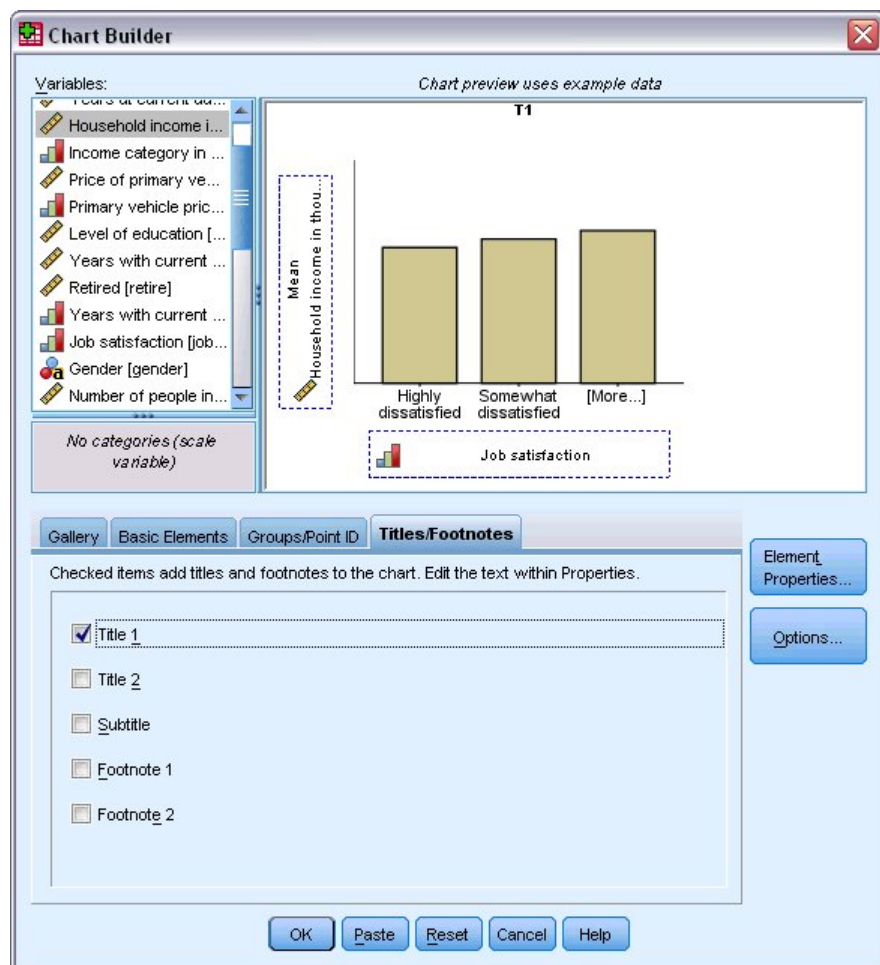


Figura 35. Titolo 1 visualizzato sull'area

Sull'area appare il titolo con l'etichetta **T1**.

3. Nella finestra Proprietà elemento, selezionare **Titolo 1** dall'elenco Modifica proprietà di.
4. Nella casella di Contenuto, digitare Reddito per soddisfazione sul lavoro, che corrisponde al testo del titolo che sarà visualizzato.



5. Fare clic su **Applica** per salvare il testo. Sebbene il testo non sia visualizzato in Builder di grafico, quando si genera il grafico esso apparirà.

## Per creare il grafico

1. Fare clic su **OK** per creare il grafico a barre.

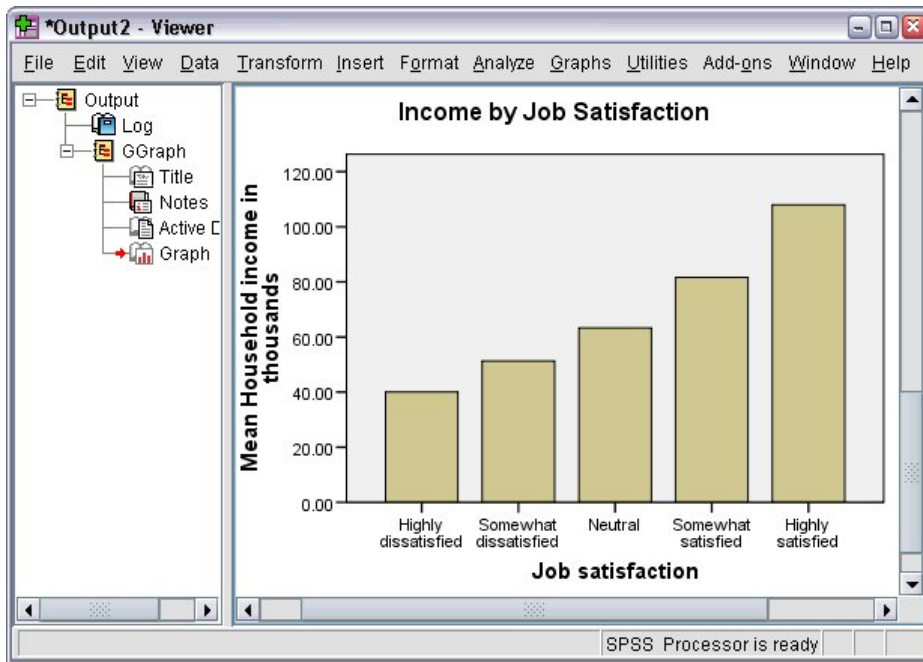


Figura 36. Grafico a barre

Il grafico a barre indica che il reddito dei rispondenti più soddisfatti del proprio lavoro tende a essere più elevato.



## Capitolo 6. Utilizzo dell'output

I risultati dell'esecuzione di una procedura statistica vengono visualizzati nel Visualizzatore. L'output può essere generato sottoforma di tabella, grafico o testo, a seconda delle opzioni scelte per l'esecuzione della procedura. In questo capitolo vengono utilizzati i file *viewertut.spv* e *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

### Utilizzo del Visualizzatore

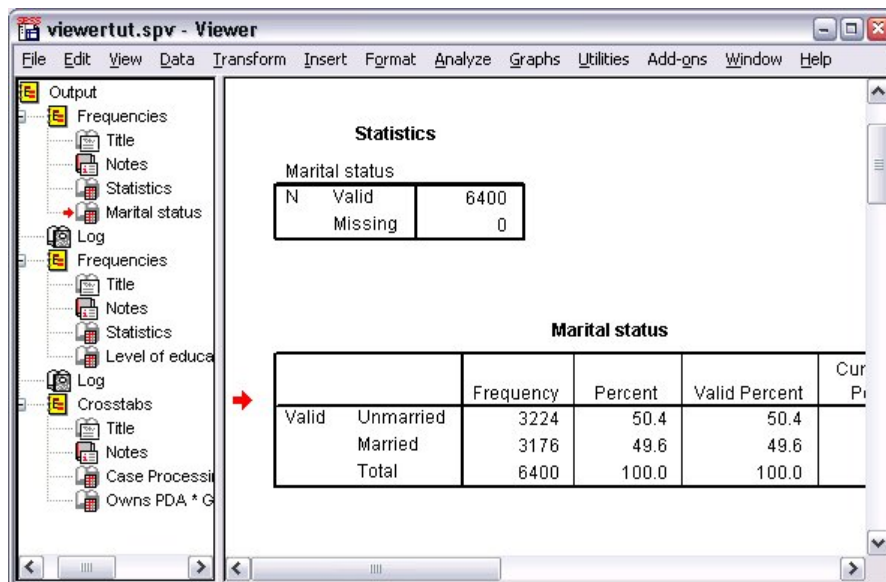


Figura 37. Visualizzatore

La finestra Visualizzatore è suddivisa in due riquadri. Il **riquadro schema**, che contiene tutte le informazioni memorizzate nel Visualizzatore. Il **riquadro di contenuto** contiene tabelle statistiche, grafici e output di testo.

Utilizzare le barre di scorrimento per navigare nel contenuto della finestra, sia in verticale che in orizzontale. Per una navigazione più semplice, fare clic su un elemento nel riquadro schema per visualizzarlo nel riquadro del contenuto.

1. Fare clic e trascinare il bordo destro del riquadro schema per modificarne l'ampiezza.  
L'icona di un libro aperto nel riquadro schema indica che è correntemente visibile nel Visualizzatore, sebbene in quel momento non si trovi nella parte visibile dell'output.
2. Per nascondere una tabella o un grafico, fare doppio clic sull'icona a forma di libro nel riquadro schema.  
L'icona a forma di libro aperto si trasforma in libro chiuso, ad indicare che le informazioni ad esso associate sono nascoste.
3. Per visualizzare nuovamente l'output nascosto, fare doppio clic sull'icona a forma di libro chiuso.  
È possibile, inoltre, nascondere tutto l'output di una procedura statistica specifica oppure tutto l'output nel Visualizzatore.
4. Fare clic sulla casella con il segno meno (-) a sinistra della procedura di cui si desidera nascondere i risultati o fare clic sulla casella accanto al primo elemento in alto nel riquadro schema per nascondere tutto l'output.

Lo schema viene compresso, ad indicare che i risultati vengono nascosti.

È possibile modificare l'ordine in cui viene visualizzato l'output.

5. Nel riquadro schema, fare clic sugli elementi che si desidera spostare.
6. Trascinare gli elementi in una nuova ubicazione nello schema.

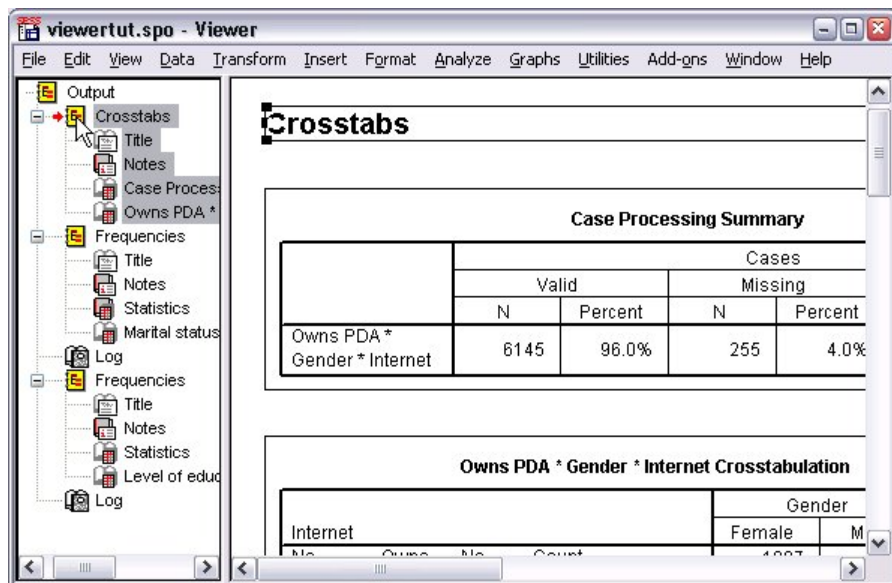


Figura 38. Output riordinato nel Visualizzatore

È anche possibile spostare gli elementi dell'output selezionandoli e trascinandoli nel riquadro dell'output.

---

## Utilizzo dell'Editor di tabelle pivot

I risultati di molte procedure statistiche vengono visualizzati nelle **tabelle pivot**.

### Accesso alle definizioni dell'output

Nell'output vengono visualizzati molti termini statistici. È possibile accedere alle definizioni relative a tali termini nel Visualizzatore.

1. Fare doppio clic sulla tabella *Tavola di contingenza Possessore di PDA \* Sesso \* Internet*.
2. Fare clic con il tasto destro del mouse su *Conteggio previsto* e scegliere **Guida rapida** dal menu a comparsa.

La definizione viene visualizzata in una finestra a comparsa.

The number of cases that would be expected in the cell if the row and column variables are statistically independent or unrelated to one another.

|          |          |                   | Gender         |        | Total  |        |
|----------|----------|-------------------|----------------|--------|--------|--------|
| Internet |          |                   | Female         | Male   |        |        |
| No       | Owns PDA | No                | 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   |        |
|          |          | Expected Count    | 2224.0         | 2285.0 | 4509.0 |        |
| Yes      | Owns PDA | No                | Count          | 513    | 512    | 1025   |
|          |          |                   | Expected Count | 512.5  | 512.5  | 1025.0 |
|          |          | % within Owns PDA | 50.0%          | 50.0%  | 100.0% |        |
|          | Yes      | Count             | 305            | 306    | 611    |        |
|          |          | Expected Count    | 305.5          | 305.5  | 611.0  |        |

Figura 39. Definizione a comparsa

## Table pivot

La presentazione delle informazioni nelle tabelle predefinite prodotte potrebbe non essere sufficientemente chiara o ordinata. Con le tabelle pivot è possibile eseguire la trasposizione di righe e colonne ("flip" della tabella), regolare l'ordine dei dati in una tabella e modificare la tabella in molti altri modi. Ad esempio, è possibile modificare una tabella corta e larga in una lunga e sottile eseguendo la trasposizione delle righe e delle colonne. La modifica del layout della tabella non influisce sui risultati. Si tratta semplicemente di un meccanismo per visualizzare le informazioni in un modo diverso, più consona alle proprie aspettative.

1. Se non è già attivata, attivare con un doppio clic la tabella *Tavola di contingenza Possessore di PDA \* Sesso \* Internet*.
2. Se la finestra Barre pivot non è visibile, dai menu selezionare:

### Pivot > Barre pivot

La funzionalità Barre pivot consente di spostare i dati tra colonne, righe e strati.

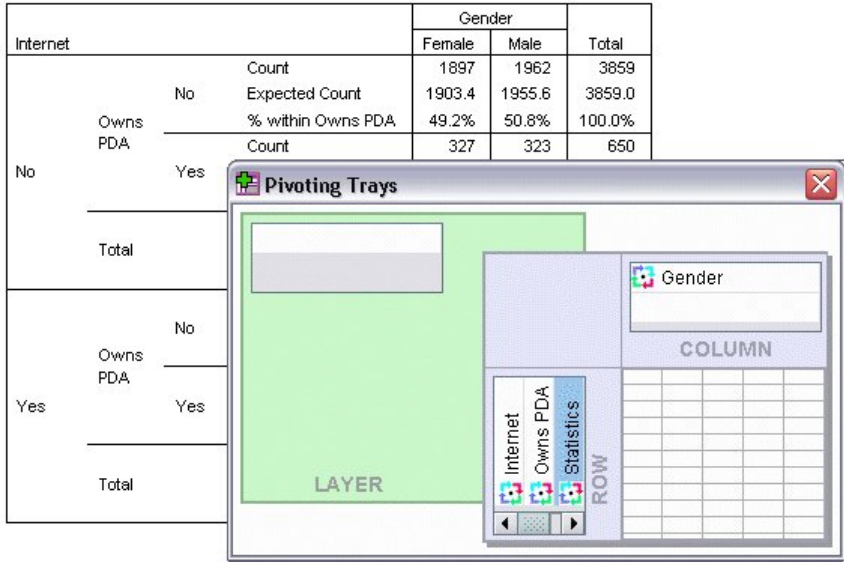


Figura 40. Barra pivot

- Trascinare l'elemento *Statistiche* dalla dimensione Riga alla dimensione di colonna, sotto *Sesso*. La tabella viene immediatamente riconfigurata per riflettere le modifiche apportate. L'ordine degli elementi nella barra pivot riflette l'ordine degli elementi nella tabella.
- Trascinare e rilasciare l'elemento *Possessore di PDA* prima dell'elemento *Internet* nella dimensione Riga per invertire l'ordine di queste due righe.

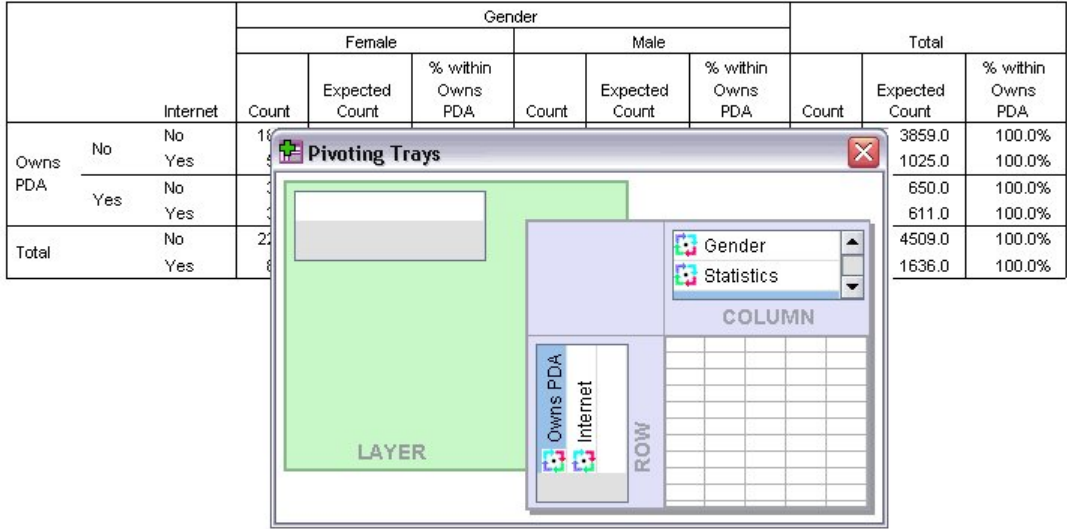


Figura 41. Scambio di righe

### Creazione e visualizzazione di strati

Gli strati possono essere utili per le tabelle di ampie dimensioni con categorie di informazioni nidificate. La creazione di strati semplifica l'aspetto della tabella facilitandone la lettura.

- Trascinare l'elemento *Sesso* dalla dimensione Colonna alla dimensione Strato.

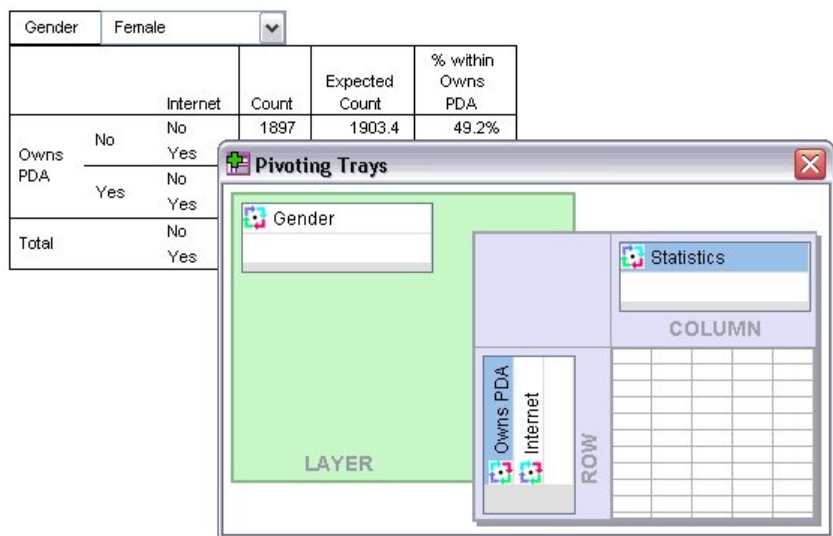


Figura 42. L'icona pivot Sesso nella dimensione Strato

Per visualizzare uno strato diverso, selezionare una categoria dall'elenco a discesa nella tabella.

## Modifica di tabelle

A meno che l'utente non abbia creato un TableLook personalizzato, con la formattazione standard vengono create tabella pivot. È possibile modificare la formattazione di qualsiasi testo all'interno di una tabella. I formati che possono essere modificati includono il nome del carattere, la dimensione del carattere, lo stile del carattere (grassetto o corsivo) e il colore.

1. Fare doppio clic sulla tabella *Livello di istruzione*.
2. Se la barra degli strumenti di formattazione non è visibile, dai menu selezionare: **Visualizza > Barra degli strumenti**
3. Fare clic sul testo di un titolo, *Livello di istruzione*.
4. Dall'elenco a discesa della dimensione del carattere sulla barra degli strumenti selezionare **12**.
5. Per cambiare il colore del testo del titolo, fare clic sullo strumento Colore testo e selezionare un nuovo colore.

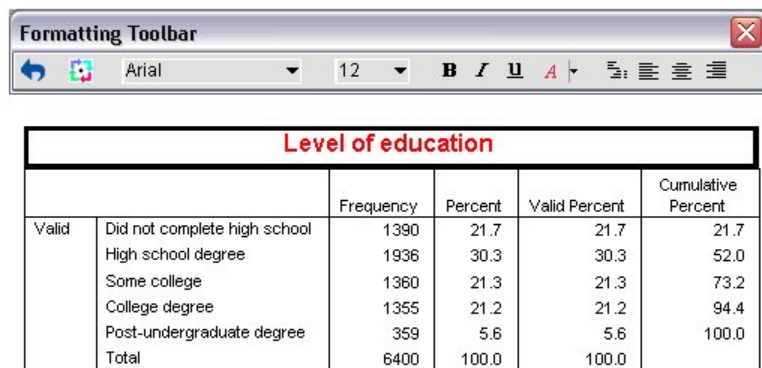


Figura 43. Testo del titolo riformattato nella tabella pivot

È anche possibile modificare il contenuto di tabelle e livelli. Ad esempio, è possibile modificare il titolo di questa tabella.

6. Fare doppio clic sul titolo.

7. Digitare Livello di istruzione per la nuova etichetta.

*Nota:* se si modificano i valori in una tabella, i totali e altre statistiche non vengono ricalcolati.

## Nascondere righe e colonne

Alcuni dei dati visualizzati in una tabella potrebbero essere superflui o addirittura rendere inutilmente complessa la tabella. Fortunatamente, è possibile nascondere righe e colonne intere senza perdere dati.

1. Se non è già attivata, attivare con un doppio clic la tabella *Livello di istruzione*.
2. Fare clic sull'etichetta di colonna *Percentuale valida* per selezionarla.
3. Dal menu Modifica o dal menu a comparsa visualizzato facendo clic con il tasto destro del mouse scegliere:

**Seleziona > Celle dati e celle etichette**

4. Dal menu Visualizza scegliere **Nascondi** o dal menu a comparsa visualizzato facendo clic con il tasto destro del mouse scegliere **Nascondi categoria**.

La colonna viene nascosta, ma non eliminata.

**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   |                    |

Figura 44. Colonna Percentuale valida nascosta in una tabella

Per visualizzare nuovamente la colonna:

5. Dai menu, scegliere:

**Visualizza > Mostra tutto**

È possibile nascondere e visualizzare le righe allo stesso modo delle colonne.

## Modifica dei formati di vista dati

È possibile modificare con facilità il formato di vista dati nelle tabelle pivot.

