

# IBM SPSS Direct Marketing 19



*Note:* Before using this information and the product it supports, read the general information under Notices a pag. 111.

This document contains proprietary information of SPSS Inc, an IBM Company. It is provided under a license agreement and is protected by copyright law. The information contained in this publication does not include any product warranties, and any statements provided in this manual should not be interpreted as such.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

© **Copyright SPSS Inc. 1989, 2010.**

---

# Prefazione

IBM® SPSS® Statistics è un sistema completo per l'analisi dei dati. Il modulo aggiuntivo opzionale Direct Marketing include le tecniche di analisi aggiuntive descritte nel presente manuale. Il modulo aggiuntivo Direct Marketing deve essere usato con il modulo Core SPSS Statistics in cui è completamente integrato.

## **Informazioni su SPSS Inc., una società del gruppo IBM**

SPSS Inc., una società del gruppo IBM, è fornitore leader mondiale nel settore del software e delle soluzioni per l'analisi predittiva. L'offerta completa dei prodotti dell'azienda (raccolta di dati, statistica, modellazione e distribuzione) consente di acquisire i comportamenti e le opinioni delle persone, prevedere i risultati delle future interazioni con i clienti ed elaborare questi dati integrando le analitiche nelle procedure aziendali. Le soluzioni SPSS Inc. consentono la gestione di attività interconnesse all'interno dell'intera organizzazione, con particolare attenzione alla convergenza di analitiche, architettura IT e procedure aziendali. Clienti commerciali, istituzionali e accademici di tutto il mondo si affidano alla tecnologia SPSS Inc. ottenendo un vantaggio competitivo in termini di attrazione, mantenimento e ampliamento della base clienti, riducendo al contempo frodi e rischi. SPSS Inc. è stata acquisita da IBM nell'ottobre 2009. Per ulteriori informazioni, visitare il sito <http://www.spss.com>.

## **Supporto tecnico**

Ai clienti che richiedono la manutenzione, viene messo a disposizione un servizio di supporto tecnico. I clienti possono contattare il supporto tecnico per richiedere assistenza per l'utilizzo dei prodotti SPSS Inc. o per l'installazione di uno degli ambienti hardware supportati. Per il supporto tecnico, visitare il sito Web di SPSS Inc. all'indirizzo <http://support.spss.com> o contattare la filiale del proprio paese indicata nel sito Web all'indirizzo <http://support.spss.com/default.asp?refpage=contactus.asp>. Ricordare che durante la richiesta di assistenza sarà necessario fornire i dati di identificazione personali, i dati relativi alla propria società e il numero del contratto di manutenzione.

## **Servizio clienti**

Per informazioni sulla spedizione o sul proprio account, contattare la filiale nel proprio paese, indicata nel sito Web all'indirizzo <http://www.spss.com/worldwide>. Tenere presente che sarà necessario fornire il numero di serie.

## ***Corsi di formazione***

SPSS Inc. organizza corsi di formazione pubblici e onsite che includono esercitazioni pratiche. Tali corsi si terranno periodicamente nelle principali città. Per ulteriori informazioni sui corsi, contattare la filiale nel proprio paese, indicata nel sito Web all'indirizzo <http://www.spss.com/worldwide>.

## ***Pubblicazioni aggiuntive***

I documenti *SPSS Statistics: Guide to Data Analysis*, *SPSS Statistics: Statistical Procedures Companion* e *SPSS Statistics: Advanced Statistical Procedures Companion*, scritti da Marija Norušis e pubblicati da Prentice Hall sono disponibili come materiale supplementare consigliato. Queste pubblicazioni descrivono le procedure statistiche nei moduli SPSS Statistics Base, Advanced Statistics e Regression. Utili sia come guida iniziale all'analisi dei dati che per applicazioni avanzate, questi manuali consentono di ottimizzare l'utilizzo delle funzionalità presenti nell'offerta IBM® SPSS® Statistics. Per ulteriori informazioni, inclusi contenuti delle pubblicazioni e capitoli di esempio, visitare il sito Web dell'autrice: <http://www.norusis.com>