1. Se non è già attivata, attivare con un doppio clic la tabella *Livello di istruzione*.
2. Fare clic sull'etichetta di colonna *Percentuale* per selezionarla.
3. Dal menu Modifica o dal menu a comparsa visualizzato facendo clic con il tasto destro del mouse scegliere:

**Seleziona > Celle dati**

4. Dal menu Formato o dal menu a comparsa visualizzato facendo clic con il tasto destro del mouse scegliere **Proprietà cella**.
5. Fare clic sulla scheda **Formatta valore**.
6. Digitare 0 nel campo Decimali per nascondere tutti i segni decimali nella colonna.



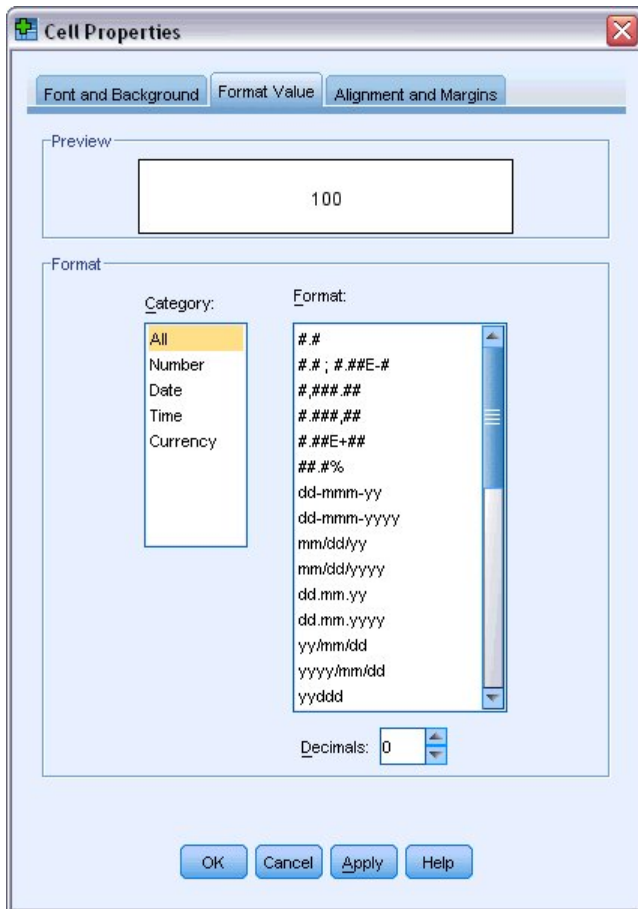


Figura 45. Proprietà cella: scheda Formatta valore

È anche possibile modificare il tipo di dati e il formato in questa finestra di dialogo.

7. Selezionare il tipo desiderato dall'elenco Categoria, quindi dall'elenco Formata selezionare il relativo formato.
8. Fare clic su **OK** o su **Applica** per applicare le modifiche.

### Education Level

|       |                              | Frequency | Percent | Cumulative<br>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     |                       |

Figura 46. Decimali nascosti nella colonna Percentuale

Nella colonna *Percentuale* i decimali sono nascosti.

## Tablelook

Il formato delle tabelle è una parte critica nel fornire risultati chiari, concisi e significativi. Se la tabella risulta difficile da leggere, le informazioni contenute in quella tabella potrebbero non essere comprese facilmente.

## Utilizzo di formati predefiniti

1. Fare doppio clic sulla tabella *Stato civile*.
2. Dai menu, scegliere:

**Formato > TableLook...**

Nella finestra di dialogo TableLook sono elencati vari stili predefiniti. Selezionare uno stile dall'elenco per visualizzarlo in anteprima nella finestra di esempio a destra.

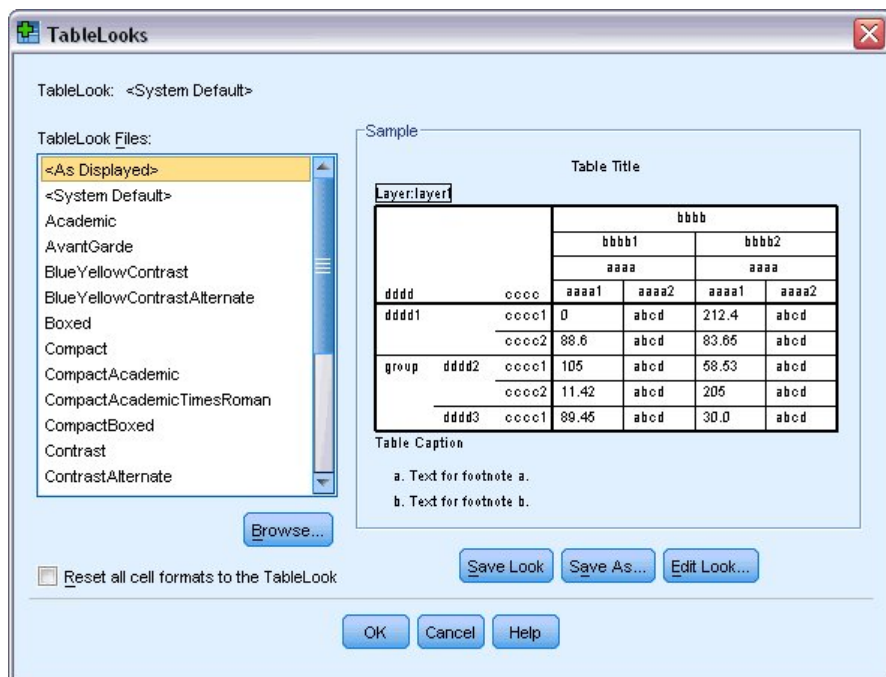


Figura 47. Finestra di dialogo Tablelook

È possibile utilizzare uno stile così com'è oppure modificarlo per adattarlo alle esigenze dell'utente.

3. Per utilizzare uno stile esistente, selezionarlo e fare clic su **OK**.

## Personalizzazione di stili TableLook

È possibile personalizzare un formato per adattarlo alle proprie specifiche esigenze. Quasi tutti gli aspetti di una tabella, dal colore dello sfondo agli stili dei bordi, possono essere personalizzati.

1. Fare doppio clic sulla tabella *Stato civile*.
2. Dai menu, scegliere:  
**Formato > TableLook...**
3. Selezionare lo stile più simile al formato desiderato e fare clic su **Modifica aspetto**.
4. Fare clic sulla scheda **Formati di cella** per visualizzare le opzioni di formattazione.

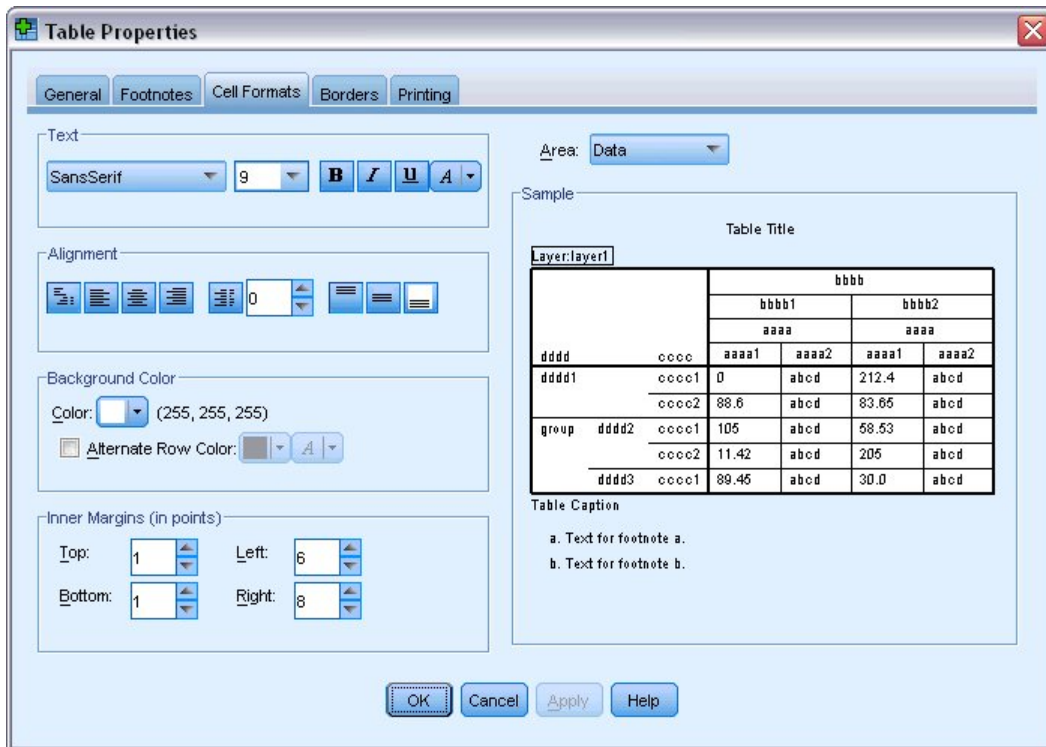


Figura 48. Finestra di dialogo Proprietà tabella

Le opzioni di formattazione riguardano il nome e la dimensione del carattere, lo stile e il colore. Sono disponibili anche ulteriori opzioni per l'allineamento, i colori di primo piano e di sfondo e le dimensioni dei margini.

La finestra di esempio a destra fornisce un'anteprima di come le modifiche di formattazione influiscono sulla tabella. Ogni area della tabella può avere stili di formattazione diversi. In tal modo, titolo e dati possono avere stili diversi. Per selezionare un'area della tabella da modificare, scegliere l'area in base al nome nell'elenco a discesa Area oppure è possibile fare clic sull'area che si desidera modificare nella finestra Campione.

5. Selezionare **Dati** dall'elenco a discesa Area.
6. Selezionare un nuovo colore dalla palette Sfondo.
7. Quindi, selezionare un nuovo colore del testo.  
Il nuovo stile è visualizzato nella finestra di esempio.

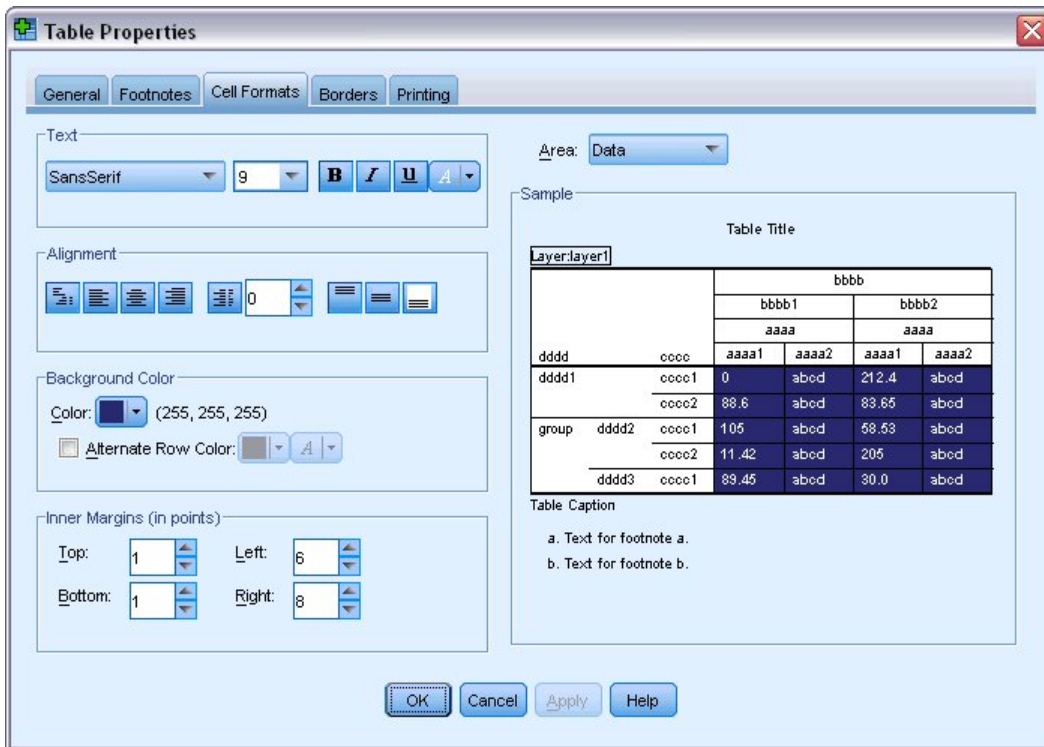


Figura 49. Modifica dei formati della cella di tabella

8. Fare clic su **OK** per chiudere la finestra di dialogo TableLook.  
È possibile salvare il nuovo stile per applicarlo alle successive tabelle.
9. Fare clic su **Salva come**.
10. Navigare alla cartella di destinazione ed immettere un nome per il nuovo stile nella casella di testo Nome file.
11. Scegliere **Salva**.
12. Fare clic su **OK** per applicare le modifiche e ritornare al Visualizzatore.

Adesso la tabella contiene la formattazione personalizzata specificata.

The screenshot shows a PivotTable titled "Marital status" within the SPSS "viewertut.spv" window. The table has four columns: Frequency, Percent, Valid Percent, and Cumulative Percent. The rows represent the marital status categories: Unmarried, Married, and Total. The data is as follows:

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

Figura 50. Personalizzare il TableLook

## Modifica dei formati di tabella predefiniti

Sebbene sia possibile modificare il formato delle tabelle già create, potrebbe essere consigliabile modificare il TableLook predefinito, in modo da non dover intervenire sul formato ogni volta che si crea una tabella.

Per modificare lo stile TableLook predefinito per le tabelle pivot, scegliere:

### Modifica > Opzioni...

1. Fare clic sulla scheda **Tabelle pivot** nella finestra di dialogo Opzioni.

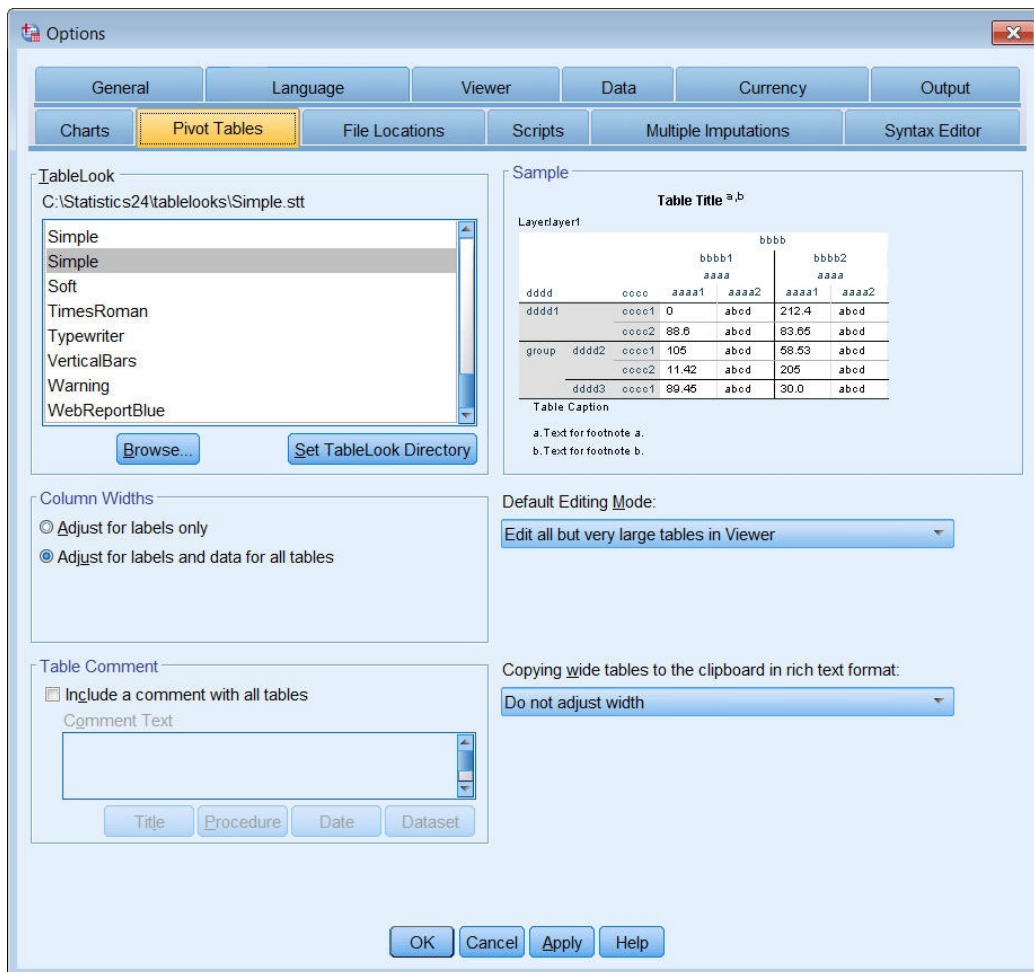


Figura 51. Finestra di dialogo Opzioni

2. Selezionare lo stile TableLook che si desidera utilizzare per tutte le nuove tabelle.  
A destra della finestra di esempio è visualizzata un'anteprima di ogni TableLook.
3. Fare clic su **OK** per salvare le impostazioni e chiudere la finestra di dialogo.

Tutte le tabelle create dopo aver modificato il TableLook predefinito saranno conformi, automaticamente, alle nuove regole di formattazione.

## Personalizzazione delle impostazioni iniziali di visualizzazione

Le impostazioni iniziali di visualizzazione includono l'allineamento degli oggetti nel Visualizzatore, che siano visibili o non visibili per impostazione predefinita, e l'ampiezza della finestra del Visualizzatore. Per modificare tali impostazioni:

1. Dai menu, scegliere:  
**Modifica > Opzioni...**
2. Fare clic sulla scheda **Visualizzatore**.

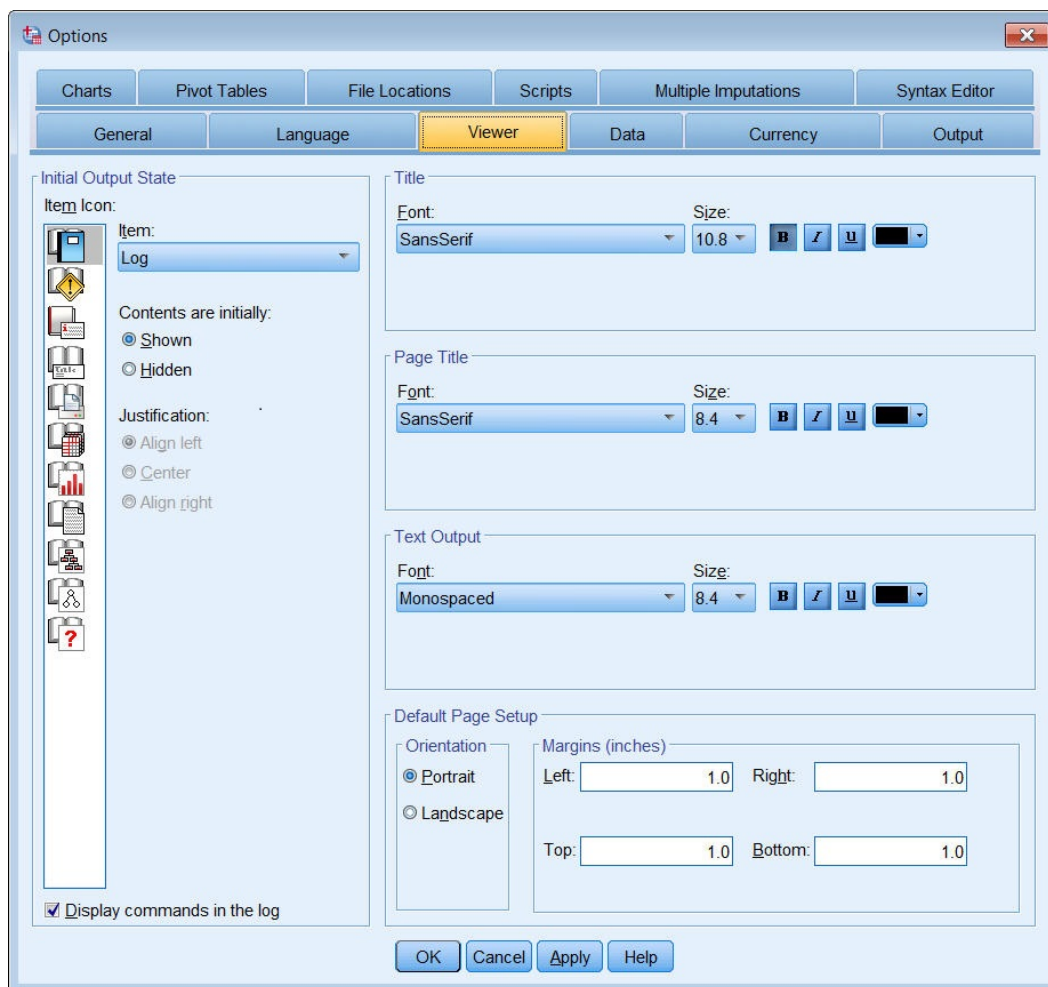


Figura 52. Opzioni del visualizzatore

Le impostazioni vengono applicate oggetto per oggetto. Ad esempio, è possibile personalizzare la modalità di visualizzazione dei grafici senza apportare modifiche alla modalità di visualizzazione delle tabelle. Selezionare semplicemente l'oggetto che si desidera personalizzare ed apportare le modifiche.

3. Fare clic sull'icona **Titolo** per visualizzarne le impostazioni.
4. Fare clic su **Centra** per visualizzare tutti i titoli al centro (orizzontale) del Visualizzatore.  
È possibile inoltre nascondere gli elementi, ad esempio i messaggi di avviso e di log, che tendono ad ingombrare l'output. Facendo doppio clic su un'icona viene automaticamente modificata la proprietà di visualizzazione dell'oggetto.
5. Fare doppio clic sull'icona **Avvisi** per nascondere i messaggi di avviso nell'output.
6. Fare clic su **OK** per salvare le modifiche e chiudere la finestra di dialogo.

## Visualizzazione delle etichette valore e variabili

Nella maggior parte dei casi, la visualizzazione delle etichette per le variabili e i valori è più efficace della visualizzazione del nome della variabile o del valore effettivo dei dati. Potrebbero esserci casi, tuttavia, in cui si desidera visualizzare sia i nomi che le etichette.