---

# Contenuto

## **Parte I: Manuale dell'utente**

<b>1</b>	<b>Direct marketing</b>	<b>1</b>
<b>2</b>	<b>Analisi RFM</b>	<b>2</b>
	Punteggi RFM dai dati della transazione . . . . .	3
	Punteggi RFM dai dati dei clienti. . . . .	5
	Categorizzazione RFM . . . . .	6
	Salvataggio dei punteggi RFM dai dati della transazione. . . . .	9
	Salvataggio dei punteggi RFM dai dati dei clienti . . . . .	11
	Output RFM . . . . .	13
<b>3</b>	<b>Analisi cluster</b>	<b>15</b>
	Impostazioni. . . . .	18
<b>4</b>	<b>Profili dei potenziali clienti</b>	<b>20</b>
	Impostazioni. . . . .	24
	Creazione di un campo Risposta categoriale. . . . .	24
<b>5</b>	<b>Tassi di risposta codice postale</b>	<b>26</b>
	Impostazioni. . . . .	31
	Creazione di un campo Risposta categoriale. . . . .	33

**6 Propensione all'acquisto 34**

Impostazioni ..... 39  
Creazione di un campo Risposta categoriale..... 41

**7 Test pacchetto di controllo 42**

**Parte II: Esempi**

**8 Analisi RFM dei dati delle transazioni 46**

Dati della transazione ..... 46  
Esecuzione dell'analisi ..... 46  
Valutazione dei risultati..... 48  
Unire i dati sui punteggi ai dati sui clienti ..... 50

**9 Analisi cluster 53**

Esecuzione dell'analisi ..... 53  
Output ..... 55  
Selezione dei record in base ai cluster ..... 63  
    Creazione di un filtro nel Viewer dei modelli di cluster ..... 64  
    Selezione dei record in base ai valori del campo Cluster ..... 66  
Riepilogo ..... 69

**10 Profili dei potenziali clienti 70**

Considerazioni sui dati ..... 70  
Esecuzione dell'analisi ..... 70  
Output ..... 73  
Riepilogo ..... 76

**11 Tassi di risposta codice postale 77**

Considerazioni sui dati ..... 77  
Esecuzione dell'analisi ..... 77  
Output ..... 80  
Riepilogo ..... 83

**12 Propensione all'acquisto 85**

Considerazioni sui dati ..... 85  
Costruzione di un modello predittivo ..... 85  
    Valutazione del modello ..... 89  
Applicazione del modello ..... 90  
Riepilogo ..... 96

**13 Test pacchetto di controllo 97**

Esecuzione dell'analisi ..... 97  
Output ..... 99  
Riepilogo ..... 99

**Appendici**

**A File di esempio 100**

**B Notices 111**

**Indice 113**





***Parte I:***  
***Manuale dell'utente***



# ***Direct marketing***

Il modulo Direct Marketing offre un insieme di strumenti destinati a migliorare i risultati delle campagne di direct marketing, identificando le caratteristiche demografiche, di acquisto e di altro tipo che definiscono i vari gruppi di consumatori e individuando come obiettivo gruppi specifici allo scopo di massimizzare i tassi di risposta positiva.

**Analisi RFM.** Questa tecnica identifica i clienti esistenti che più probabilmente risponderanno a una nuova offerta. [Per ulteriori informazioni, vedere l'argomento Analisi RFM in il capitolo 2 a pag. 2.](#)

**Analisi cluster.** Questo è uno strumento esplorativo progettato per rivelare i raggruppamenti naturali (o cluster) all'interno dei dati. Ad esempio, è in grado di identificare gruppi diversi di clienti in base a varie caratteristiche demografiche e di acquisto. [Per ulteriori informazioni, vedere l'argomento Analisi cluster in il capitolo 3 a pag. 15.](#)

**Profili dei potenziali clienti.** Questa tecnica utilizza i risultati di una campagna precedente o di prova per creare profili descrittivi. È possibile utilizzare i profili per definire gruppi specifici di contatti nelle campagne successive. [Per ulteriori informazioni, vedere l'argomento Profili dei potenziali clienti in il capitolo 4 a pag. 20.](#)

**Tassi di risposta codice postale.** Questa tecnica utilizza i risultati di una campagna precedente per calcolare i tassi di risposta relativi al codice postale. Tali tassi possono essere utilizzati per identificare codici postali specifici nelle campagne successive. [Per ulteriori informazioni, vedere l'argomento Tassi di risposta codice postale in il capitolo 5 a pag. 26.](#)

**Propensione all'acquisto.** Questa tecnica utilizza i risultati di un mailing di prova o di campagne precedenti per generare punteggi di propensione. I punteggi indicano i contatti che risponderanno con maggiore probabilità. [Per ulteriori informazioni, vedere l'argomento Propensione all'acquisto in il capitolo 6 a pag. 34.](#)

**Test pacchetto di controllo.** Questa tecnica confronta le campagne di marketing per rilevare eventuali differenze significative nell'efficacia di diversi pacchetti o offerte. [Per ulteriori informazioni, vedere l'argomento Test pacchetto di controllo in il capitolo 7 a pag. 42.](#)

## ***Analisi RFM***

L'analisi RFM è una tecnica utilizzata per identificare i clienti esistenti che più probabilmente risponderanno a una nuova offerta. Questa tecnica è comunemente usata nel direct marketing. L'analisi RFM si basa sulla seguente semplice teoria:

- Il fattore più importante per l'identificazione dei clienti che probabilmente risponderanno a una nuova offerta è la **data dell'ultimo acquisto**. I clienti che hanno effettuato un acquisto più di recente effettueranno più probabilmente un altro acquisto rispetto ai clienti che hanno effettuato acquisti in passato.
- Il secondo fattore più importante è la **frequenza**. I clienti che hanno effettuato più acquisti in passato risponderanno con maggiore probabilità rispetto a quelli che hanno effettuato meno acquisti.
- Il terzo fattore più importante è l'importo totale della spesa, ovvero il **valore monetario**. I clienti che hanno speso di più (in totale per tutti gli acquisti) in passato risponderanno con maggiore probabilità rispetto a quelli che hanno speso meno.

### ***Funzionamento dell'analisi RFM***

- Ai clienti viene assegnato un punteggio per la data dell'ultimo acquisto basato sulla data dell'acquisto più recente o sull'intervallo di tempo trascorso dall'acquisto più recente. Il punteggio si basa su una semplice classificazione dei valori in un numero limitato di categorie. Ad esempio, se si utilizzano cinque categorie, i clienti con le date di acquisto più recenti ricevono una classificazione di 5 e quelli con le date di acquisto più lontane nel tempo ricevono una classificazione di 1.
- In questo modo, i clienti ricevono una classificazione della frequenza, con i valori più elevati che rappresentano una maggiore frequenza degli acquisti. Ad esempio, in uno schema di classificazione a cinque categorie, i clienti che acquistano più spesso ricevono una classificazione della frequenza pari a 5.
- Infine, i clienti vengono classificati in base al valore monetario, con i valori più elevati che ricevono la classificazione più alta. Continuando con l'esempio a cinque categorie, i clienti che hanno speso di più ricevono una classificazione del valore monetario pari a 5.

Il risultato è quattro punteggi per ogni cliente: data ultimo acquisto, frequenza, valore monetario e punteggio RFM combinato, ovvero i tre singoli punteggi concatenati in un solo valore. I clienti "migliori", quelli che più probabilmente rispondono a un'offerta, sono quelli con i punteggi RFM combinati più alti. Ad esempio, in una classificazione a cinque categorie, vi è un totale di 125 possibili punteggi RFM combinati e il massimo punteggio RFM combinato è 555.