1. Dai menu, scegliere:  
**Modifica > Opzioni...**
2. Selezionare la scheda **Etichette di output**.

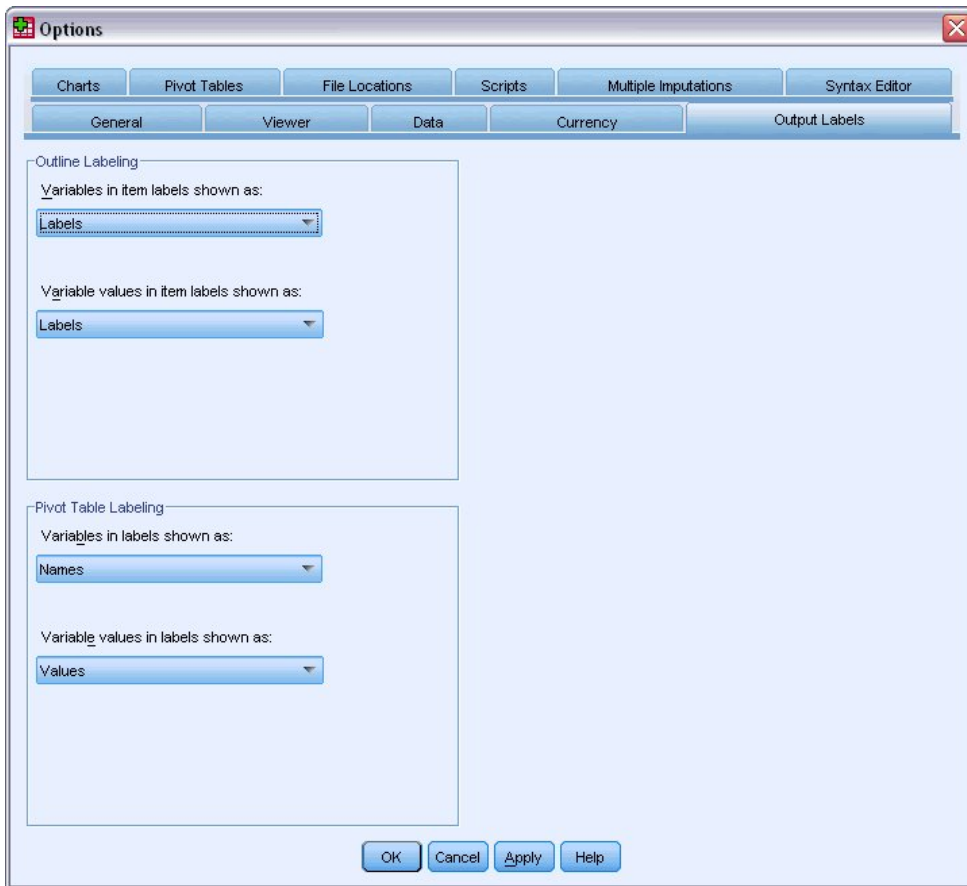


Figura 53. Impostazioni di Etichette delle tabelle pivot

È possibile specificare impostazioni diverse per i riquadri schema e contenuto. Per visualizzare, ad esempio, le etichette nello schema e i nomi delle variabili e i valori dei dati nel contenuto:

3. Nel gruppo Etichette delle tabelle pivot selezionare **Nomi** dall'elenco a discesa Variabili nelle etichette di riga o colonna per visualizzare i nomi delle variabili anziché le etichette.
4. Quindi selezionare **Valori** dall'elenco a discesa Valori nelle etichette di riga o colonna per visualizzare i valori dei dati anziché le etichette.

Le tabelle seguenti generate nella sessione rispecchieranno tali modifiche.

**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         |                    |

Figura 54. Nomi e valori variabile visualizzati

## Utilizzo dell'output in altre applicazioni

I risultati possono essere utilizzati in molte applicazioni. Ad esempio, è possibile includere una tabella o un grafico in una presentazione o in un report.

Gli esempi seguenti sono specifici di Microsoft Word ma funzionano anche in altre applicazioni di elaborazione testi.



## Incollare i risultati come tabelle Word

È possibile incollare tabelle pivot in Word come se fossero tabelle Word native. Tutti gli attributi della tabella, quali le dimensioni e il colore dei caratteri vengono preservati. Dato che la tabella viene incollata nel formato tabelle di Word, è possibile modificarla in Word proprio come qualsiasi altra tabella.

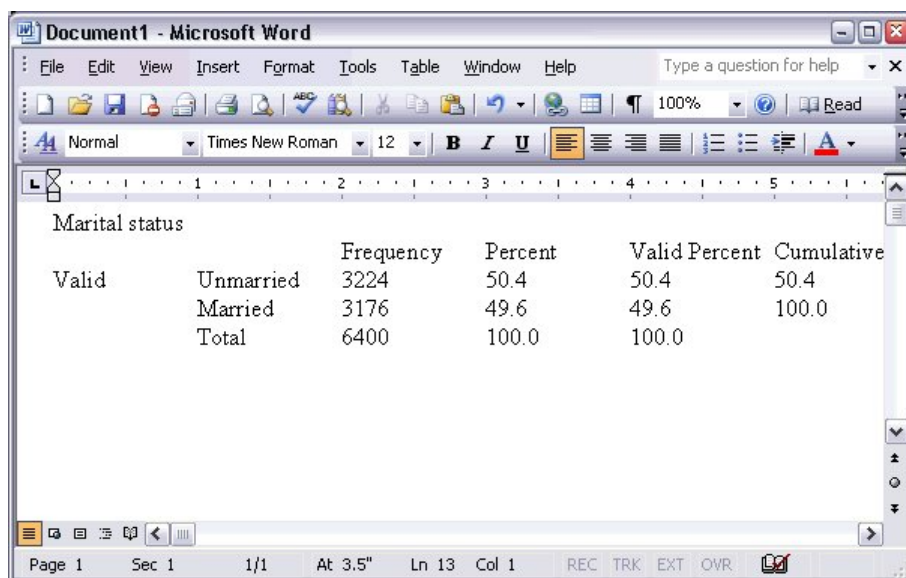
1. Fare clic su una tabella nel Visualizzatore per selezionarla.
2. Dai menu, scegliere:  
**Modifica > Copia**
3. Aprire l'applicazione di elaborazione testi.
4. Dai menu dell'applicazione scegliere:  
**Modifica > Incolla speciale...**
5. Selezionare **Testo formattato (RTF)** nella finestra di dialogo Incolla speciale.
6. Fare clic su **OK** per incollare i risultati nel documento corrente.

La tabella viene ora visualizzata nel documento. È possibile applicare la formattazione personalizzata, modificare i dati e ridimensionare la tabella per adeguarla alle necessità dell'utente.

## Incollare i risultati come testo

È possibile copiare le tabelle pivot ad altre applicazioni come testo normale. Gli stili di formattazione non vengono preservati con questo metodo, ma è possibile modificare i dati della tabella dopo averli incollati nell'applicazione di destinazione.

1. Fare clic su una tabella nel Visualizzatore per selezionarla.
2. Dai menu, scegliere:  
**Modifica > Copia**
3. Aprire l'applicazione di elaborazione testi.
4. Dai menu dell'applicazione scegliere:  
**Modifica > Incolla speciale...**
5. Selezionare **Testo non formattato** nella finestra di dialogo Incolla speciale.
6. Fare clic su **OK** per incollare i risultati nel documento corrente.



The screenshot shows a Microsoft Word window titled "Document1 - Microsoft Word". The ribbon is set to "Table" and the "Table" menu is open. The pivot table is displayed 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         |            |

Figura 55. Tabella pivot visualizzata in Word

Ogni colonna della tabella viene separata da tabulazioni. È possibile modificare la larghezza delle colonne adattando le tabulazioni nell'applicazione di elaborazione testi.

## Esportazione dei risultati in file di Microsoft Word, PowerPoint ed Excel

È possibile esportare i risultati a un file Excel, Microsoft Word o PowerPoint. È possibile esportare elementi selezionati o tutti gli elementi nel Visualizzatore. In questa sezione vengono utilizzati i file *msouttut.spv* e *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

*Nota:* l'esportazione in PowerPoint è disponibile solo su sistemi operativi Windows e non è disponibile con la versione Student.

Nel riquadro schema del Visualizzatore è possibile selezionare elementi specifici che si desidera esportare oppure esportare tutti gli elementi o tutti gli elementi visibili.

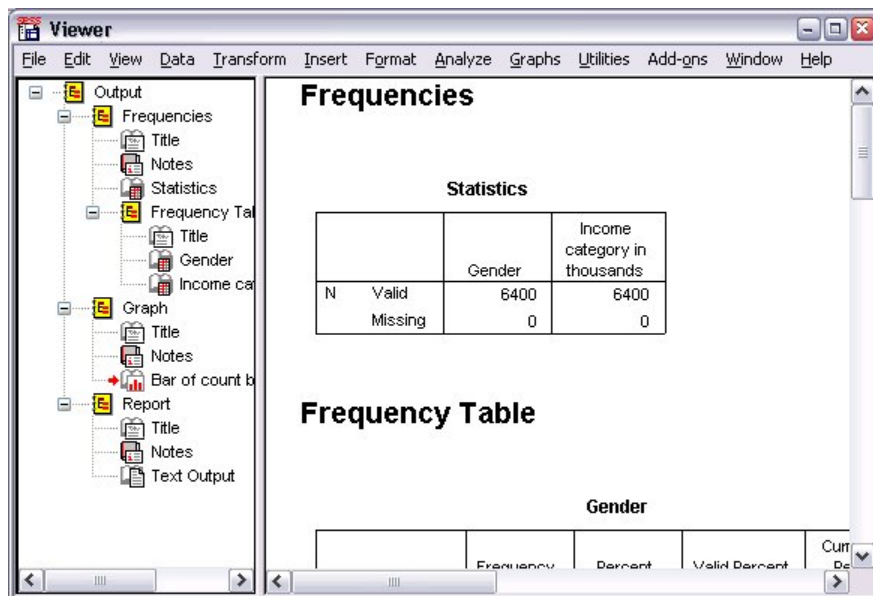


Figura 56. Visualizzatore

1. Dai menu del Visualizzatore scegliere:

**File > Esporta...**

Anziché esportare tutti gli oggetti nel Visualizzatore, è possibile scegliere di esportare solo gli oggetti visibili (libri aperti nel riquadro schema) oppure gli oggetti selezionati nel riquadro schema. Se non sono stati selezionati oggetti nel riquadro schema, l'opzione che consente di esportare gli oggetti selezionati non sarà disponibile.

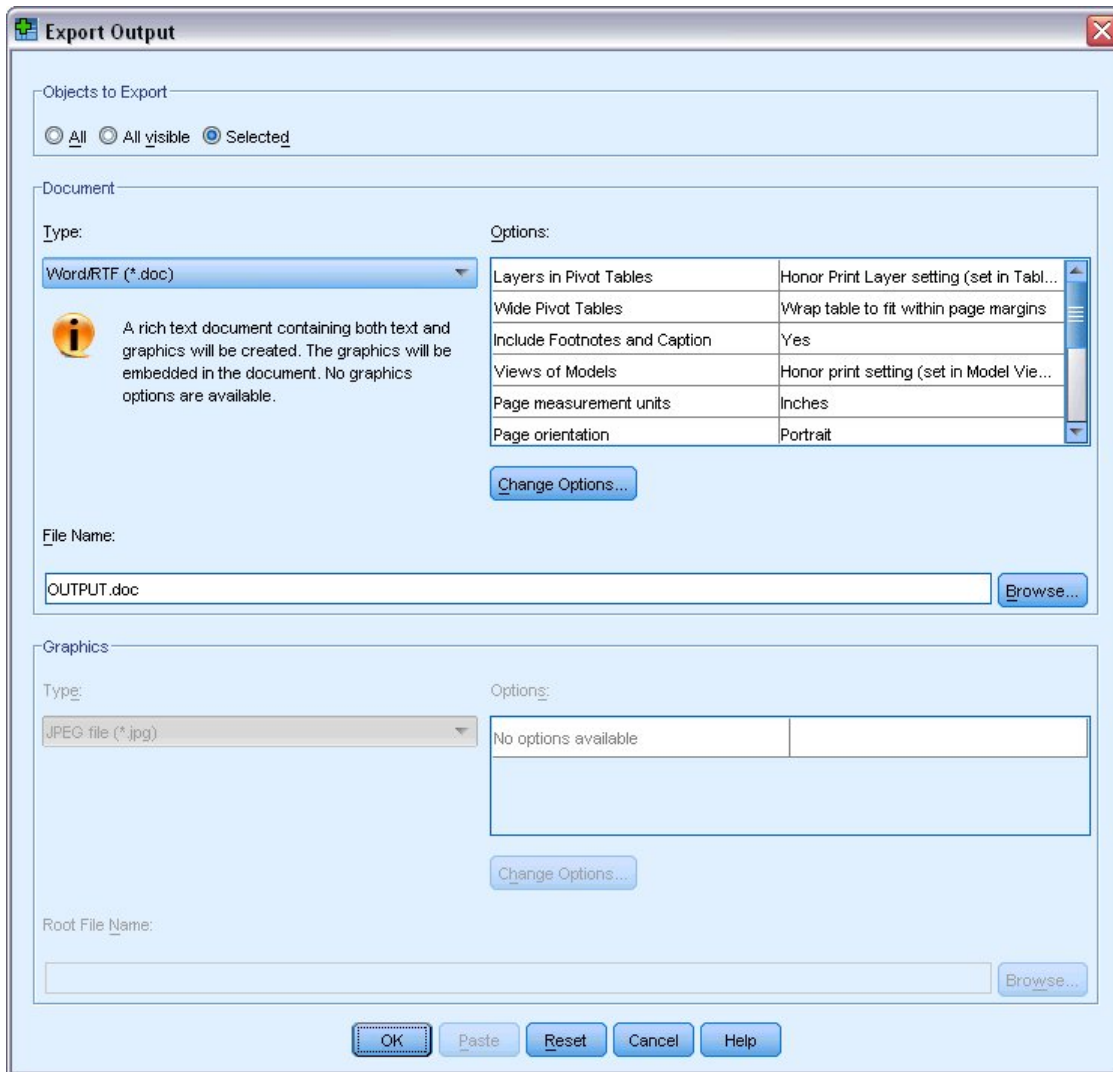


Figura 57. Finestra di dialogo Esporta output

2. Nel gruppo Oggetti da esportare, selezionare **Tutto**.
3. Dall'elenco a discesa Tipo file, selezionare **File Word/RTF (\*.doc)**.
4. Fare clic su **OK** per generare il file di Word.

Quando viene aperto il file risultante con Word, è possibile vedere come sono stati esportati i risultati. Le note, che sono oggetti non visibili, vengono visualizzate in Word, perché è stato scelto di esportare tutti gli oggetti.

Le tabelle pivot diventano tabelle di Word e preservano la formattazione della tabella pivot originale, compresi i caratteri, i colori, i bordi e così via.

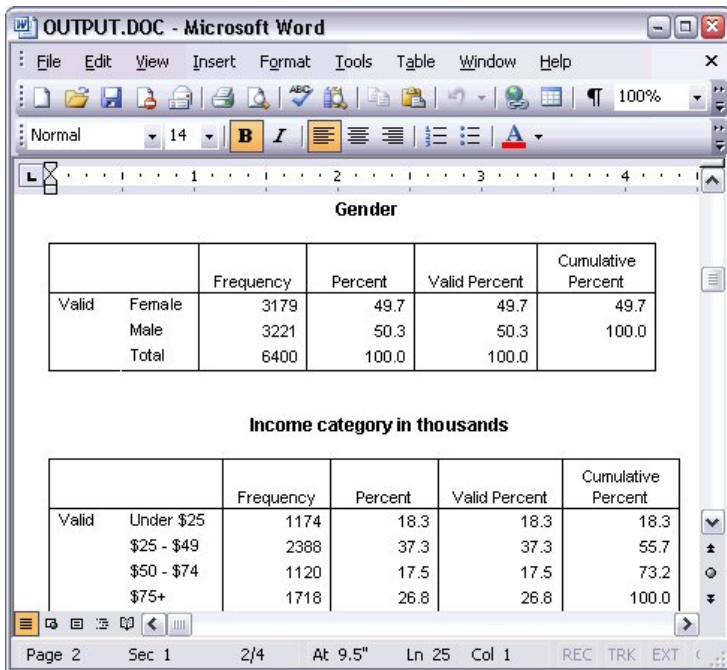


Figura 58. Tabelle pivot in Word

I grafici vengono incorporati nel documento di Word come immagini grafiche.

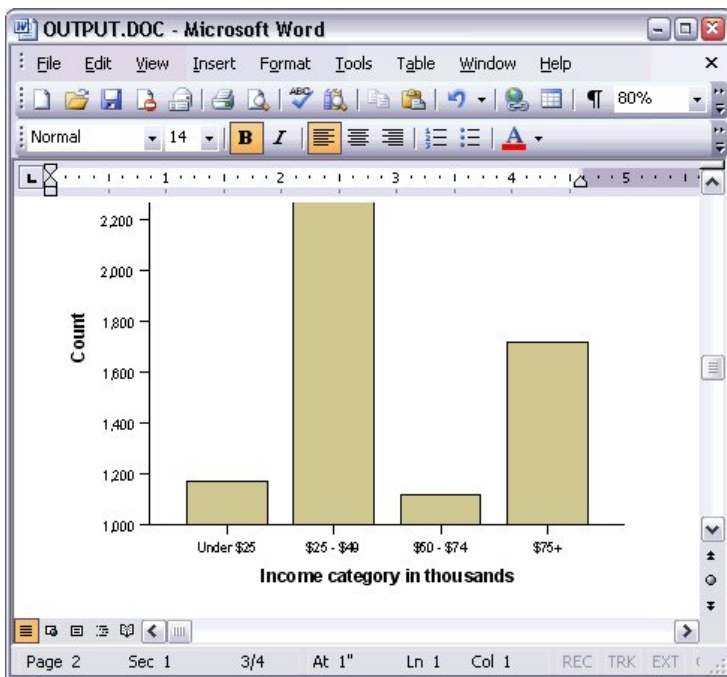


Figura 59. Grafici in Word

L'output testuale viene visualizzato con lo stesso carattere utilizzato per l'oggetto di testo nel Visualizzatore. Per il corretto allineamento, per l'output testuale è necessario utilizzare un carattere a spaziatura fissa (proporzionale).

Page 1

| Income category in thousands | Age in years Mean | Level of education Mean |
|------------------------------|-------------------|-------------------------|
| Under \$25                   | 38                | 2                       |
| \$25 - \$49                  | 39                | 3                       |
| \$50 - \$74                  | 43                | 3                       |
| \$75+                        | 49                | 3                       |
| Grand Total                  | 42                | 3                       |

Page 3 Sec 1 3/4 At 1" Ln 1 Col 1 REC TRK EXT

Figura 60. Output testuale in Word

Se si esegue l'esportazione in un file di PowerPoint, ogni elemento esportato viene inserito in una presentazione separata. Le tabelle pivot esportate in PowerPoint diventano tabelle di Word e mantengono la formattazione della tabella pivot originale, compresi i caratteri, i colori, i bordi e così via.

Microsoft PowerPoint - [OUTPUT.ppt]

Outline Slides x

- 1 Frequencies: Sta
- 2 Frequencies: Ger
- 3 Frequencies: Inc category in thou
- 4 Graph: Bar of coi inccat

### Frequencies: Gender

Gender

|       |        | Frequency | Percent | Valid Percent | Cumulative Percent |
|-------|--------|-----------|---------|---------------|--------------------|
| Valid | Female | 3179      | 49.7    | 49.7          | 49.7               |
|       | Male   | 3221      | 50.3    | 50.3          | 100.0              |
|       | Total  | 6400      | 100.0   | 100.0         |                    |

Click to add notes

Slide 2 of 4 Default Design

Figura 61. Tabelle pivot in PowerPoint

I grafici selezionati per l'esportazione in PowerPoint vengono incorporati nel file di PowerPoint.

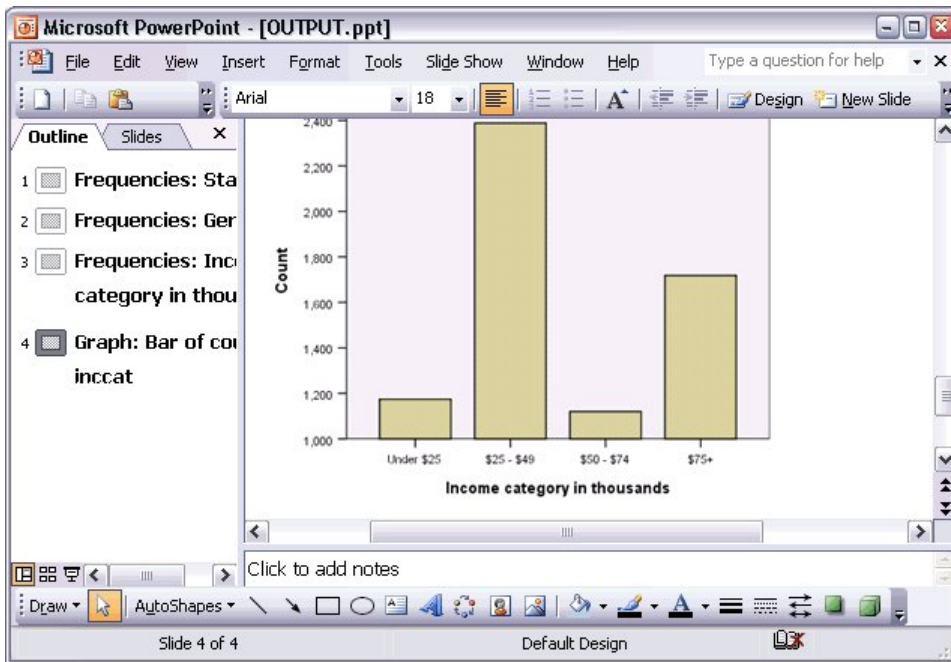


Figura 62. Grafici in PowerPoint

Nota: l'esportazione in PowerPoint è disponibile solo su sistemi operativi Windows e non è disponibile con la versione Student.

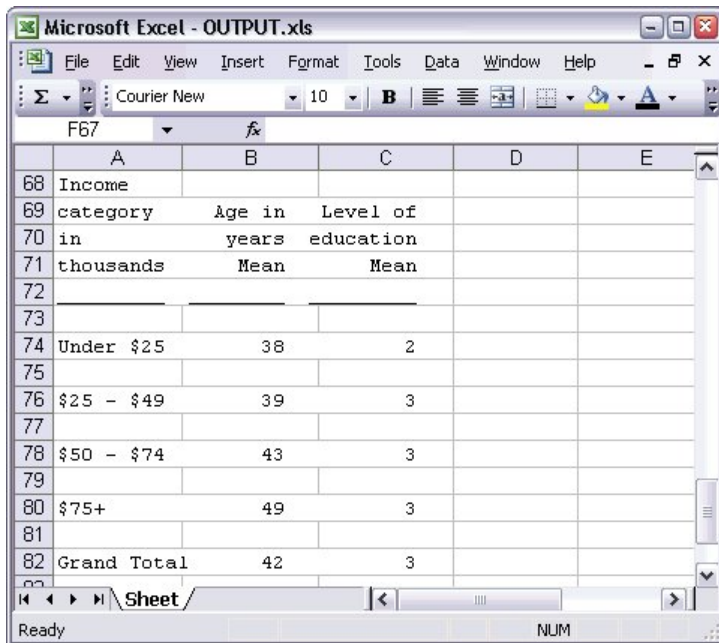
Se l'esportazione è verso un file Excel, i risultati verranno esportati in modo diverso.

| Gender                       |             |           |         |               |                    |  |
|------------------------------|-------------|-----------|---------|---------------|--------------------|--|
|                              |             | 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         |                    |  |
| Income category in thousands |             |           |         |               |                    |  |
|                              |             | Frequency | Percent | Valid Percent | Cumulative Percent |  |
| Valid                        | Under \$25  | 1,174     | 18.3    | 18.3          | 18.3               |  |
|                              | \$25 - \$49 | 2,368     | 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         |                    |  |

Figura 63. Output.xls in Excel

Celle, colonne e righe della tabella pivot diventano celle, colonne e righe di Excel.