**Considerazioni sui dati**

- Se le righe di dati rappresentano delle transazioni (ogni riga rappresenta una singola transazione e possono esservi più transazioni per ciascun cliente), utilizzare RFM dai dati della transazione. [Per ulteriori informazioni, vedere l'argomento Punteggi RFM dai dati della transazione a pag. 3.](#)
- Se le righe di dati rappresentano clienti con informazioni di riepilogo per tutte le transazioni (con colonne contenenti valori per l'importo totale speso, il numero totale di transazioni e la data della transazione più recente), utilizzare RFM dai dati dei clienti. [Per ulteriori informazioni, vedere l'argomento Punteggi RFM dai dati dei clienti a pag. 5.](#)

Figura 2-1

Confronto tra dati della transazione e dati dei clienti

Le righe sono le transazioni			
ID	Gender	Date	Amount
1	Male	9/25/2005	21
2	Male	1/15/2006	297
4	Male	2/5/2006	249
4	Male	5/7/2005	172
6	Male	4/16/2005	164
6	Male	4/12/2005	286
7	Female	7/12/2005	400
9	Male		
9	Male		
9	Male		
10	Female		
10	Female		

Le righe sono i clienti con i riepiloghi transazioni				
ID	Gender	Most Recent	Total Amount	Number of Purchases
1	Male	9/25/2005	21	1
2	Male	1/15/2006	297	1
4	Male	2/5/2006	421	2
6	Male	4/16/2005	450	2

**Punteggi RFM dai dati della transazione****Considerazioni sui dati**

Le variabili presenti nell'insieme di dati devono contenere le seguenti informazioni:

- Una variabile o combinazione di variabili che identifica ciascun caso (cliente).
- Una variabile con la data di ciascuna transazione.
- Una variabile con il valore monetario di ciascuna transazione.

Figura 2-2  
Dati della transazione RFM

ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

### Creazione di punteggi RFM dai dati della transazione

- ▶ Dai menu, scegliere:  
Direct marketing > Scegli tecnica
- ▶ Selezionare Aiutami a individuare i contatti migliori (Analisi RFM) e fare clic su Continua.
- ▶ Selezionare Dati transazioni, quindi fare clic su Continua.

Figura 2-3  
Dati transazioni, scheda Variabili

Analisi RFM dei dati delle transazioni

Variabili Categorizzazione Salva Output

Ciascuna riga contiene dati per una transazione. Le transazioni verranno aggregate dal cliente per l'analisi.

Variabili:

- Linea di prodotti [ProductLine]
- Numero di prodotto [ProductNumber]

Data transazione:

Data di acquisto [Date]

Importo transazioni:

Importo dell'acquisto [Amount]

Metodo di riepilogo:

Totale

Identificatori cliente:

ID cliente [ID]

OK Incolla Reimposta Annulla Aiuto

- ▶ Selezionare la variabile che contiene le date delle transazioni.

- ▶ Selezionare la variabile che contiene l'importo monetario per ciascuna transazione.
- ▶ Selezionare il metodo di riepilogo degli importi delle transazioni per ciascun cliente: somma (totale di tutte le transazioni), media, mediana o massimo (importo massimo delle transazioni).
- ▶ Selezionare la variabile o combinazione di variabili che identifica ciascun cliente in modo univoco. Ad esempio, i casi possono essere identificati da un codice ID univoco o da una combinazione di cognome e nome.

## ***Punteggi RFM dai dati dei clienti***

### ***Considerazioni sui dati***

Le variabili presenti nell'insieme di dati devono contenere le seguenti informazioni:

- La data dell'acquisto più recente o un intervallo di tempo dalla data dell'acquisto più recente. Questo dato verrà utilizzato per calcolare i punteggi relativi alla data dell'ultimo acquisto.
- Numero totale di acquisti. Questo dato verrà utilizzato per calcolare i punteggi relativi alla frequenza.
- Valore monetario di riepilogo per tutti gli acquisti. Questo dato verrà utilizzato per calcolare i punteggi relativi al valore monetario. In genere, si tratta della somma (totale) di tutti gli acquisti, ma può essere la media, l'importo massimo o un'altra misura di riepilogo.

Figura 2-4

*Dati dei clienti RFM*

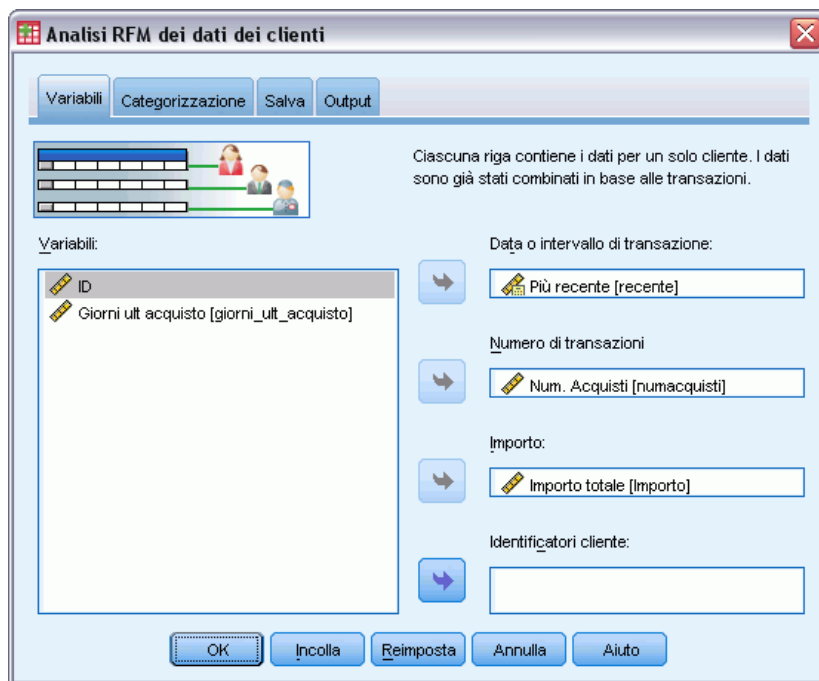
ID	TotalAmount	MostRecent	NumberOfPurchases
1	485.00	09/04/2006	5
2	350.00	11/10/2005	4
3	233.00	06/04/2005	2
4	936.00	08/18/2006	7
5	359.00	07/07/2006	3
6	249.00	07/16/2006	3
7	1089.00	02/15/2006	7
8	423.00	08/21/2006	4
9	689.00	08/31/2006	7
10	325.00	10/13/2005	3

Per scrivere i punteggi RFM in un nuovo insieme di dati, l'insieme di dati attivo deve inoltre contenere una variabile o combinazione di variabili che identifichi ciascun caso (cliente).

### ***Creazione di punteggi RFM dai dati dei clienti***

- ▶ Dai menu, scegliere:  
Direct marketing > Scegli tecnica
- ▶ Selezionare Aiutami a individuare i contatti migliori (Analisi RFM) e fare clic su Continua.
- ▶ Selezionare Dati clienti, quindi fare clic su Continua.

Figura 2-5  
Dati clienti, scheda Variabili



- ▶ Selezionare la variabile che contiene la data della transazione più recente o un numero che rappresenta un intervallo di tempo calcolato a partire dalla transazione più recente.
- ▶ Selezionare la variabile che contiene il numero totale delle transazioni per ciascun cliente.
- ▶ Selezionare la variabile che contiene l'importo monetario di riepilogo per ciascun cliente.
- ▶ Se si desidera scrivere i punteggi RFM in un nuovo insieme di dati, selezionare la variabile o combinazione di variabili che identifica in modo univoco ciascun cliente. Ad esempio, i casi possono essere identificati da un codice ID univoco o da una combinazione di cognome e nome.