Ogni riga nell'output testuale è una riga del file di Excel, con tutti i contenuti di una riga raccolti in una singola cella.



|    | 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         |   |   |

Figura 64. Output testuale in Excel

## Esportazione dei risultati in formato PDF

È possibile esportare tutti gli elementi o solo gli elementi selezionati nel Visualizzatore in un file PDF (portable document format).

1. Dai menu della finestra del Visualizzatore contenente il risultato che si desidera esportare in PDF scegliere:  
**File > Esporta...**
2. Nella finestra di dialogo Esporta output, nell'elenco a discesa Tipo file in Esporta formato, scegliere **Portable Document Format**.



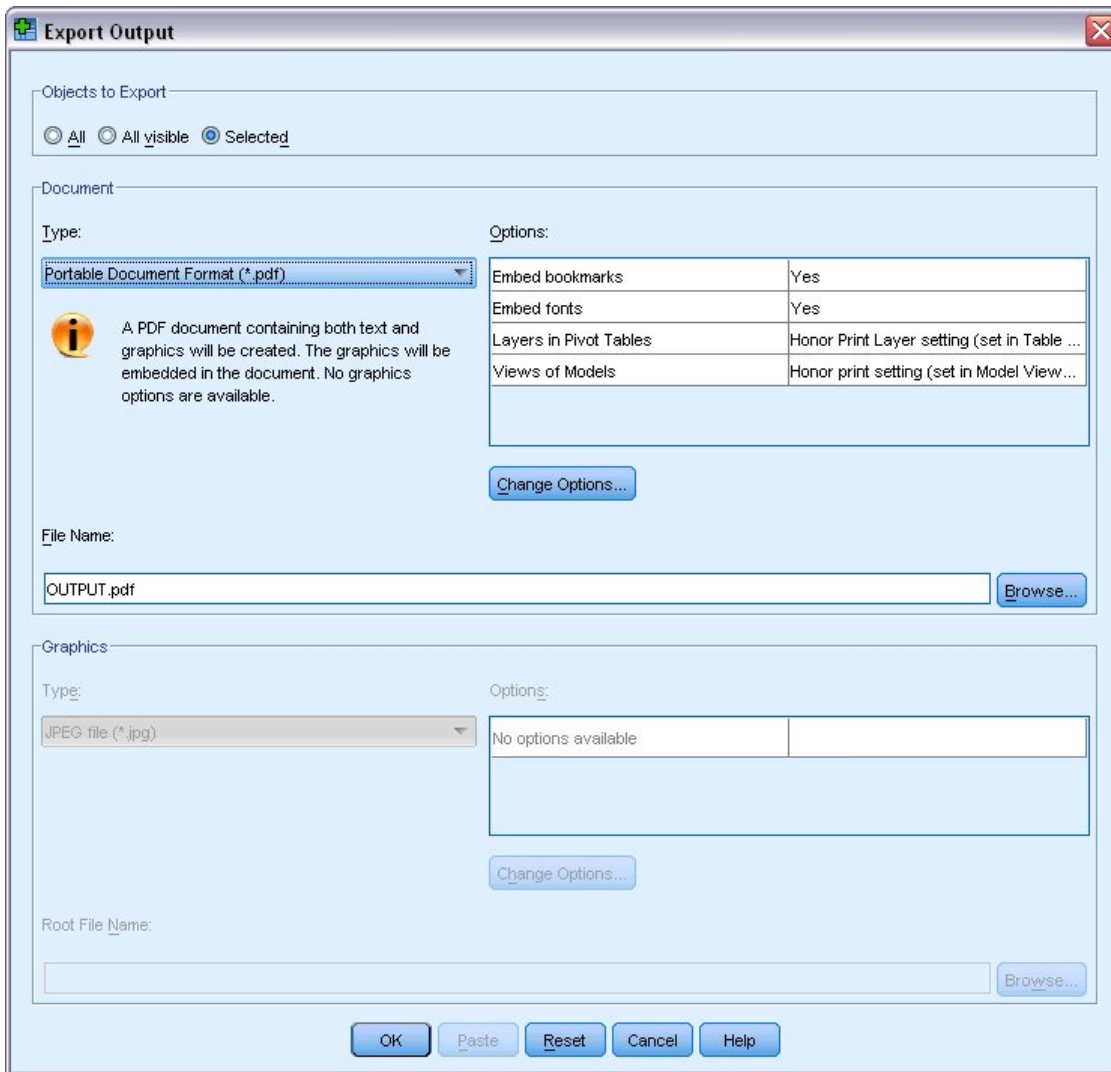


Figura 65. Finestra di dialogo Esporta output

- Il riquadro schema del documento del Visualizzatore viene convertito in segnalibri nel file PDF per semplicità di navigazione.
- Le dimensioni della pagina, l'orientamento, i margini, il contenuto e la visualizzazione di intestazioni e piè di pagina nonché le dimensioni del grafico stampato nei documenti PDF sono opzioni controllate dell'impostazione della pagina (menu File, Imposta pagina nella finestra del Visualizzatore).
- La risoluzione (DPI) del documento PDF corrisponde alla risoluzione corrente della stampante predefinita o correntemente selezionata (modificabile in Imposta pagina). La risoluzione massima è 1200 DPI. Se l'impostazione della stampante è maggiore, la risoluzione del documento PDF sarà 1200 DPI. *Nota:* i documenti ad elevata risoluzione possono produrre scarsi risultati sulle stampanti a bassa risoluzione.



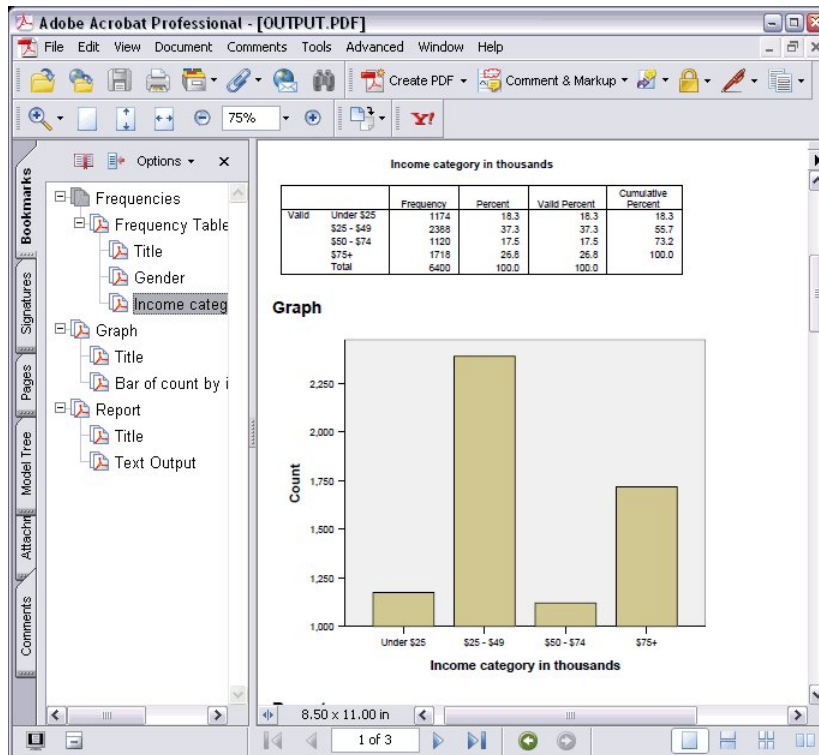


Figura 66. File PDF con segnalibri

## Esportazione dei risultati in formato HTML

Si possono esportare i risultati anche in formato HTML (hypertext markup language). Durante le operazioni di salvataggio come HTML, l'intero output non grafico viene esportato in un file HTML unico.

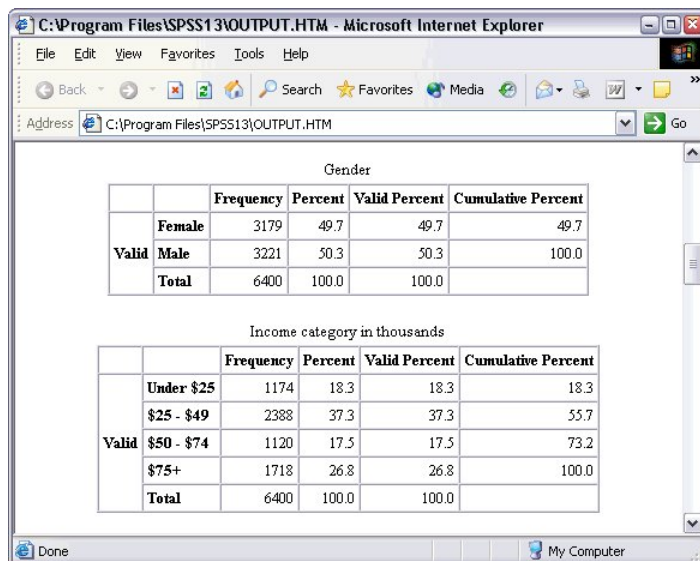


Figura 67. Output.htm nel browser Web

Quando si esegue l'esportazione in HTML, è possibile esportare anche i grafici ma non in un unico file.

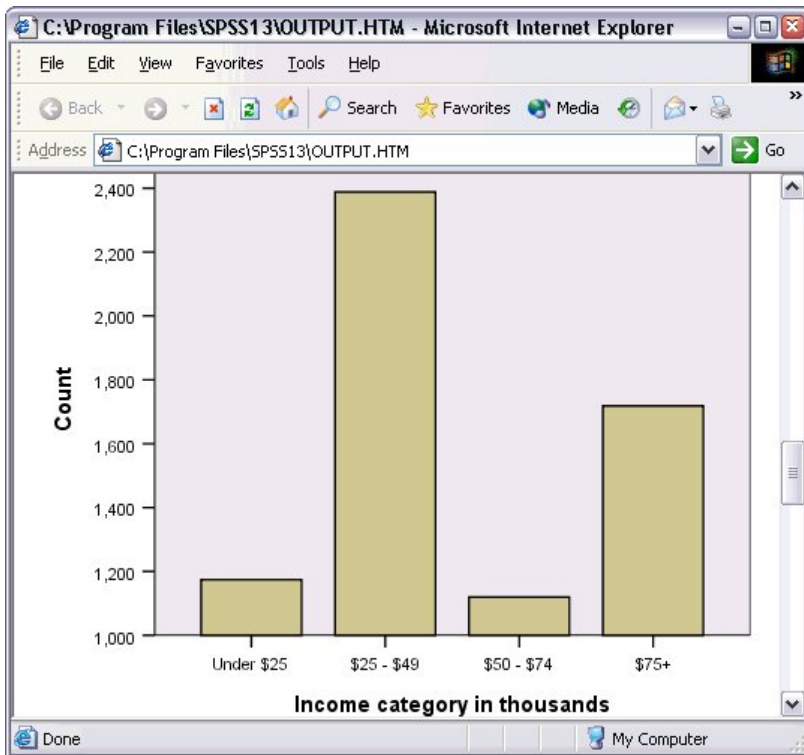


Figura 68. Grafico in HTML

Ogni grafico verrà salvato come un file del formato specificato e nel documento HTML verranno inseriti riferimenti ai file grafici indicati. È disponibile anche un'opzione per esportare tutti i grafici (o quelli selezionati) in file di grafici separati.

---

## Capitolo 7. Utilizzo della sintassi

Il potente linguaggio a comandi consente di salvare e automatizzare molte delle attività più comuni. Inoltre, offre alcune funzionalità a cui non è possibile accedere tramite i menu e le finestre di dialogo. È possibile accedere alla maggior parte dei comandi tramite i menu e le finestre di dialogo. Tuttavia, alcuni comandi e opzioni sono disponibili solo utilizzando il linguaggio dei comandi. Questo linguaggio consente inoltre di salvare i lavori in un file di sintassi, in modo da poter ripetere l'analisi successivamente.

Un file di sintassi dei comandi è un semplice file di testo che contiene comandi di sintassi IBM SPSS Statistics. È possibile aprire una finestra della sintassi e digitarvi direttamente i comandi, ma spesso è più semplice affidare parte o tutto il lavoro alle finestre di dialogo.

Gli esempi di questo capitolo utilizzano il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

*Nota:* la sintassi dei comandi non è disponibile con la versione Student.

---

### Incollare la sintassi

Il modo più semplice per creare la sintassi è utilizzare il pulsante Incolla situato nella maggior parte delle finestre di dialogo.

1. Aprire il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.
2. Dai menu, scegliere:  
**Analizza > Statistiche descrittive > Frequenze...**
3. Selezionare *Stato civile [statociv]* e spostare la variabile nell'elenco Variabili.
4. Fare clic su **Grafici**.
5. Nella finestra di dialogo Grafici, selezionare **Grafici a barre**.
6. Nel gruppo Valori nel grafico, selezionare **Percentuali**.
7. Fare clic su **Continua**. Fare clic su **Incolla** per copiare la sintassi creata come risultato delle selezioni della finestra di dialogo per l'Editor di sintassi.

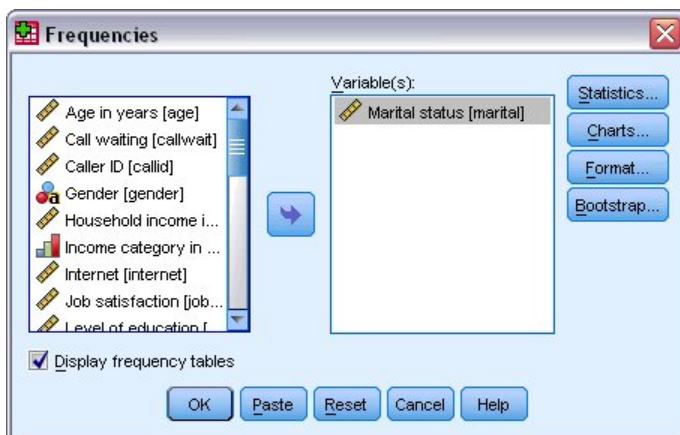


Figura 69. Finestra di dialogo Frequenze

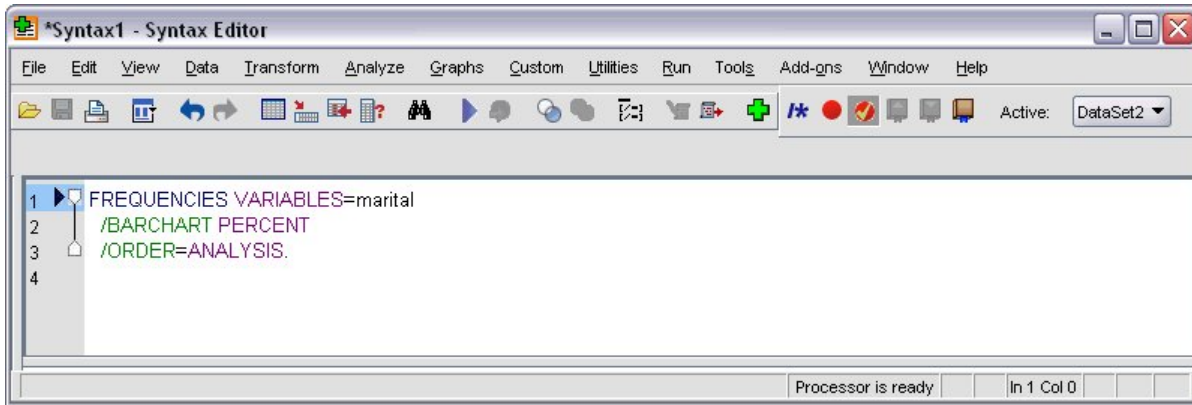


Figura 70. Sintassi di frequenze

8. Per eseguire la sintassi correntemente visualizzata, scegliere:  
**Esegui > Selezione**

---

## Modificare la sintassi

Nella finestra della sintassi è possibile modificare la sintassi. Ad esempio, è possibile modificare il sottocomando /BARCHART in modo da visualizzare le frequenze anziché le percentuali. (I sottocomandi sono indicati da una barra). Se si conosce la parola chiave per la visualizzazione delle frequenze, è possibile immetterla direttamente. In caso contrario, è possibile ottenere un elenco delle parole chiave disponibili per il sottocomando posizionando il cursore in un punto qualsiasi dopo il nome del sottocomando e premendo Ctrl+Barra spaziatrice. In questo modo si visualizza il controllo di autocompletamento per il sottocomando.

Eliminare la parola chiave PERCENT dal sottocomando BARCHART.

Premere Ctrl-Barra spaziatrice.

1. Fare clic sull'elemento etichettato **FREQ** per le frequenze. Facendo clic su un elemento nel controllo autocompletamento lo si inserisce nella posizione corrente del cursore.  
 Per impostazione predefinita, il controllo di autocompletamento proporrà un elenco di termini disponibili durante la digitazione. Ad esempio, si potrebbe decidere di includere un grafico a torta insieme a un grafico a barre. Il grafico a torta viene specificato con un sottocomando distinto.
2. Premere Invio dopo la parola chiave **FREQ** e digitare una barra (/) per indicare l'inizio di un sottocomando.

L'Editor di sintassi richiede all'utente l'elenco dei sottocomandi per il comando corrente.

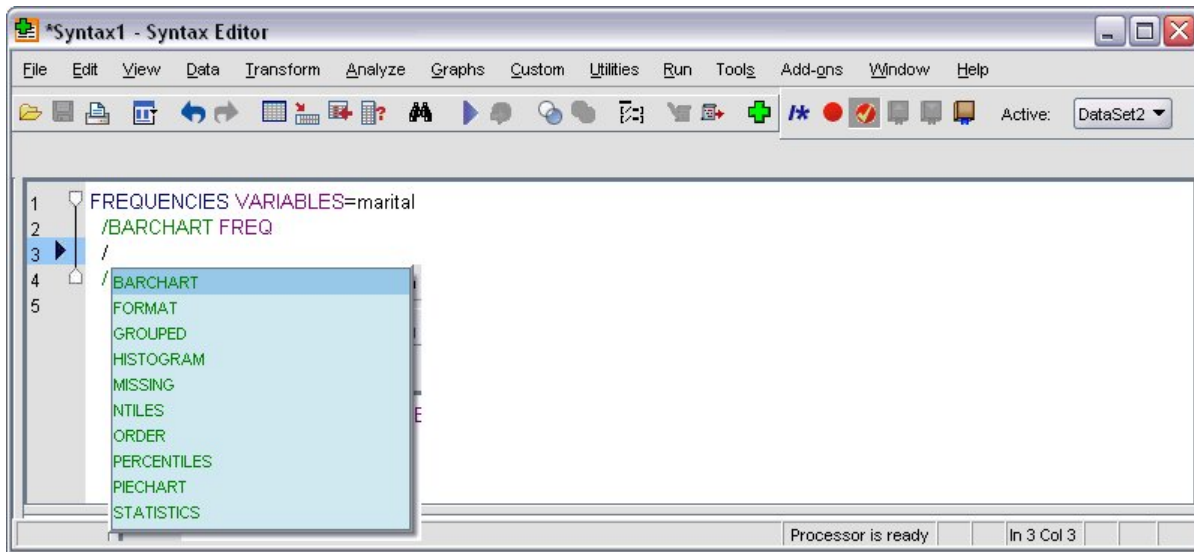


Figura 71. Controllo di autocompletamento che visualizza i sottocomandi

Per ottenere una guida più dettagliata per il comando corrente, premere il tasto F1. Questo porta l'utente direttamente alle informazioni di riferimento della sintassi del comando per il comando corrente.

Si è forse notato come il testo visualizzato nella finestra della sintassi sia colorato. La codifica colori consente di individuare rapidamente i termini non riconosciuti, dal momento che solo i termini riconosciuti sono colorati. Ad esempio, il sottocomando FORMAT viene erroneamente digitato come FRMAT. Per impostazione predefinita, i sottocomandi vengono colorati in verde, ma il testo FRMAT apparirà senza colori, poiché non è un termine riconosciuto.

---

## Apertura ed esecuzione di un file di sintassi

1. Per aprire un file di sintassi salvato, dai menu scegliere:  
**File > Apri > Sintassi...**  
Viene visualizzata una finestra di dialogo standard per l'apertura dei file.
2. Selezionare un file di sintassi. Se non vengono visualizzati file di sintassi, accertarsi che venga selezionato **Sintassi (\*.sps)** come tipo di file da visualizzare.
3. Fare clic su **Apri**.
4. Utilizzare il menu **Esegui** nella finestra della sintassi per eseguire i comandi.

Se i comandi sono validi per un tipo di file specifico, prima di eseguirli è necessario aprire il file di dati oppure inserire un comando che apre tale file. Questo tipo di comando può essere incollato dalle finestre di dialogo nelle quali è possibile aprire i file di dati.

---

## Utilizzo dei punti di interruzione

I punti di interruzione consentono di interrompere l'esecuzione della sintassi dei comandi in punti specifici all'interno della sintassi e di riprendere l'esecuzione quando si è pronti. In questo modo è possibile visualizzare l'output o i dati in un punto intermedio di un lavoro di sintassi, oppure di eseguire una sintassi di comando che visualizza le informazioni sullo stato corrente dei dati, ad esempio FREQUENCIES. I punti di interruzione possono essere impostati solo a livello di un comando, non su righe specifiche all'interno di un comando.

Per inserire un punto di interruzione in un comando:

1. Fare clic in un punto qualsiasi dell'area alla sinistra del testo associato al comando.

Il punto di interruzione è rappresentato da un cerchio rosso nell'area alla sinistra del testo del comando e sulla stessa riga del nome del comando indipendentemente dal punto in cui si è fatto clic con il mouse.

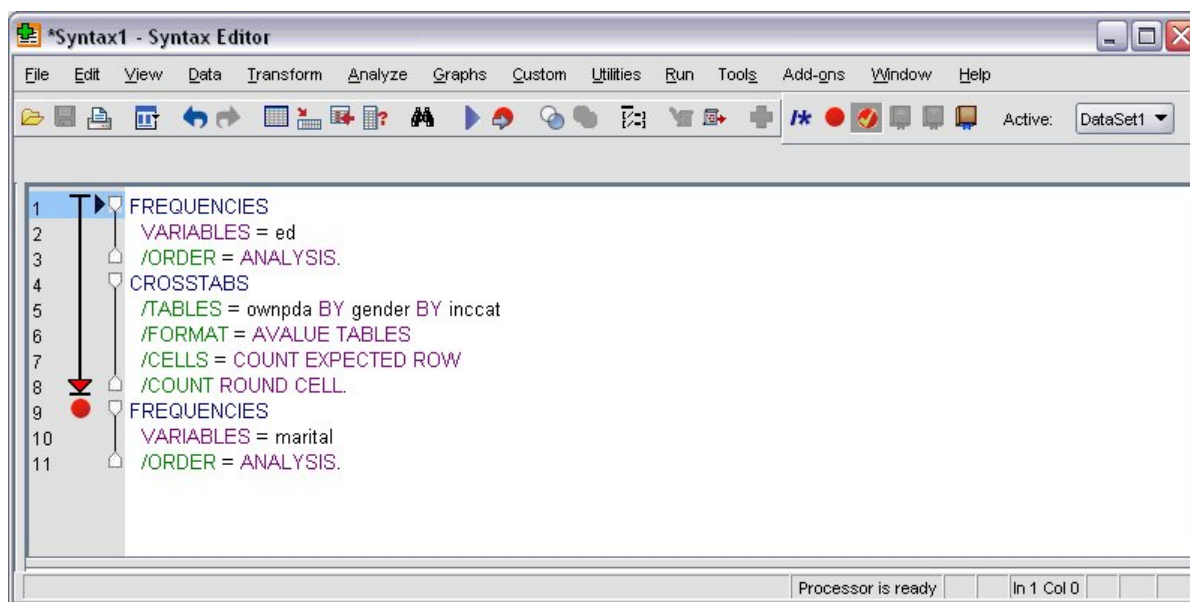


Figura 72. Esecuzione interrotta in corrispondenza di un punto di interruzione

Quando si esegue la sintassi dei comandi che contiene punti di interruzione, l'esecuzione viene interrotta prima di ciascun comando contenente un punto di interruzione.

La freccia che punta verso il basso alla sinistra del testo del comando mostra l'avanzamento dell'esecuzione della sintassi. Abbraccia l'area dall'esecuzione del primo comando fino all'esecuzione dell'ultimo comando ed è particolarmente utile quando si esegue la sintassi dei comandi contenente punti di interruzione.

Per riprendere l'esecuzione in seguito a un punto di interruzione:

2. Dai menu dell'Editor di sintassi, scegliere:

**Esegui > Continua**

---

## Capitolo 8. Modifica dei valori dei dati

I dati da cui si parte non sempre sono organizzati nel modo più consono alle esigenze analitiche o di reporting. Ad esempio, è possibile che si desideri eseguire una delle operazioni seguenti:

- Creare una variabile categoriale da una variabile di scala.
- Unire più categorie di risposta in un'unica categoria.
- Creare una variabile nuova che rappresenta la differenza calcolata tra due variabili esistenti.
- Calcolare il periodo di tempo tra due date.

In questo capitolo viene utilizzato il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

---

### Creazione di una variabile categoriale da una variabile di scala.

Varie variabili categoriali presenti nel file di dati *demo.sav* derivano infatti da variabili di scala presenti nello stesso file di dati. Ad esempio, la variabile *catredd* è semplicemente il *reddito* raggruppato in quattro categorie. Questa variabile categoriale utilizza i valori numero intero 1–4 per rappresentare le seguenti categorie di reddito (in migliaia): inferiore a \$25, \$25–\$49, \$50–\$74 e \$75 o superiore.

Per creare la variabile categoriale *inccat*:

1. Dai menu nella finestra dell'Editor dei dati, scegliere:

**Trasforma > Raccolta visuale...**

Nella finestra di dialogo iniziale Raccolta visuale, selezionare le variabili ordinali e/o di scala per cui si desiderano creare nuove variabili raccolte. Con il termine **raccolta** si intende l'operazione che consente di raggruppare due o più valori contigui nella stessa categoria.

Poiché Raccolta visuale si basa sui valori reali del file di dati per eseguire scelte di raccolta corrette, prima deve leggere il file di dati. Poiché tale operazione potrebbe richiedere tempi lunghi se il file di dati contiene un numero molto elevato di casi, la finestra di dialogo iniziale consente inoltre di limitare il numero di casi da leggere ("esaminare"). Non si tratta di un'operazione necessaria per il file di dati campione. Sebbene i casi siano oltre 6.000, non occorre molto tempo per esaminarli.

2. Trascinare *Reddito familiare in migliaia [reddito]* dall'elenco Variabili e rilasciarlo nell'elenco Variabili da raccogliere, quindi fare clic su **Continua**.

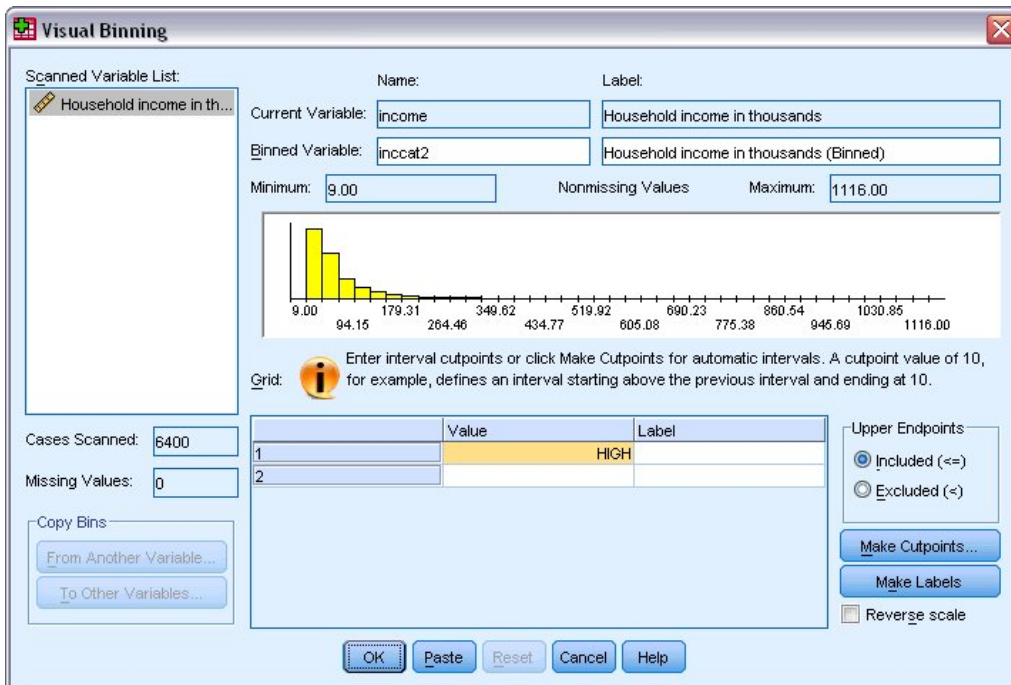


Figura 73. Finestra di dialogo principale Raccolta visuale

3. Nella finestra di dialogo principale Raccolta visuale, selezionare *Reddito familiare in migliaia [reddito]* nell'elenco delle variabili scansionate.

Un istogramma visualizza la distribuzione della variabile selezionata (che in questo caso è molto asimmetrica).

4. Inserire `catredd2` per la nuova variabile raccolta e *Categoria reddito [in migliaia]* per l'etichetta di variabile.
5. Fare clic su **Crea punti divisione**.
6. Selezionare **Intervalli di larghezza uguale**.
7. Inserire 25 per l'ubicazione del primo punto di divisione, 3 per il numero di punti di divisione e 25 per la larghezza.

Il numero di categorie raccolte corrisponde al numero di punti di divisione più uno. Quindi, in questo esempio, la nuova variabile raccolta avrà quattro categorie: le prime tre contengono ognuna intervalli di 25 (in migliaia) e l'ultima contiene tutti i valori sopra al valore del punto di divisione più alto, uguale a 75 (in migliaia).

8. Fare clic su **Applica**.

I valori visualizzati adesso nella griglia rappresentano i punti di divisione definiti, ovvero i punti finali superiori di ogni categoria. Le linee verticali nell'istogramma indicano inoltre le ubicazioni dei punti di divisione.

Per impostazione predefinita, i valori di questi punti di divisione vengono inclusi nelle categorie corrispondenti. Ad esempio, il primo valore pari 25 includerebbe tutti i valori inferiori o uguali a 25. Ma in questo esempio si desiderano categorie che corrispondano a meno di 25, 25–49, 50–74 e 75 o più.

9. Nel gruppo Punti finali superiori, selezionare **Escluso (<)**.
10. Quindi fare clic su **Crea etichette**.



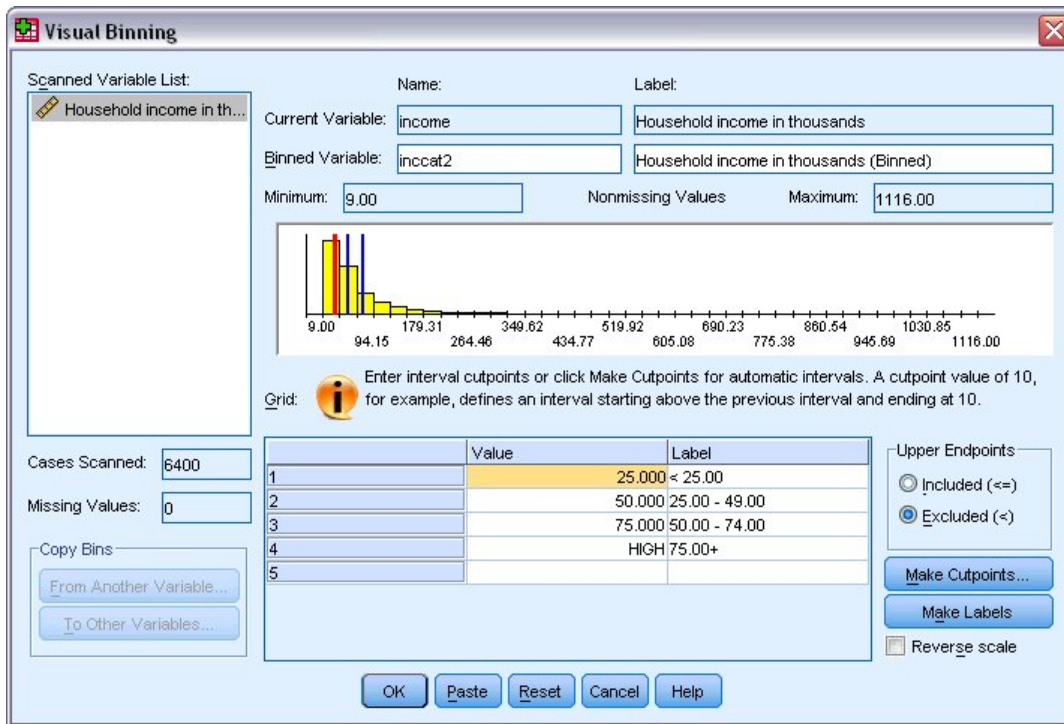


Figura 74. Etichette di variabile generate automaticamente

Questa operazione genera automaticamente etichette valore descrittive per ogni categoria. Poiché i valori effettivi assegnati alla nuova variabile raccolta sono semplicemente numeri interi sequenziali che iniziano con 1, le etichette valore posso risultare molto utili.

È inoltre possibile specificare o modificare manualmente i punti di divisione e le etichette nella griglia, modificare le ubicazioni dei punti di divisione trascinando le linee dei punti di divisione nell'istogramma ed eliminare i punti di divisione trascinando e rilasciando le linee dei punti di divisione all'esterno dell'istogramma.

11. Fare clic su **OK** per creare la nuova variabile raccolta.

La nuova variabile viene visualizzata nell'Editor dei dati. Poiché la variabile viene aggiunta al termine del file, viene visualizzata nella colonna più a destra della Vista dati e nell'ultima riga della Vista variabile.

## Calcolo di nuove variabili

È possibile calcolare nuove variabili basandosi su equazioni altamente complesse, utilizzando una grande varietà di funzioni matematiche. Tuttavia, in questo esempio verrà semplicemente calcolata una nuova variabile che è la differenza tra i valori di due variabili esistenti.

Il file di dati *demo.sav* contiene una variabile per l'età corrente del rispondente e una variabile per il numero di anni trascorsi nell'attuale posto di lavoro. Tuttavia, non contiene una variabile per l'età del rispondente nel momento in cui iniziò tale lavoro. È possibile creare una nuova variabile che è la differenza calcolata tra l'età corrente e il numero di anni trascorsi nell'attuale posto di lavoro, che dovrebbe essere l'età approssimativa in cui il rispondente ha iniziato tale lavoro.

1. Dai menu nella finestra dell'Editor dei dati, scegliere:  
**Trasforma > Calcola variabile...**
2. Per la Variabile obiettivo, digitare `iniziolavoro`.
3. Selezionare *Età in anni [età]* nell'elenco delle variabili origine e fare clic sul pulsante freccia per copiarla nella casella di testo Espressione numerica.

4. Fare clic sul pulsante meno (-) sulla calcolatrice nella finestra di dialogo (o premere il tasto meno sulla tastiera).
5. Selezionare *Anni di permanenza nell'impiego attuale [impiego]* e fare clic sul pulsante freccia per copiarla nell'espressione.

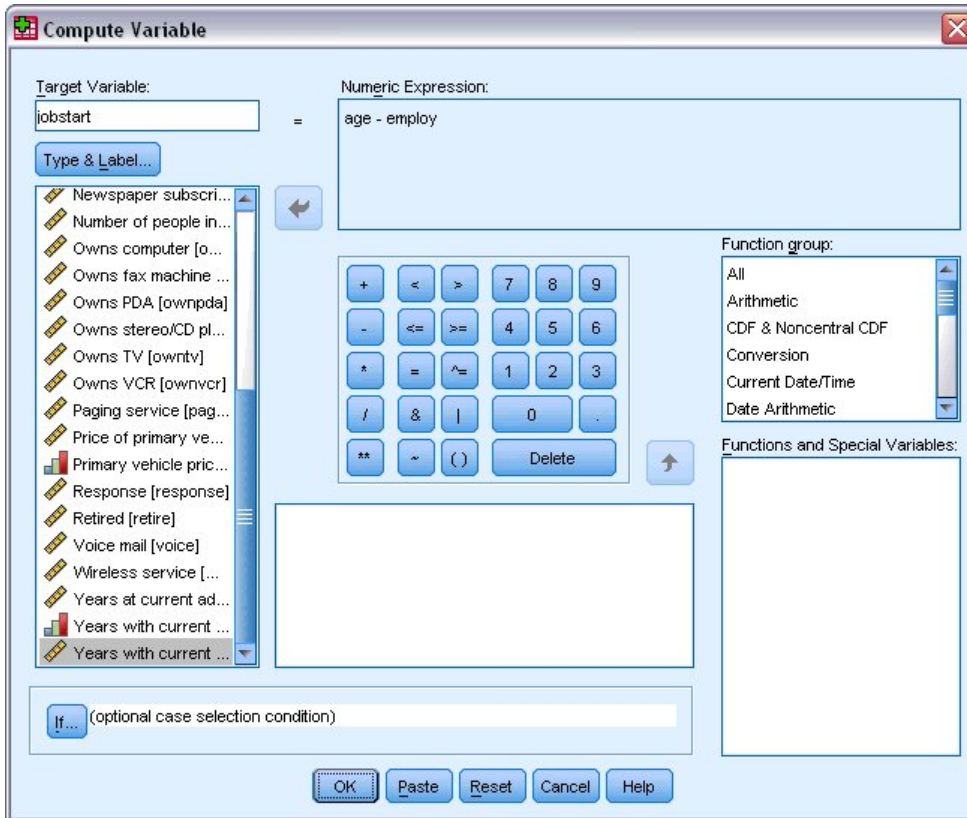


Figura 75. Finestra di dialogo Calcola variabile

*Nota:* fare attenzione a selezionare la variabile di impiego corretta. Esiste anche una versione categoriale ricodificata della variabile, che *non* è quella desiderata. L'espressione numerica deve essere *età-impiego*, non *età-catimpiego*.

6. Fare clic su **OK** per calcolare la nuova variabile.

La nuova variabile viene visualizzata nell'Editor dei dati. Poiché la variabile viene aggiunta al termine del file, viene visualizzata nella colonna più a destra della Vista dati e nell'ultima riga della Vista variabile.

## Utilizzo delle funzioni nelle espressioni

È inoltre possibile utilizzare le funzioni predefinite nelle espressioni. Sono disponibili oltre 70 funzioni incorporate, tra le quali:

- Funzioni aritmetiche
- Funzioni statistiche
- Funzioni di distribuzione
- Funzioni logiche
- Funzioni di aggregazione e di estrazione relative alla data e all'ora
- Funzioni per valori mancanti
- Funzioni incrociate
- Funzioni di stringa

Le funzioni sono organizzate in gruppi divisi in modo logico, ad esempio uno per le operazioni aritmetiche e un altro per il calcolo delle metriche statistiche. Per comodità, anche alcune variabili di sistema di uso comune, come *\$TIME* (ora e data correnti) sono incluse in gruppi di funzioni appropriate.

Incollare una funzione in un'espressione

Per incollare una funzione in un'espressione:

1. Posizionare il cursore all'interno dell'espressione, nel punto in cui si desidera visualizzare la funzione.
2. Selezionare il gruppo appropriato dall'elenco Gruppo di funzioni. Il gruppo con etichetta **Tutti** fornisce un elenco di tutte le funzioni e di tutte le variabili di sistema disponibili.
3. Fare doppio clic sulla funzione desiderata nell'elenco Funzioni e Variabili speciali, oppure selezionare la funzione e fare clic sulla freccia accanto all'elenco Gruppo di funzioni.

La funzione verrà inserita nell'espressione. Se si seleziona una parte dell'espressione e quindi si inserisce la funzione, la parte selezionata diventerà il primo argomento della funzione.

Modifica di una funzione di un'espressione

La funzione non è completa fino a quando non si inseriscono gli argomenti, rappresentati da punti interrogativi nella funzione incollata. Il numero dei punti interrogativi indica il numero minimo di argomenti necessari per completare la funzione.

1. Evidenziare i punti interrogativi nella funzione incollata.
2. Inserire gli argomenti. Se si tratta di nomi di variabili, è possibile incollarli dall'elenco di variabili.

## Utilizzo delle espressioni condizionali

Le espressioni condizionali consentono di applicare le trasformazioni ai sottoinsiemi di casi selezionati. Un'espressione condizionale restituisce un valore vero, falso o mancante per ciascun caso. Se il risultato dell'espressione condizionale è vero, la trasformazione verrà applicata al caso. Se il risultato è falso o mancante, la trasformazione non verrà applicata al caso.

Per specificare un'espressione condizionale:

1. Fare clic su **Se** nella finestra di dialogo Calcola variabile. Viene aperta la finestra di dialogo Condizioni.
2. Selezionare **Includi se il caso soddisfa la condizione**.
3. Specificare l'espressione condizionale.

La maggior parte delle espressioni condizionali contiene almeno un operatore relazionale, ad esempio

`age>=21`

o

`income*3<100`

Nel primo esempio, vengono selezionati solo i casi con valore 21 o superiore per *Età [age]*. Nel secondo esempio, *Reddito familiare in migliaia [income]* moltiplicato per 3 deve essere inferiore a 100 affinché un caso possa essere selezionato.

È inoltre possibile collegare due o più espressioni condizionali mediante gli operatori logici, ad esempio:

`age>=21 | ed>=4`

o

income\*3<100 & ed=5

Nel primo esempio, vengono selezionati i casi che soddisfano la condizione *Età [age]* o *Livello di istruzione [ed]*. Nel secondo esempio, entrambe le condizioni *Reddito familiare in migliaia [income]* e *Livello di istruzione [ed]* devono essere soddisfatte affinché un caso possa essere selezionato.

## Utilizzo di date e orari

Numerose attività eseguite comunemente con date e orari possono essere realizzate facilmente utilizzando la Procedura guidata data e ora. La procedura guidata offre le funzioni seguenti:

- Creazione di una variabile di data/ora da una variabile stringa che contiene una data o un orario.
- Creazione di una variabile di data/ora dall'unione di variabili contenenti varie parti della data o dell'ora.
- Aggiunta o sottrazione di valori da variabili di data/ora, inclusa aggiunta o sottrazione di due variabili di data/ora.
- Estrazione di una parte di una variabile di data o ora; ad esempio, il giorno del mese da una variabile di data/ora con formato mm/gg/aaaa.

Gli esempi di questa sezione utilizzano il file di dati *upgrade.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

Per utilizzare la Procedura guidata data e ora

1. Dai menu, scegliere:

**Trasforma > Procedura guidata Data e ora...**



Figura 76. Schermata introduttiva della Procedura guidata data e ora

Nella schermata introduttiva della Procedura guidata data e ora è visualizzato un insieme di attività di carattere generale. Le attività non applicabili ai dati correnti sono disabilitate. Ad esempio, il file di dati *upgrade.sav* non contiene variabili stringa, quindi l'attività di creazione di una variabile di data da una stringa è disattivata.

Se non si ha familiarità con le operazioni relative a date e ore in IBM SPSS Statistics, selezionare **Visualizzare informazioni sulla rappresentazione di date e ore** e fare clic su **Successivo**. Verrà visualizzata una schermata che fornisce una breve panoramica sulle variabili di data e ora e un collegamento, tramite il pulsante Aiuto, a informazioni più dettagliate.

## Calcolo del periodo di tempo tra due date

Una delle attività più comuni relative alle date è il calcolo del periodo di tempo tra due date. Si consideri ad esempio un'azienda produttrice di software interessata ad analizzare gli acquisti di licenze di aggiornamento tramite la determinazione del numero di anni dall'ultimo acquisto di un aggiornamento da parte di ogni cliente. Il file di dati *upgrade.sav* contiene una variabile per la data in cui ogni cliente ha acquistato l'ultimo aggiornamento ma non il numero di anni trascorsi da tale data. Una nuova variabile che rappresenti il periodo di tempo in anni tra la data dell'ultimo aggiornamento e quella della prossima release del prodotto fornirebbe una misura di questa quantità.

Per calcolare il periodo di tempo tra due date:

1. Selezionare **Eseguire calcoli con date e ore** nella schermata introduttiva della Procedura guidata data e ora e fare clic su **Avanti**.
2. Selezionare **Calcolare il numero di unità di tempo tra due date** e fare clic su **Avanti**.

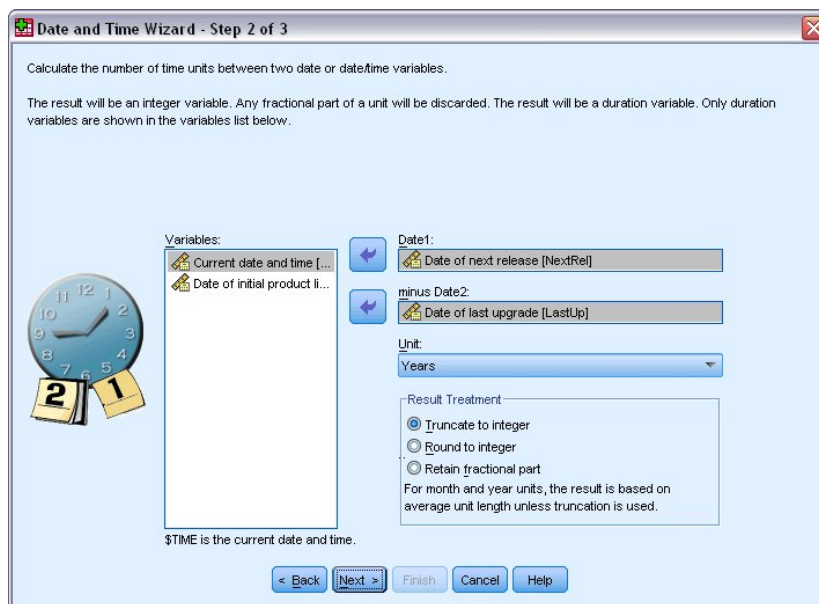


Figura 77. Calcolo del periodo di tempo tra due date: Fase 2

3. Nella fase 2, selezionare *Data della prossima release* per Data1.
4. Selezionare *Data dell'ultimo aggiornamento* per Data2.
5. Selezionare **Anni** per l'Unità e **Tronca a intero** per il Trattamento risultato. Si tratta di selezioni predefinite.
6. Fare clic su **Avanti**.
7. Nella fase 3, immettere *AnniUltimoAgg* per il nome della variabile risultato. Le variabili risultato non possono avere lo stesso nome di una variabile esistente.
8. Digitare *Anni dall'ultimo aggiornamento* come etichetta per la variabile risultato. Le etichette di variabile per le variabili risultato sono facoltative.
9. Non modificare la selezione predefinita di **Crea la variabile ora** e fare clic su **Fine** per creare la nuova variabile.

La nuova variabile, *AnniUltAgg*, visualizzata nell'Editor dei dati, rappresenta il numero intero di anni tra le due date. Le parti frazionarie di un anno vengono troncate.

## Aggiunta di una durata a una data

È possibile aggiungere o sottrarre a una data durate come 10 giorni o 12 mesi. Continuando con l'esempio della sezione precedente, si consideri di voler determinare la data di termine del contratto iniziale di assistenza tecnica di ogni cliente. Il file di dati *upgrade.sav* contiene una variabile per il numero di anni di durata del contratto di assistenza e una variabile per la data di acquisto iniziali. È quindi possibile determinare la data finale del supporto iniziale aggiungendo gli anni di assistenza alla data di acquisto.

Per aggiungere una durata a una data:

1. Selezionare **Eseguire calcoli con date e ore** nella schermata introduttiva della Procedura guidata data e ora e fare clic su **Avanti**.
2. Nella fase 1, selezionare **Aggiungere o sottrarre una durata da una data** e fare clic su **Avanti**.

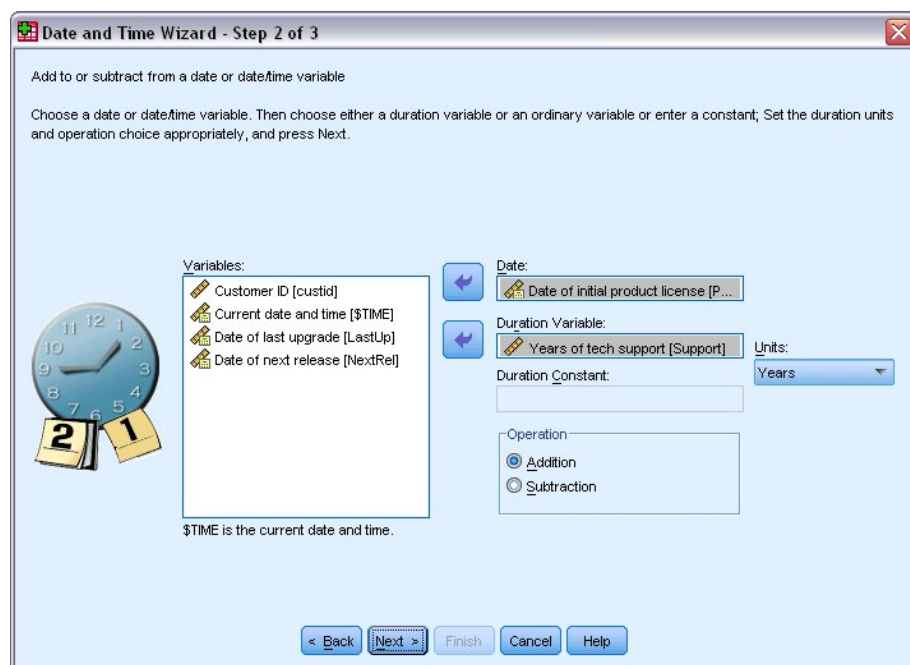


Figura 78. Aggiunta di una durata ad una data: Fase 2

3. Selezionare *Data della licenza prodotto iniziale* come Data.
4. Nella fase 2, selezionare *Anni di assistenza tecnica* per la variabile di durata.  
Poiché *Anni di assistenza tecnica* è semplicemente una variabile numerica, è necessario indicare le unità da utilizzare quando si aggiunge tale variabile come una durata.
5. Selezionare **Anni** dall'elenco a discesa Unità.
6. Fare clic su **Avanti**.
7. Nella fase 3, immettere *DataFineAss* per il nome della variabile risultato. Le variabili risultato non possono avere lo stesso nome di una variabile esistente.
8. Inserire *Data finale assistenza* come etichetta della variabile risultato. Le etichette di variabile per le variabili risultato sono facoltative.
9. Fare clic su **Fine** per creare la nuova variabile.

La nuova variabile viene visualizzata nell'Editor dei dati.

---

## Capitolo 9. Ordinamento e selezione di dati

I file di dati non sempre sono organizzati nella forma ottimale per esigenze specifiche. È tuttavia disponibile una vasta gamma di trasformazioni di file che consentono di preparare i dati per l'analisi, ad esempio:

- **Ordinamento dati.** È possibile ordinare i dati in base al valore di una o più variabili.
- **Selezione di sottoinsiemi di casi.** È possibile limitare l'analisi a un sottoinsieme di casi oppure eseguire contemporaneamente più analisi in diversi sottoinsiemi.

Gli esempi di questo capitolo utilizzano il file di dati *demo.sav*. Per ulteriori informazioni, consultare l'argomento Capitolo 10, "File di esempio", a pagina 83.

---

### Ordinamento di dati

In alcuni tipi di analisi è spesso utile e a volte necessario ordinare i dati, ovvero ordinare le righe del file di dati.

Per riordinare la sequenza dei casi del file di dati in base al valore di una o più variabili di ordinamento:

1. Dai menu, scegliere:

**Dati > Ordina casi...**

Si visualizza la Finestra di dialogo Ordina casi.

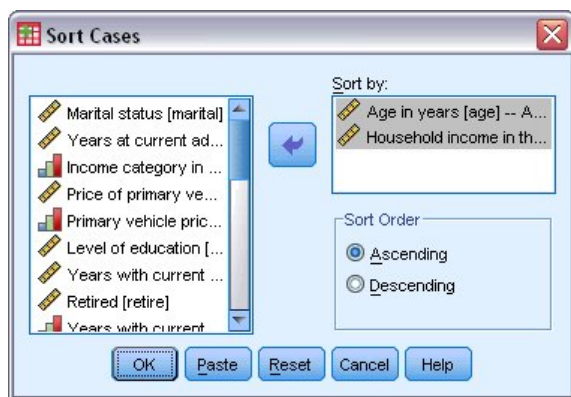


Figura 79. Finestra di dialogo Ordina casi.

2. Aggiungere le variabili *Età in anni [età]* e *Reddito familiare in migliaia [reddito]* all'elenco Ordinamento per.

Se si selezionano più variabili di ordinamento, l'ordinamento dei casi sarà determinato dall'ordine in cui tali variabili sono visualizzate nell'elenco Ordinamento per. In questo esempio, in base all'elenco Ordinamento per, i casi verranno ordinati in base al valore di *Reddito familiare in migliaia [reddito]* all'interno delle categorie di *Età in anni [età]*. Per le variabili stringa viene utilizzato un criterio di ordinamento tale che le lettere maiuscole precedano le minuscole corrispondenti. Ad esempio, la stringa *Sì* precederà la stringa *sì*.



## Elaborazione file suddiviso

Per suddividere il file di dati in gruppi distinti per l'analisi:

1. Dai menu, scegliere:

**Dati > File suddiviso...**

Viene visualizzata la finestra di dialogo File suddiviso.

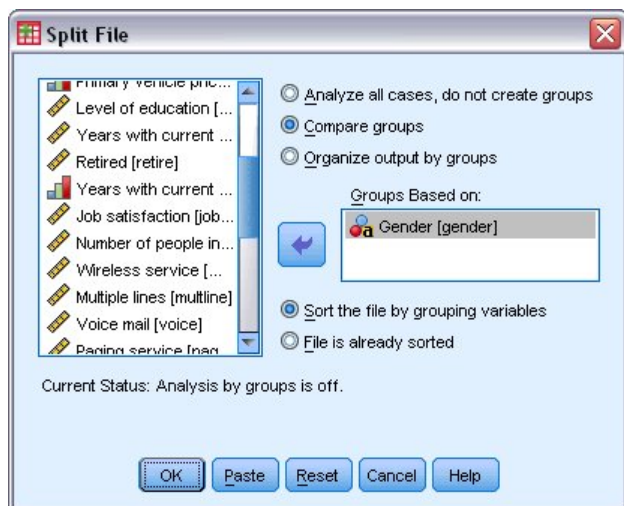


Figura 80. Finestra di dialogo File suddiviso

2. Selezionare **Confronta gruppi** o **Distingui i risultati per gruppo**. (Nell'esempio seguente vengono illustrate le differenze tra queste due opzioni).
3. Selezionare *Sesso [sesso]* per suddividere il file in gruppi separati in base a tali variabili.

È possibile utilizzare variabili numeriche, stringa corta e stringa lunga come variabili di raggruppamento. Per ciascun sottogruppo definito dalle variabili di raggruppamento verrà eseguita un'analisi distinta. Se si selezionano più variabili di raggruppamento, la modalità di raggruppamento dei casi sarà determinata dall'ordine in cui tali variabili sono visualizzate nell'elenco Gruppi basati su.

Se si seleziona **Confronta gruppi**, i risultati di tutti i gruppi di file suddivisi verranno inclusi nella stessa tabella, come indicato nella tabella delle statistiche di riepilogo seguente, generata dalla procedura Frequenze.

| 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 |

Figura 81. Output di file suddiviso con una singola tabella pivot



Se si seleziona **Organizza output per gruppi** e si esegue la procedura Frequenze, vengono create due tabelle pivot: una per le femmine e una per i maschi.

**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

Figura 82. Output di file suddiviso con una tabella pivot per le femmine

**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

Figura 83. Output di file suddiviso con una tabella pivot per i maschi

## Ordinamento dei casi per l'elaborazione file suddiviso

La procedura File suddiviso crea un nuovo sottogruppo ogni volta che rileva un valore diverso per una delle variabili di raggruppamento. È quindi importante ordinare i casi in base ai valori delle variabili di raggruppamento prima di avviare l'elaborazione file suddiviso.

Per impostazione predefinita la procedura File suddiviso ordina automaticamente il file di dati in base ai valori delle variabili di raggruppamento. Se al file è già applicato l'ordinamento corretto, è possibile ridurre il tempo di elaborazione selezionando **Il file è già ordinato**.

## Attivazione e disattivazione dell'elaborazione file suddiviso

Dopo essere stata avviata, l'elaborazione file suddiviso rimane attiva per il resto della sessione fino a quando non viene disattivata.

- **Analizza tutti i casi.** Questa opzione disattiva l'elaborazione file suddiviso.
- **Confronta gruppi e Organizza output per gruppi.** Questa opzione attiva l'elaborazione file suddiviso.

Se l'elaborazione del file suddiviso è attiva, nella barra di stato nella parte inferiore della finestra dell'applicazione verrà visualizzato il messaggio **File suddiviso attivo**.

---

## Selezione di sottoinsiemi di casi

È possibile limitare l'analisi a un sottogruppo specifico in base a criteri che includono variabili ed espressioni complesse. È inoltre possibile selezionare un campione casuale di casi. I criteri in base ai quali viene definito un sottogruppo includono:

- Valori di variabili e intervalli
- Intervalli di date e orari
- Numeri caso (righe)
- Espressioni aritmetiche

- Espressioni logiche
- Funzioni

Per selezionare un sottoinsieme di casi per l'analisi:

1. Dai menu, scegliere:

**Dati > Seleziona casi...**

Verrà aperta la finestra di dialogo Seleziona casi.

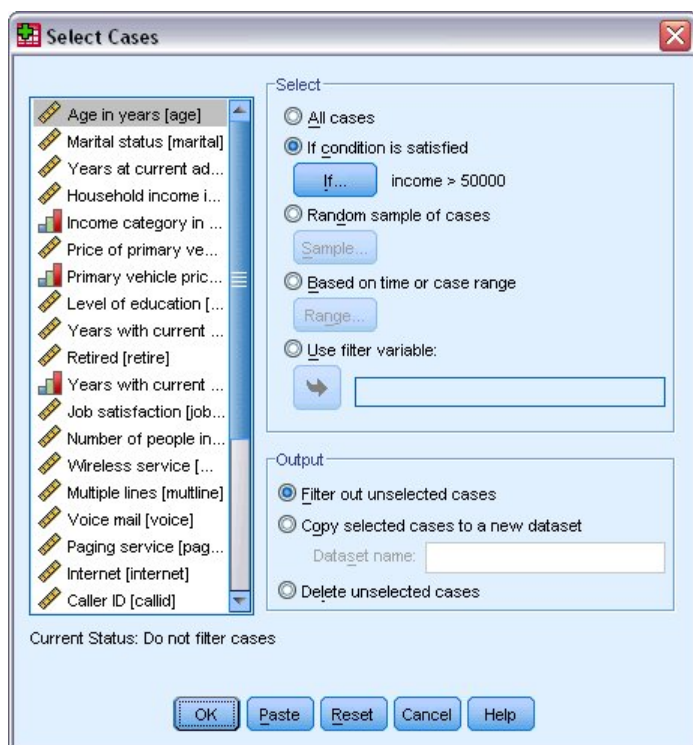


Figura 84. Finestra di dialogo Seleziona casi

## Per selezionare sottoinsiemi di casi in base a un'espressione condizionale

Per selezionare casi in base a un'espressione condizionale:

1. Selezionare **Se la condizione è soddisfatta** e fare clic su **Se** nella finestra di dialogo Seleziona casi.

Verrà aperta la finestra di dialogo Seleziona casi: Se.

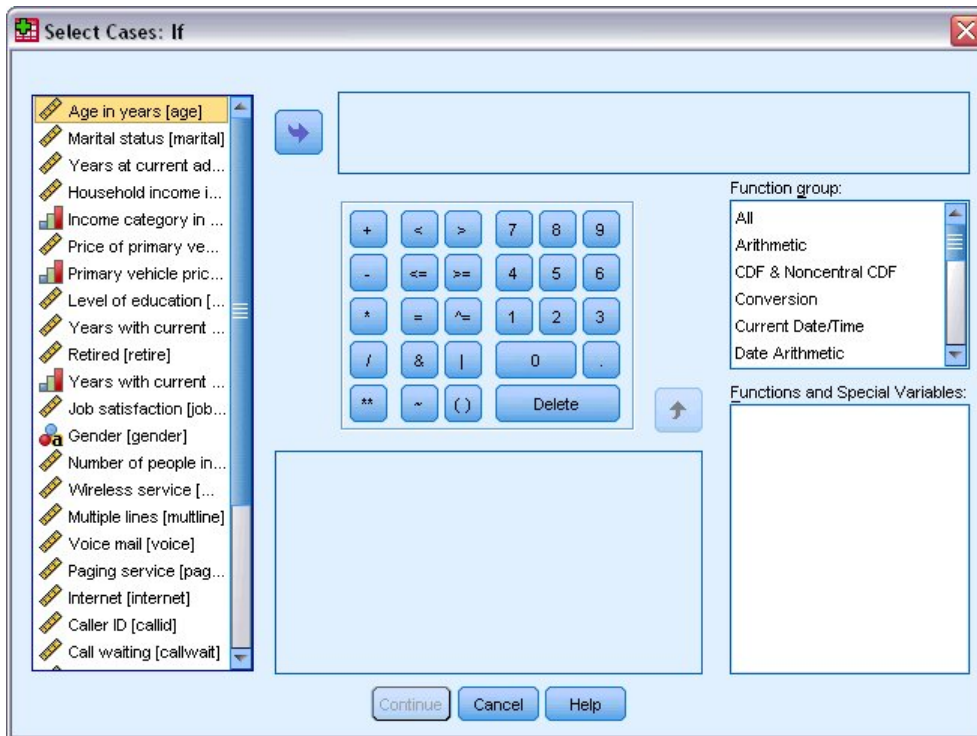


Figura 85. Finestra di dialogo Seleziona casi: Se

L'espressione condizionale può utilizzare nomi di variabili esistenti, costanti, operatori aritmetici, operatori logici, operatori relazionali e funzioni. È possibile digitare e modificare l'espressione nella casella di testo allo stesso modo del testo nella finestra dell'output. È inoltre possibile utilizzare la calcolatrice, l'elenco delle variabili e l'elenco delle funzioni per incollare elementi nell'espressione. Per ulteriori informazioni, consultare l'argomento "Utilizzo delle espressioni condizionali" a pagina 71.

## Per selezionare un campione casuale di dati

Per ottenere un campione casuale di dati:

1. Selezionare **Campione casuale di casi** nella finestra di dialogo Seleziona casi.
2. Fare clic su **Campione**.

Verrà aperta la finestra di dialogo Seleziona casi: Campione casuale.

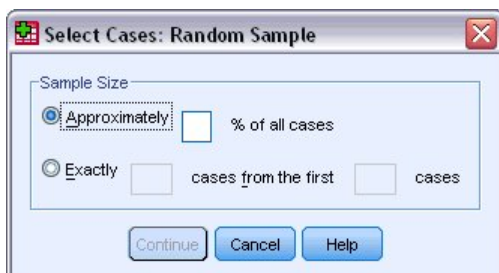


Figura 86. Finestra di dialogo Seleziona casi: Campione casuale

È possibile selezionare una delle opzioni seguenti per la dimensione del campione:

- **Approssimativamente.** Una percentuale specificata dall'utente. Questa opzione genera un campione casuale che include approssimativamente la percentuale di casi specificata.

- **Esattamente.** Un numero di casi specificato dall'utente. È necessario specificare anche il numero di casi in base al quale generare il campione. Il secondo numero deve essere uguale o minore del numero totale di casi del file di dati. Se questo numero supera il numero totale di casi del file di dati, in proporzione il campione conterrà un numero di dati inferiore a quello specificato.

## Selezione di un intervallo di tempo o di casi

Per selezionare un intervallo di casi in base alla data, all'ora o al numero di osservazione (riga):

1. Selezionare **Basato su intervallo di tempo o di casi** e fare clic su **Intervallo** nella finestra di dialogo Seleziona casi.

Verrà aperta la finestra di dialogo Seleziona casi: intervallo, in cui è possibile selezionare un intervallo di numeri di osservazione (riga).

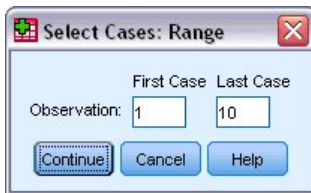


Figura 87. Finestre di dialogo Seleziona casi: intervallo

- **Primo caso.** Specificare i valori della data e/o dell'orario iniziale relativi all'intervallo. Se non sono definite variabili della data, specificare il numero dell'osservazione iniziale (numero di riga nell'Editor dei dati a meno che non sia attivo File suddiviso). Se non si specifica un valore per Ultimo caso, verranno selezionati tutti i casi a partire dalla data o dell'orario iniziale fino al termine della serie storica.
- **Ultimo caso.** Specificare i valori della data e/o dell'orario finale relativi all'intervallo. Se non sono definite variabili della data, specificare il numero dell'osservazione finale (numero di riga nell'Editor dei dati a meno che non sia attivo File suddiviso). Se non si specifica un valore per Primo caso, verranno selezionati tutti i casi a partire dall'inizio della serie storica fino alla data o all'orario finale.

Per i dati di serie storiche con variabili di data definite, è possibile selezionare un intervallo di date e/o orari basato sulle variabili di data definite. Ciascun caso rappresenta osservazioni eseguite in momenti diversi e l'ordinamento del file è cronologico.



Figura 88. Finestre di dialogo Seleziona casi: intervallo (serie storica)

Per generare variabili di data per i dati di serie storiche:

2. Dai menu, scegliere:  
**Dati > Definisci date...**

## Trattamento dei casi non selezionati

Per il trattamento dei casi non selezionati sono disponibili le opzioni seguenti:

- **Escludi casi non selezionati.** I casi non selezionati non saranno inclusi nell'analisi, ma rimarranno nel dataset. e sarà possibile utilizzarli successivamente nella sessione se si disattivano le funzioni di filtro.

Se si seleziona un campione casuale o si selezionano casi in base a un'espressione condizionale, verrà creata una variabile denominata *filtro\_\$* con un valore 1 per i casi selezionati e un valore 0 per i casi non selezionati.

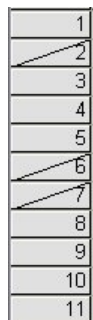
- **Copia casi selezionati in un nuovo dataset.** I casi selezionati vengono copiati in un nuovo dataset senza modificare quello originale. I casi non selezionati non vengono inclusi nel nuovo dataset e vengono lasciati nello stato originale nel dataset originale.
- **Elimina casi non selezionati.** I casi non selezionati verranno cancellati dal dataset. Per recuperare i casi cancellati è necessario uscire dal file senza salvare le modifiche e quindi riaprire il file. Se si salvano le modifiche al file di dati, l'eliminazione dei casi sarà definitiva.

*Nota:* se si eliminano i casi non selezionati e si salva il file, i casi non possono essere recuperati.

---

## Criteri di selezione dei casi

Se si è selezionato un sottoinsieme di casi senza eliminare i casi non selezionati, nell'Editor dei dati questi ultimi verranno contrassegnati da una linea diagonale sopra il numero di riga.



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

Figura 89. Criteri di selezione dei casi



---

## Capitolo 10. File di esempio

Il file di esempio installato con il prodotto si trova nella sottocartella *Samples* della cartella di installazione. È presente una cartella separata all'interno della sottocartella *Samples* per ciascuna delle seguenti lingue: inglese, francese, tedesco, italiano, giapponese, coreano, polacco, russo, cinese semplificato, spagnolo e cinese tradizionale.

Non tutti i file di esempio sono disponibili in tutte le lingue. Se un file di esempio non è disponibile in una lingua, la cartella di tale lingua contiene una versione inglese del file.

### Descrizioni

Questa sezione contiene brevi descrizioni dei file di esempio utilizzati negli esempi riportati in tutta la documentazione.

- **accidents.sav.** File di dati ipotetici che prende in esame una compagnia di assicurazioni impegnata nello studio dei fattori di rischio correlati all'età e al sesso per gli incidenti automobilistici che si verificano in una determinata regione. Ciascun caso corrisponde a una classificazione incrociata della categoria relativa età e del sesso.
- **adl.sav.** File di dati ipotetici che prende in esame l'impegno richiesto per determinare i vantaggi di un tipo di terapia proposto per i pazienti con problemi di cuore. I medici hanno assegnato in modo casuale i pazienti con problemi di cuore di sesso femminile a uno di due gruppi. Al primo gruppo è stata assegnata la terapia fisica standard; al secondo gruppo, un'ulteriore terapia di supporto psicologico. Dopo tre mesi di trattamenti, a ciascuna capacità dei pazienti che consente di riprendere le normali attività giornaliere è stato assegnato un punteggio come variabile ordinale.
- **advert.sav.** File di dati ipotetici che prende in esame l'impegno di un rivenditore al dettaglio che desidera esaminare la relazione tra il denaro speso per la pubblicità e le vendite risultanti. Finora sono stati raccolti i dati delle vendite precedenti e i relativi costi pubblicitari.
- **aflatoxin.sav.** File di dati ipotetici che prende in esame il test di raccolti di mais con presenza di Aflatossina, un veleno la cui concentrazione varia notevolmente nei raccolti. Una macchina per la lavorazione dei cereali ha ricevuto 16 campioni da ciascuno degli otto raccolti di mais e ha misurato i livelli di Aflatossina in parti per miliardo (PPB).
- **anorectic.sav.** Per trovare una sintomatologia standardizzata del comportamento anoressico/bulimico, i ricercatori <sup>1</sup> hanno condotto uno studio basato su 55 adolescenti affetti da disordini alimentari conosciuti. Ogni paziente è stato visitato quattro volte in quattro anni, per un totale di 220 visite. Durante ogni visita, ai pazienti sono stati assegnati punteggi per ciascuno dei 16 sintomi. I punteggi relativi ai sintomi sono assenti per il paziente 71 alla visita 2, il paziente 76 alla visita 2 e il paziente 47 alla visita 3, con 217 osservazioni valide.
- **bankloan.sav.** File di dati ipotetici che prende in esame l'impegno di una banca nel tentativo di ridurre il tasso di inadempienza nel rimborso di un prestito. Il file contiene informazioni finanziarie e demografiche su 850 vecchi e potenziali clienti. I primi 700 casi riguardano i clienti a cui sono stati concessi dei prestiti precedentemente. Gli ultimi 150 casi riguardano i potenziali clienti che la banca deve classificare come rischi di credito positivi o negativi.
- **bankloan\_binning.sav.** File di dati ipotetici che contiene informazioni finanziarie e demografiche su 5000 vecchi clienti.

---

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

- **behavior.sav.** In un classico esempio <sup>2</sup>, è stato chiesto a 52 studenti di classificare una combinazione di 15 situazioni e 15 comportamenti utilizzando una scala da 0="molto appropriato" a 9="molto inadeguato". I valori medi riferiti ai partecipanti sono stati considerati dissimilarità.
- **behavior\_ini.sav.** Questo file di dati contiene la configurazione iniziale di una soluzione a due dimensioni per *behavior.sav*.
- **brakes.sav.** File di dati ipotetici che prende in esame il controllo di qualità di un'industria che produce freni a disco per automobili con elevata prestazione. Il file di dati contiene le misurazioni del diametro di 16 dischi da ciascuna delle otto macchine di produzione. L'obiettivo finale è ottenere un diametro dei dischi pari a 322 millimetri.
- **breakfast.sav.** In uno studio classico <sup>3</sup>, è stato chiesto a 21 studenti MBA della Wharton School e ai loro consorti di classificare 15 cibi da colazione in ordine di preferenza, dove il valore 1 corrispondeva all'alimento preferito in assoluto e il valore 15 a quello meno preferito. Le loro preferenze sono state registrate per sei diversi scenari, che comprendevano tutti gli scenari compresi tra "Preferenza generale" e "Solo snack con bibita".
- **breakfast-overall.sav.** Questo file contiene le preferenze degli alimenti della colazione solo per il primo scenario, "Preferenza generale".
- **broadband\_1.sav.** File di dati ipotetici che contiene il numero di sottoscrittori, per area, di un provider di servizi a banda larga nazionale. Il file di dati contiene il numero dei sottoscrittori mensili di 85 aree in un periodo di quattro anni.
- **broadband\_2.sav.** Questo file è identico al file *broadband\_1.sav*, ma contiene i dati per ulteriori tre mesi.
- **car\_insurance\_claims.sav.** Un dataset presentato e analizzato altrove <sup>4</sup> riguarda le richieste di risarcimento auto. La quantità media di richieste di risarcimento può essere adattata come avente una distribuzione gamma, utilizzando una funzione di collegamento inverso per correlare la media della variabile dipendente a una combinazione lineare di età del contraente della polizza e tipo e anni del veicolo. Il numero delle richieste di risarcimento specificato può essere utilizzato come peso di scaling.
- **car\_sales.sav.** Questo file di dati ipotetici contiene le stime sulle vendite, i prezzi di listino e le specifiche fisiche di numerose marche e modelli di veicoli. I prezzi di listino e le specifiche fisiche sono state ottenute dal sito *edmunds.com* e dai siti dei produttori.
- **car\_sales\_uprepared.sav.** Questa è una versione modificata di *car\_sales.sav* che non include le versioni trasformate dei campi.
- **carpet.sav.** In un famoso esempio <sup>5</sup>, una società interessata a commercializzare un nuovo prodotto per la pulizia della moquette desidera esaminare l'influenza di cinque fattori sulla preferenza del consumatore (disegno del package, nome del marchio, prezzo, approvazione di un *ente riconosciuto* e garanzia soddisfatti o rimborsati). Esistono tre livelli di fattori per il disegno del package, che differiscono per l'ubicazione della spazzola dell'applicatore; tre marchi (*K2R*, *Glory* e *Bissell*); tre livelli di prezzo e due livelli (no o sì) per ciascuno degli ultimi due fattori. Dieci consumatori sono classificati in 22 profili definiti da questi fattori. La variabile *Preferenza* contiene come classificazione il rango medio per ogni profilo. Classificazioni basse corrispondono a una preferenza elevata. La variabile riflette una misura globale della preferenza per ogni profilo.
- **carpet\_prefs.sav.** Questo file di dati si basa sullo stesso esempio del file *carpet.sav*, ma contiene le classificazioni effettive raccolte da ciascuno dei 10 clienti. Ai clienti è stato chiesto di classificare 22 profili di prodotti in ordine di preferenza. Le variabili da *PREF1* a *PREF22* contengono gli ID dei profili associati, come definito nel file *carpet\_plan.sav*.
- **catalog.sav.** File di dati ipotetico che contiene le cifre sulle vendite mensili di tre prodotti venduti da una società di vendita per corrispondenza. Il file include anche i dati di cinque possibili variabili predittore.

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

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

4. P. McCullagh e J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

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



- **catalog\_seasfac.sav.** Questo file di dati è uguale al file *catalog.sav* con l'eccezione che contiene un insieme di fattori stagionali calcolati dalla procedura Decomposizionale stagionale insieme a variabili di dati.
- **cellular.sav.** File di dati ipotetici che prende in esame l'impegno di un'azienda di telefonia cellulare nel tentativo di ridurre il tasso di abbandono, ovvero l'abbandono dei clienti. Vengono applicati agli account dei punteggi di propensione all'abbandono, che vanno da 0 a 100. Gli account che totalizzano 50 o più potrebbero essere propensi a cambiare fornitore.
- **ceramics.sav.** File di dati ipotetici che prende in esame l'impegno di un produttore che desidera stabilire se una nuova lega premium ha una maggiore resistenza al calore rispetto alla lega standard. Ciascun caso rappresenta il test separato di una delle leghe. È indicata la temperatura massima alla quale può essere sottoposto il cuscinetto.
- **cereal.sav.** File di dati ipotetici che prende in esame le preferenze relative agli alimenti della colazione di un campione di 880 persone. Il file riporta anche l'età, il sesso e lo stato civile del campione e se le persone conducono uno stile di vita attivo (in base a un'attività sportiva con frequenza di due volte alla settimana). Ogni caso rappresenta un rispondente separato.
- **clothing\_defects.sav.** File di dati ipotetici che prende in esame il processo di controllo di qualità di un'industria di abbigliamento. Per ciascun lotto prodotto nella fabbrica, gli ispettori prelevano un campione di abiti per contare il numero dei capi che non sono accettabili per la vendita.
- **coffee.sav.** Questo file di dati contiene informazioni sulle immagini percepite di sei marche di caffè freddo<sup>6</sup>. Per ciascuno dei 23 attributi dell'immagine del caffè freddo, sono state selezionate tutte le marche descritte da tale attributo. Le sei marche sono indicate dalle sigle AA, BB, CC, DD, EE e FF per tutelare la confidenzialità dei dati.
- **contacts.sav.** File di dati ipotetici che prende in esame l'elenco dei contatti di un gruppo di rappresentanti di vendita di computer aziendali. Ciascun contatto è classificato in base al reparto della società in cui lavora e dalle relative categorie aziendali. Il file riporta anche l'importo dell'ultima vendita effettuata, il tempo trascorso dall'ultima vendita e le dimensioni della società del contatto.
- **creditpromo.sav.** File di dati ipotetici che prende in esame l'impegno di un grande magazzino nel tentativo di valutare l'efficacia di una recente promozione con carta di credito. A tale scopo, sono stati selezionati 500 titolari di carta in modo casuale. Alla metà di questi è stato inviato un annuncio promozionale che comunica la riduzione del tasso d'interesse nel caso di acquisti effettuati entro i tre mesi successivi. All'altra metà è stato inviato un annuncio stagionale standard.
- **customer\_dbase.sav.** File di dati ipotetici che prende in esame l'impegno di una società nel tentativo di utilizzare le informazioni contenute nel proprio database dei dati per creare offerte speciali per i clienti che più probabilmente risponderanno all'offerta. È stato selezionato in modo casuale un sottoinsieme della base dei clienti a cui è stata inviata l'offerta speciale e sono state registrate le risposte ricevute.
- **customer\_information.sav.** File di dati ipotetici contenente le informazioni postali del cliente, ad esempio il nome e l'indirizzo.
- **customer\_subset.sav.** Un sottoinsieme di 80 casi da *customer\_dbase.sav*.
- **debate.sav.** File di dati ipotetici che prende in esame le risposte appaiate a un'indagine da parte dei partecipanti a un dibattito politico prima e dopo il dibattito. Ogni caso rappresenta un rispondente separato.
- **debate\_aggregate.sav.** File di dati ipotetici che aggrega le risposte contenute nel file *debate.sav*. Ciascun caso corrisponde a una classificazione incrociata della preferenza prima e dopo il dibattito.
- **demo.sav.** File di dati ipotetici che prende in esame un database di clienti che hanno fatto acquisti al fine di inviare offerte mensili tramite il metodo del direct mailing. Viene registrata la risposta dei clienti, sia che abbiano aderito all'offerta o meno, insieme a diverse informazioni demografiche.
- **demo\_cs\_1.sav.** File di dati ipotetici che prende in esame la prima fase che una società intraprende per compilare un database con informazioni ricavate dalle indagini. Ogni caso rappresenta una diversa città. Sono registrate anche le informazioni sulla regione, provincia, distretto e città.

---

6. R. Kennedy, C. Riquier e 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.

- **demo\_cs\_2.sav.** File di dati ipotetici che prende in esame la seconda fase che una società intraprende per compilare un database con informazioni ricavate dalle indagini. Ogni caso rappresenta una diversa unità di abitazione, ricavata dalle città selezionate nella prima fase. Sono registrate anche le informazioni sulla regione, provincia, distretto, città, suddivisione e unità. Il file include inoltre informazioni sul campionamento ottenute dai primi due stadi del disegno.
- **demo\_cs.sav.** File di dati ipotetici che contiene informazioni sulle indagini raccolte utilizzando un disegno di campionamento complesso. Ogni caso rappresenta una diversa unità di abitazione. Sono registrate diverse informazioni demografiche e sul campionamento.
- **diabetes\_costs.sav.** Questo è un file di dati ipotetici che contiene informazioni gestite da una compagnia di assicurazioni su titolari di polizze che hanno il diabete. Ogni caso rappresenta un titolare di polizza differente.
- **dietstudy.sav.** File di dati ipotetici che contiene il risultato di uno studio ipotetico sulla dieta chiamato "Stillman diet" <sup>7</sup>. Ogni caso rappresenta un diverso soggetto e ne riporta il peso prima e dopo la dieta in libbre e i livelli dei trigliceridi in mg/100 ml.
- **dmdata.sav.** Questo è un file di dati ipotetici che contiene informazioni demografiche e relative all'acquisto per una società di direct marketing. *dmdata2.sav* contiene informazioni per un sottoinsieme di contatti che hanno ricevuto un mailing di test e *dmdata3.sav* contiene informazioni sui restanti contatti che non hanno ricevuto il mailing di test.
- **dvdplayer.sav.** File di dati ipotetici che prende in esame lo sviluppo di un nuovo lettore DVD. Utilizzando un prototipo, il personale addetto al marketing ha raccolto dati sui gruppi di interesse. Ogni caso rappresenta un diverso utente che è stato sottoposto all'indagine e include informazioni demografiche personali dell'utente e sulle risposte che ha fornito riguardo al prototipo.
- **german\_credit.sav.** Questo file di dati contiene informazioni ricavate dal dataset "German Credit" del Repository of Machine Learning Databases <sup>8</sup> presso la University of California, Irvine.
- **grocery\_1month.sav.** Questo file di dati ipotetici corrisponde al file di dati *grocery\_coupons.sav* con gli acquisti settimanali organizzati in modo che ogni caso corrisponda a un cliente separato. Alcune delle variabili che cambiano settimanalmente non vengono riportate nei risultati; l'importo speso registrato corrisponde ora alla somma degli importi spesi durante le quattro settimane dello studio.
- **grocery\_coupons.sav.** File di dati ipotetici che contiene i dati di indagine raccolti da una catena di drogherie interessata alle abitudini di acquisto dei suoi clienti. Ciascun cliente viene seguito per quattro settimane e ciascun caso corrisponde a una settimana per cliente con informazioni sul luogo degli acquisti e i tipi di acquisti, incluso l'importo speso nelle drogherie durante la settimana.
- **guttman.sav.** Bell <sup>9</sup> ha presentato una tabella per illustrare i possibili gruppi sociali. Guttman <sup>10</sup> ha utilizzato una parte di tale tabella, in cui cinque variabili che descrivono elementi come l'interazione sociale, i sentimenti di appartenenza a un gruppo, la prossimità fisica dei membri e il grado di formalità della relazione, sono state incrociate con cinque gruppi sociali teorici, compresi folla (ad esempio, le persone presenti a una partita di calcio), uditorio (ad esempio, di uno spettacolo teatrale o di una lezione universitaria), pubblico (ad esempio televisivo), calca (come una folla, ma con un'interazione molto maggiore), gruppi primari (intimi), gruppi secondari (volontari) e la comunità moderna (unione non stretta derivante da una prossimità fisica elevata e dall'esigenza di servizi specializzati).
- **health\_funding.sav.** File di dati ipotetici che contiene i dati sui fondi di assistenza sanitaria (importo per 100 persone), sui tassi di malattie (tasso per 10.000 persone) e sulle visite ai fornitori di assistenza sanitaria (tasso per 10.000 persone). Ogni caso rappresenta una diversa città.

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

8. C. L. Blake e C. J. Merz. 1998. "UCI Repository of machine learning databases." Disponibile all'indirizzo <http://www.ics.uci.edu/~mllearn/MLRepository.html>.

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

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

- **hivassay.sav.** File di dati ipotetici che prende in esame l'impegno di un'industria farmaceutica nel tentativo di sviluppare un'analisi che riesca a rilevare in tempi brevi l'infezione da virus HIV. I risultati dell'analisi sono otto sfumature di colore rosso sempre più intenso; le sfumature più intense indicano la maggiore probabilità di infezione. Una prova di laboratorio è stata condotta su 2000 campioni di sangue. La metà di questi è risultata infetta al virus HIV, l'altra metà non è risultata infetta.
- **hourlywagedata.sav.** File di dati ipotetici che prende in esame la paga oraria degli infermieri occupati presso uffici e ospedali e in base ai diversi livelli di esperienza.
- **insurance\_claims.sav.** Questo è un file di dati ipotetici che prende in esame una compagnia di assicurazioni che desidera creare un modello per segnalare richieste di risarcimento sospette e potenzialmente fraudolente. Ogni caso rappresenta una richiesta di risarcimento separata.
- **insure.sav.** File di dati ipotetici che prende in esame una compagnia di assicurazioni impegnata nello studio dei fattori di rischio, che indicano l'eventualità che un cliente presenti una domanda di indennizzo in un contratto assicurativo sulla vita della durata di dieci anni. Ogni caso nel file di dati rappresenta una coppia di contratti. In un contratto sono contenute informazioni su una richiesta di risarcimento, l'altro sull'età e sul sesso.
- **judges.sav.** File di dati ipotetici che prende in esame il punteggio assegnato, da giurie qualificate (più un appassionato) a 300 prestazioni sportive. Ciascuna riga rappresenta una diversa prestazione; i giudici hanno esaminato la stessa prestazione.
- **kinship\_dat.sav.** Rosenberg e Kim <sup>11</sup> si prefiggono di analizzare 15 termini indicanti parentela (zia, fratello, cugino, padre, nipote femmina di nonni, nonno, nonna, nipote maschio di nonni, madre, nipote maschio di zii), nipote femmina di zii, sorella, figlio, zio). Hanno richiesto a quattro gruppi di studenti universitari (due composti da femmine e due da maschi) di ordinare questi termini in base alla similarità. A due gruppi (uno femminile e uno maschile) è stato richiesto di effettuare l'ordinamento due volte, con il secondo ordinamento basato su un criterio diverso rispetto al primo. Pertanto, è stato ottenuto un totale di sei "origini". Ogni origine corrisponde ad una matrice di prossimità 15 x 15, le cui celle sono uguali al numero di persone in un'origine meno il numero di volte in cui gli oggetti sono stati partizionati insieme in quell'origine.
- **kinship\_ini.sav.** Questo file di dati contiene la configurazione iniziale di una soluzione a tre dimensioni per *kinship\_dat.sav*.
- **kinship\_var.sav.** Questo file di dati contiene variabili indipendenti relative a  *Sesso*,  *generazione* e  *grado* di separazione che possono essere utilizzate per interpretare le dimensioni di una soluzione per *kinship\_dat.sav*. In modo specifico, tali variabili possono essere utilizzate per limitare lo spazio della soluzione a una combinazione lineare di tali variabili.
- **marketvalues.sav.** Questo file di dati è relativo alle vendite immobiliari in un nuovo complesso residenziale in Algonquin, Ill., durante gli anni 1999–2000. Queste vendite sono una questione di record pubblico.
- **nhis2000\_subset.sav.** Il National Health Interview Survey (NHIS) è un'indagine di grandi dimensioni condotta sulla popolazione civile americana. Le interviste vengono realizzate di persona e si basano su un campione rappresentativo di famiglie a livello nazionale. Per ogni membro di una famiglia vengono raccolte osservazioni e informazioni di carattere demografico relative allo stato di salute. Questo file di dati contiene un sottoinsieme delle informazioni ottenute dall'indagine del 2000. National Center for Health Statistics. National Health Interview Survey, 2000. File di dati e documentazione di utilizzo pubblico. [ftp://ftp.cdc.gov/pub/Health\\_Statistics/NCHS/Datasets/NHIS/2000/](ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/NHIS/2000/). Accesso 2003.
- **ozone.sav.** I dati includono 330 osservazioni basate su sei variabili meteorologiche per quantificare la concentrazione dell'ozono dalle variabili rimanenti. I precedenti ricercatori, <sup>12</sup> e <sup>13</sup>, hanno rilevato non linearità tra queste variabili, che impediscono un approccio di regressione standard.

---

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

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

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

- **pain\_medication.sav.** File di dati ipotetici che contiene i risultati di un test clinico per stabilire la cura antinfiammatoria per il trattamento del dolore generato dall'artrite cronica. Di particolare interesse, il test ha evidenziato il tempo che impiega il farmaco ad avere effetto e il confronto con altri farmaci esistenti.
- **patient\_los.sav.** File di dati ipotetici che contiene informazioni sul trattamento dei pazienti ricoverati per sospetto di infarto del miocardio. Ogni caso corrisponde a un diverso paziente e contiene diverse variabili correlate alla degenza nell'ospedale.
- **patlos\_sample.sav.** File di dati ipotetici che contiene informazioni sul trattamento di un campione di pazienti curato con trombolitici durante la degenza per infarto del miocardio. Ogni caso corrisponde a un diverso paziente e contiene diverse variabili correlate alla degenza nell'ospedale.
- **poll\_cs.sav.** File di dati ipotetici che prende in esame i sondaggi per stabilire il livello di sostegno pubblico nei confronti di un disegno di legge prima che diventi una legge vera e propria. I casi corrispondono ai votanti registrati. Ciascun caso riporta informazioni sulla contea, sul comune e sul quartiere in cui vive il votante.
- **poll\_cs\_sample.sav.** File di dati ipotetici che contiene un campione dei votanti elencati nel file *poll\_cs.sav*. Il campione è stato prelevato in base al disegno specificato nel file di piano *poll\_csplan* e questo file di dati contiene le probabilità di inclusione e i pesi del campione. Tuttavia, notare che poiché fa uso del metodo PPS (probability-proportional-to-size, probabilità proporzionale alla dimensione), esiste anche un file contenente le probabilità di selezione congiunte (*poll\_jointprob.sav*). Le ulteriori variabili corrispondenti ai dati demografici dei votanti e alla loro opinione sul disegno di legge, sono state raccolte e aggiunte al file di dati dopo aver acquisito il campione.
- **property\_assess.sav.** File di dati ipotetici che prende in esame l'impegno di un perito di una contea nel tentativo di mantenere gli accertamenti sui valori delle proprietà aggiornati in base alle risorse limitate. I casi rappresentano le proprietà vendute nella contea nello scorso anno. Ogni caso nel file di dati contiene informazioni sul comune in cui si trova la proprietà, il perito che per ultimo ha visitato la proprietà, il tempo trascorso dall'accertamento, la valutazione fatta in tale momento e il valore di vendita della proprietà.
- **property\_assess\_cs.sav.** File di dati ipotetici che prende in esame l'impegno di un perito di uno stato nel tentativo di mantenere aggiornati gli accertamenti sui valori delle proprietà in base alle risorse limitate. I casi corrispondono alle proprietà nello stato. Ogni caso nel file di dati include informazioni sulla contea, il comune e il quartiere in cui risiede la proprietà, la data dell'ultimo accertamento e la valutazione fatta in tale data.
- **property\_assess\_cs\_sample.sav.** File di dati ipotetici che contiene un campione delle proprietà elencate nel file *property\_assess\_cs.sav*. Il campione è stato prelevato in base al disegno specificato nel file di piano *property\_assess\_csplan* e questo file di dati contiene le probabilità di inclusione e i pesi del campione. L'ulteriore variabile *Valore corrente* è stata raccolta e aggiunta al file di dati dopo aver acquisito il campione.
- **recidivism.sav.** File di dati ipotetici che prende in esame l'impegno delle Forze dell'Ordine nel tentativo di valutare il tasso di recidività nella propria area di giurisdizione. Ogni caso corrisponde a un precedente trasgressore e include le informazioni demografiche, alcuni dettagli sul primo crimine, il tempo trascorso fino al secondo arresto e se tale arresto è avvenuto entro due anni dal primo.
- **recidivism\_cs\_sample.sav.** File di dati ipotetici che prende in esame l'impegno delle Forze dell'Ordine nel tentativo di valutare il tasso di recidività nella propria area di giurisdizione. Ogni caso corrisponde a un precedente trasgressore, rilasciato dopo il primo arresto durante il mese di giugno 2003, e include le informazioni demografiche, alcuni dettagli sul primo crimine e i dati del secondo arresto, se si è verificato entro la fine di giugno 2006. I trasgressori sono stati selezionati da dipartimenti campione in base al piano di campionamento specificato in *recidivism\_cs\_csplan*; poiché fa uso del metodo PPS (probability-proportional-to-size, probabilità proporzionale alla dimensione), esiste anche un file contenente le probabilità di selezione congiunte (*recidivism\_cs\_jointprob.sav*).
- **rfm\_transactions.sav.** File di dati ipotetici contenente i dati della transazione di acquisto, inclusa la data di acquisto, gli articoli acquistati e il valore monetario di ciascuna transazione.
- **salesperformance.sav.** File di dati ipotetici che prende in esame la valutazione di due nuovi corsi di formazione alle vendite. Sessanta dipendenti, divisi in tre gruppi, ricevono tutti la formazione

standard. In più, al gruppo 2 viene assegnato un corso di formazione tecnica e al gruppo 3 un'esercitazione pratica. Al termine del corso di formazione, ciascun dipendente viene sottoposto a un esame e il punteggio conseguito viene registrato. Ciascun caso nel file di dati rappresenta un diverso partecipante. Il file di dati include il gruppo a cui è assegnato il partecipante e il punteggio conseguito all'esame finale.

- **satisf.sav.** File di dati ipotetici che prende in esame un'indagine sulla soddisfazione dei clienti condotta da una società di vendita al dettaglio presso 4 negozi. Sono stati intervistati 582 clienti e ciascun caso rappresenta le risposte ottenute da un singolo cliente.
- **screws.sav.** Questo file di dati contiene informazioni sulle caratteristiche di viti, bulloni, dadi e puntine<sup>14</sup>.
- **shampoo\_ph.sav.** File di dati ipotetici che prende in esame il processo di controllo di qualità di un'industria di prodotti per capelli. A intervalli di tempo regolari, vengono misurati sei diversi lotti prodotti e ne viene registrato il relativo pH. L'intervallo di obiettivo è 4.5–5.5.
- **ships.sav.** Ad esempio, un dataset presentato e analizzato altrove<sup>15</sup> riguarda i danni subiti dalle navi da carico a causa delle onde. I conteggi degli incidenti possono essere presentati con un tasso di Poisson in base al tipo di nave, al periodo di costruzione e al periodo di servizio. I mesi di servizio aggregati di ciascuna cella della tabella generata dalla classificazione incrociata dei fattori fornisce i valori di esposizione al rischio.
- **site.sav.** File di dati ipotetici che prende in esame l'impegno di una società nella scelta di nuovi siti in cui espandere la propria presenza. La società ha incaricato due consulenti separati che, oltre a valutare i siti e presentare un report completo, devono classificarli come potenzialmente "molto adatti", "adatti" o "poco adatti".
- **smokers.sav.** Questo file di dati è un estratto del 1998 National Household Survey of Drug Abuse e rappresenta un campione probabile di famiglie americane. (<http://dx.doi.org/10.3886/ICPSR02934>) Pertanto, il primo passo nell'analisi di questo file di dati deve essere soppesare i dati per rispecchiare le tendenze della popolazione.
- **stocks.sav** Questo file di dati ipotetici contiene i prezzi delle azioni e il volume per un anno.
- **stroke\_clean.sav.** Questo file di dati ipotetici contiene lo stato di un database medico dopo averne eseguito la pulizia utilizzando le procedure nell'opzione Data Preparation.
- **stroke\_invalid.sav.** Questo file di dati ipotetici contiene riporta lo stato iniziale di un database medico e contiene numerosi errori di immissione dati.
- **stroke\_survival.** Questo file di dati ipotetici riguarda i tempi di sopravvivenza per i pazienti che, dopo avere completato un programma riabilitativo in seguito a un ictus postischemico, affrontano alcune sfide. Dopo l'attacco, viene annotata l'occorrenza dell'infarto miocardico, dell'ictus ischemico o emorragico e viene registrata l'ora dell'evento. Questo campione viene troncato a sinistra perché include solo i pazienti che sono sopravvissuti fino al termine del programma riabilitativo post-ictus.
- **stroke\_valid.sav.** File di dati ipotetici che riporta lo stato di un database medico dopo il controllo dei valori eseguito con la procedura Convalida i dati. Il database contiene comunque casi potenzialmente anomali.
- **survey\_sample.sav.** File di dati che contiene i dati dell'indagine, compresi i dati demografici e varie misure dell'atteggiamento. Si basa su un sottoinsieme di variabili tratte dal 1998 NORC General Social Survey, benché alcuni valori dei dati siano stati modificati e siano state aggiunte variabili dummy a scopo dimostrativo.
- **tcm\_kpi.sav.** Questo è un file di dati ipotetici che contiene valori di indicatori di prestazioni chiave settimanali per un business. Contiene anche i dati settimanali per un numero di metriche controllabili durante lo stesso periodo di tempo.
- **tcm\_kpi\_upd.sav.** Questo file di dati è identico al file *tcm\_kpi.sav*, ma contiene i dati per quattro settimane extra.

---

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

15. P. McCullagh e J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.

- **telco.sav.** File di dati ipotetici che prende in esame l'impegno di un'azienda di telecomunicazioni nel tentativo di ridurre il tasso di abbandono, ovvero l'abbandono dei propri clienti. Ciascun caso rappresenta un cliente separato e riporta diverse informazioni demografiche e sull'uso del servizio.
- **telco\_extra.sav.** Questo file di dati è simile al file *telco.sav*, ma le variabili "tenure" e spesa del cliente trasformata tramite logaritmo sono state sostituite dalle variabili di spesa del cliente trasformata tramite logaritmo standardizzate.
- **telco\_missing.sav.** Questo file di dati è un sottoinsieme del file di dati *telco.sav*, ma alcuni dei valori dei dati demografici sono stati sostituiti con valori mancanti.
- **testmarket.sav.** File di dati ipotetici che prende in esame i piani di una catena di fast food per aggiungere un nuovo prodotto al proprio menu. Sono previste tre campagne promozionali del nuovo prodotto. Il prodotto viene introdotto in diversi mercati selezionati in modo casuale. Per ogni sede viene utilizzata una promozione differente registrando le vendite settimanali della nuova voce per le prime quattro settimane. Ogni caso rappresenta un luogo e una settimana diversi.
- **testmarket\_1month.sav.** Questo file di dati ipotetici corrisponde al file *testmarket.sav* con le vendite settimanali organizzate in modo che ogni caso corrisponda a un luogo separato. Alcune delle variabili che cambiano settimanalmente non vengono riportate nei risultati; le vendite registrate corrispondono ora alla somma delle vendite conseguite durante le quattro settimane dello studio.
- **tree\_car.sav.** File di dati ipotetici che contiene dati demografici e sul prezzo di acquisto dei veicoli.
- **tree\_credit.sav.** File di dati ipotetici che contiene dati demografici e sulla cronologia dei mutui di una banca.
- **tree\_missing\_data.sav.** File di dati ipotetici che contiene dati demografici e sulla cronologia dei mutui di una banca con un numero elevato di valori mancanti.
- **tree\_score\_car.sav.** File di dati ipotetici che contiene dati demografici e sul prezzo di acquisto dei veicoli.
- **tree\_textdata.sav.** File di dati semplice con due variabili destinato principalmente per mostrare lo stato predefinito delle variabili prima dell'assegnazione dei livelli di misurazione e delle etichette valore.
- **tv-survey.sav.** File di dati ipotetici che prende in esame un'indagine condotta da una emittente televisiva che deve stabilire se estendere la durata di un programma di successo. A un campione di 906 rispondenti è stato chiesto se preferisce guardare il programma con diverse condizioni. Ciascuna riga rappresenta un diverso rispondente e ciascuna colonna una diversa condizione.
- **ulcer\_recurrence.sav.** Questo file contiene informazioni parziali su uno studio svolto per mettere a confronto l'efficacia di due terapie preventive per la recidiva delle ulcere. Fornisce un ottimo esempio di dati acquisiti a intervalli ed è stato presentato e analizzato in altri luoghi <sup>16</sup>.
- **ulcer\_recurrence\_recoded.sav.** In questo file sono contenute le informazioni del file *ulcer\_recurrence.sav* riorganizzate per consentire di presentare la probabilità degli eventi per ciascun intervallo dello studio, anziché solo al termine. È stato presentato e analizzato in altri luoghi <sup>17</sup>.
- **verd1985.sav.** Questo file di dati prende in esame un'indagine <sup>18</sup>. Sono state registrate le risposte di quindici soggetti a otto variabili. Le variabili di interesse sono suddivise in tre insiemi. L'insieme 1 include *età* e *statociv*, l'insieme 2 include *andom* e *giornale* e l'insieme 3 include *musica* e *vicinato*. *Andom* viene scalata come nominale multipla ed *età* come ordinale; tutte le altre variabili vengono scalate come nominali singole.
- **virus.sav.** File di dati ipotetici che prende in esame l'impegno di un ISP (Internet Service Provider) nel tentativo di determinare gli effetti che un virus può generare nelle sue reti. Si è tenuta traccia della percentuale (approssimativa) di traffico e-mail infettato da virus sulla rete in un lasso di tempo, dal momento dell'individuazione fino alla soppressione della minaccia.

16. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

17. Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.

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

- **wheeze\_steubenville.sav.** Questo file è un sottoinsieme di uno studio longitudinale degli effetti che l'inquinamento provoca sulla salute dei bambini <sup>19</sup>. I dati contengono misure binarie ripetute del livello di asma dei bambini della città di Steubenville, Ohio, di 7, 8, 9 e 10 anni. I dati indicano anche se la mamma dei bambini era fumatrice durante il primo anno dello studio.
- **workprog.sav.** File di dati ipotetici che prende in esame un programma di lavoro governativo il cui obiettivo è fornire attività più adatte alle persone diversamente abili. È stato seguito un campione di potenziali partecipanti al programma, alcuni dei quali sono stati selezionati in modo casuale e altri no. Ogni caso rappresenta un diverso partecipante al programma.
- **worldsales.sav** Questo file di dati ipotetici contiene i ricavi delle vendite per continente e prodotto.

---

19. Ware, J. H., D. W. Dockery, A. Spiro III, F. E. Speizer e 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.





---

## Informazioni particolari

Queste informazioni sono state sviluppate per prodotti e servizi offerti negli Stati Uniti. Questo materiale potrebbe essere disponibile da IBM in altre lingue. Tuttavia, può essere necessario possedere una copia del prodotto o versione del prodotto in tale lingua per potervi accedere.

IBM può non offrire i prodotti, i servizi o le funzioni presentati in questo documento in altri paesi. Consultare il proprio rappresentante locale IBM per informazioni sui prodotti ed i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento ad un prodotto, programma o servizio IBM non implica o intende dichiarare che solo quel prodotto, programma o servizio IBM può essere utilizzato. Qualsiasi prodotto funzionalmente equivalente al prodotto, programma o servizio che non violi alcun diritto di proprietà intellettuale IBM può essere utilizzato. È comunque responsabilità dell'utente valutare e verificare la possibilità di utilizzare altri prodotti, programmi o servizi non IBM.

IBM può avere applicazioni di brevetti o brevetti in corso relativi all'argomento descritto in questo documento. La fornitura del presente documento non concede alcuna licenza a tali brevetti. È possibile inviare per iscritto richieste di licenze a:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

Per richieste di licenze relative ad informazioni double-byte (DBCS), contattare il Dipartimento di Proprietà Intellettuale IBM nel proprio paese o inviare richieste per iscritto a:

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

IBM (INTERNATIONAL BUSINESS MACHINES CORPORATION) FORNISCE LA PRESENTE PUBBLICAZIONE "NELLO STATO IN CUI SI TROVA" SENZA GARANZIE DI ALCUN TIPO, ESPRESSE O IMPLICITE, IVI INCLUSE, A TITOLO DI ESEMPIO, GARANZIE IMPLICITE DI NON VIOLAZIONE, DI COMMERCIALIZZABILITÀ E DI IDONEITÀ PER UNO SCOPO PARTICOLARE. Alcune giurisdizioni non consentono la rinuncia ad alcune garanzie espresse o implicite in determinate transazioni, pertanto, la presente dichiarazione può non essere applicabile.

Queste informazioni potrebbero contenere imprecisioni tecniche o errori tipografici. Le modifiche alle presenti informazioni vengono effettuate periodicamente; tali modifiche saranno incorporate nelle nuove pubblicazioni della pubblicazione. IBM si riserva il diritto di apportare miglioramenti e/o modifiche al prodotto o al programma descritto nel manuale in qualsiasi momento e senza preavviso.

I riferimenti in queste informazioni a siti Web non IBM vengono forniti solo per comodità e non implicano in alcun modo l'approvazione di tali siti web. I materiali presenti in tali siti web non sono parte dei materiali per questo prodotto IBM e l'utilizzo di tali siti è a proprio rischio.

IBM può utilizzare o distribuire qualsiasi informazione fornita in qualsiasi modo ritenga appropriato senza incorrere in alcun obbligo verso l'utente.

Coloro che detengano la licenza su questo programma e desiderano avere informazioni su di esso allo scopo di consentire: (i) uno scambio di informazioni tra programmi indipendenti ed altri (compreso questo) e (ii) l'utilizzo reciproco di tali informazioni, dovrebbe rivolgersi a:

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119  
Armonk, NY 10504-1785  
US*

Tali informazioni possono essere disponibili, in base ad appropriate clausole e condizioni, includendo in alcuni casi, il pagamento di una tassa.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale concesso in licenza disponibile sono forniti da IBM in base alle clausole dell'Accordo per Clienti IBM (IBM Customer Agreement), dell'IBM IPLA (IBM International Program License Agreement) o qualsiasi altro accordo equivalente tra le parti.

Gli esempi client e dati delle prestazioni citati sono presentati esclusivamente a scopo illustrativo. Gli effettivi risultati delle prestazioni possono variare in base alle configurazioni e alle condizioni operative specifiche.

Le informazioni relative a prodotti non IBM sono ottenute dai fornitori di quei prodotti, dagli annunci pubblicati e da altre fonti disponibili al pubblico. IBM non ha testato quei prodotti e non può confermarne la precisione della prestazione, la compatibilità o qualsiasi altro reclamo relativo ai prodotti non IBM. Le domande sulle funzionalità dei prodotti non IBM devono essere indirizzate ai fornitori di tali prodotti.

Le dichiarazioni relative all'orientamento o alle intenzioni future di IBM sono soggette a modifica o a ritiro senza preavviso e rappresentano solo mete e obiettivi.

Questa pubblicazione contiene esempi di dati e prospetti utilizzati quotidianamente nelle operazioni aziendali. Pertanto, per maggiore completezza, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti i nomi contenuti nel manuale sono fittizi e ogni riferimento a persone o aziende reali è puramente casuale.

#### LICENZA DI COPYRIGHT:

Queste informazioni contengono programmi applicativi di esempio in linguaggio sorgente, che illustrano tecniche di programmazione su varie piattaforme operative. È possibile copiare, modificare e distribuire questi programmi di esempio sotto qualsiasi forma senza alcun pagamento a IBM, allo scopo di sviluppare, utilizzare, commercializzare o distribuire i programmi applicativi in conformità alle API (application programming interface) a seconda della piattaforma operativa per cui i programmi di esempio sono stati scritti. Questi esempi non sono stati testati approfonditamente tenendo conto di tutte le condizioni possibili. IBM, quindi, non può garantire o sottintendere l'affidabilità, l'utilità o il funzionamento di questi programmi. I programmi di esempio sono forniti "NELLO STATO IN CUI SI TROVANO", senza alcun tipo di garanzia. IBM non intende essere responsabile per alcun danno derivante dall'uso dei programmi di esempio.

Ogni copia o qualsiasi parte di questi programmi di esempio o qualsiasi lavoro derivato, devono contenere le seguenti informazioni relative alle leggi sul diritto d'autore:

© (nome della società) (anno). Parti di questo codice derivano dai Programmi di Esempio di IBM Corp.

© Copyright IBM Corp. \_immettere l'anno o gli anni\_. Tutti i diritti riservati.

---

## Marchi

IBM, il logo IBM e [ibm.com](http://ibm.com) sono marchi di International Business Machines Corp., registrati in molte giurisdizioni del mondo. Altri nomi di prodotti e servizi possono essere marchi di IBM o di altre società. Un elenco corrente dei marchi IBM è disponibile sul web in "Copyright and trademark information" all'indirizzo [www.ibm.com/legal/copytrade.shtml](http://www.ibm.com/legal/copytrade.shtml).

Adobe, il logo Adobe, PostScript e il logo PostScript sono marchi o marchi registrati di Adobe Systems Incorporated negli Stati Uniti e/o in altri paesi.

Intel, il logo Intel, Intel Inside, il logo Intel Inside, Intel Centrino, il logo Intel Centrino, Celeron, Intel Xeon, Intel SpeedStep, Itanium e Pentium sono marchi o marchi registrati di Intel Corporation o delle sue consociate negli Stati Uniti e in altri paesi.

Linux è un marchio registrato di Linus Torvalds negli Stati Uniti e/o negli altri paesi.

Microsoft, Windows, Windows NT, e il logo Windows sono marchi di Microsoft Corporation negli Stati Uniti e/o in altri paesi.

UNIX è un marchio della The Open Group negli Stati Uniti e/o negli altri paesi.

Java e tutti i marchi e i logo basati su Java sono marchi o marchi registrati di Oracle e/o delle sue associate.



---

## Indice analitico

### A

Access (Microsoft) 11

### C

calcolo di nuove variabili 69  
casi  
    ordinamento 75, 77  
    selezione 77  
conteggi  
    tabelle di conteggi 27  
creare etichette di variabile 21

### D

dati categoriali 27  
    misure di riepilogo 27  
dati continui 27  
dati di intervallo 27  
dati di rapporto 27  
dati di scala 27  
dati di stringa  
    inserimento di dati 20  
dati nominali 27  
dati numerici 19  
dati ordinali 27  
dati qualitativi 27  
dati quantitativi 27

### E

Editor dei dati  
    inserimento di dati non numerici 20  
    inserimento di dati numerici 19  
elaborazione file suddiviso 76  
esportazione dei risultati  
    HTML 61  
esportazione risultati  
    a Excel 54  
    a Word 54  
    in PowerPoint 54  
espressioni condizionali 71  
etichette di variabile  
    creazione 21  
etichette valori  
    assegnazione 23  
    controllo visualizzazione nel  
        Visualizzatore 23  
    variabili numeriche 23  
Excel (Microsoft)  
    esportazione dei risultati 54

### F

file di database  
    lettura 11  
file di dati testuali  
    lettura 14

file di esempio  
    ubicazione 83  
file di Excel  
    lettura 8  
file di fogli elettronici  
    lettura 8  
    lettura dei nomi di variabili 8  
file di sintassi  
    apertura 65  
finestre della sintassi  
    autocompletamento 64  
    codifica colori 64  
esecuzione di comandi 63, 65  
incollare i comandi 63  
modifica di comandi 64  
punti di interruzione 65  
funzioni nelle espressioni 70

### G

grafici  
    barra 28, 33  
    creazione di grafici 33  
    istogrammi 30  
grafici a barre 28

### H

HTML  
    esportazione dei risultati 61

### I

incollare la sintassi  
    da una finestra di dialogo 63  
inserimento di dati 19, 20  
    non numerici 20  
    numerico 19  
istogrammi 30

### L

livello di misurazione 27

### M

misure di riepilogo  
    dati categoriali 27  
    variabili di scala 29  
modifica di tabelle pivot 43

### N

nascondere righe e colonne in tabelle  
    pivot 44

### O

ordinamento dei casi 75

### P

PowerPoint (Microsoft)  
    esportazione dei risultati 54  
Procedura guidata Data e ora 72  
Procedura guidata di database 11  
Procedura guidata di importazione di  
    testo 14

### R

ricodifica di valori 67

### S

selezione di casi 77  
sintassi 63  
sottoinsiemi di casi  
    basati su date e orari 80  
    campione casuale 79  
    eliminazione di casi non  
        selezionati 80  
    espressioni condizionali 78  
    filtro di casi non selezionati 80  
    se la condizione è soddisfatta 78  
    selezione 77  
spostamento  
    elementi del Visualizzatore 39  
    elementi delle tabelle pivot 41  
strati  
    creazione in tabelle pivot 42  
strumento Aiuto sintassi 64

### T

tabella delle frequenze 27  
tabelle pivot  
    accesso alle definizioni 40  
    barre pivot 41  
    formati di celle 44  
    formattazione 43  
    modifica 43  
    nascondere righe e colonne 44  
    nascondere segni decimali 44  
    strati 42  
    tipi di dati cella 44  
    trasposizione di righe e colonne 41  
tipi di dati  
    per variabili 22  
trasposizione (capovolgimento) di righe e  
    colonne nelle tabelle pivot 41

## V

- valori mancanti
  - mancanti di sistema 23
  - per variabili non numeriche 24
  - per variabili numeriche 24
- valori mancanti di sistema 23
- variabili 19
  - etichette 21
  - tipi di dati 22
- variabili di data e ora 72
- variabili di scala
  - misure di riepilogo 29
- Visualizzatore
  - nascondere e visualizzare l'output 39
  - spostamento dell'output 39

## W

- Word (Microsoft)
  - esportazione dei risultati 54





Stampato in Italia