## Categorizzazione RFM

Il processo di raggruppamento di un grande numero di valori numerici in un piccolo numero di categorie viene talvolta definito **categorizzazione**. Nell'analisi RFM, i bin sono le categorie classificate. La scheda Categorizzazione consente di modificare il metodo utilizzato per assegnare ai bin i valori relativi a data dell'ultimo acquisto, frequenza e valore monetario.



Figura 2-6  
Scheda Categorizzazione RFM

Analisi RFM dei dati dei clienti

Variabili Categorizzazione Salva Output

Metodo di categorizzazione

Nidificata

I clienti vengono prima categorizzati per data dell'ultimo acquisto. Successivamente, i clienti in ogni bin di data dell'ultimo acquisto vengono categorizzati per frequenza. Quindi, i clienti in ogni bin di data dell'ultimo acquisto-frequenza vengono categorizzati per valore monetario.

Indipendente

I clienti vengono categorizzati indipendentemente da data dell'ultimo acquisto, frequenza e valore monetario.

Numero di bin

Data ultimo acquisto: 5

Frequenza: 5

Valore monetario: 5

Pari merito

Assegna pari merito allo stesso bin

Assegna in modo casuale i pari merito ai bin adiacenti

OK Incolla Reimposta Annulla Aiuto

### **Metodo di categorizzazione**

**Nidificata.** Nella categorizzazione nidificata, si assegna una classificazione semplice ai valori della data dell'ultimo acquisto. All'interno di ogni classificazione della data dell'ultimo acquisto, si assegna ai clienti una classificazione della frequenza e, all'interno di ogni rango della frequenza, si assegna una classificazione per il valore monetario. Ciò tende a uniformare una distribuzione più omogenea dei punteggi RFM combinati, ma ha lo svantaggio di rendere i punteggi relativi a frequenza e valore monetario più difficili da interpretare. Ad esempio, una classificazione di frequenza pari a 5 per un cliente con data dell'ultimo acquisto classificata 5 potrebbe non significare la stessa cosa di una classificazione di frequenza pari a 5 per un cliente con classificazione della data dell'ultimo acquisto pari a 4, poiché la classificazione della frequenza dipende dalla classificazione della data dell'ultimo acquisto.

**Indipendente.** Vengono assegnate classificazioni semplici ai valori di data dell'ultimo acquisto, frequenza e valore monetario. Le tre classificazioni vengono assegnate indipendentemente l'una dall'altra. L'interpretazione di ciascuno dei tre componenti RFM quindi non è ambigua: un punteggio di 5 per la frequenza per un cliente ha lo stesso significato del punteggio di 5 per la frequenza per un altro cliente, indipendentemente dai punteggi relativi alla data dell'ultimo acquisto. Per i campioni più piccoli, ciò ha lo svantaggio di produrre una distribuzione meno uniforme dei punteggi RFM combinati.

### **Numero di bin**

Numero di categorie (bin) da utilizzare per ciascun componente per creare i punteggi RFM. Il numero totale di punteggi RFM combinati possibili è il prodotto di tre valori. Ad esempio, 5 bin di data dell'ultimo acquisto, 4 bin di frequenza e 3 bin di valore monetario creano un totale di 60 possibili punteggi RFM combinati, nell'intervallo tra 111 e 543.

- Il valore predefinito è 5 per ciascun componente, che crea 125 possibili punteggi RFM combinati, da 111 a 555.
- Il numero massimo consentito di bin per ogni componente del punteggio è nove.

### **Pari merito**

Un caso “pari merito” si ha semplicemente con due o più valori uguali per data dell'ultimo acquisto, frequenza o valore monetario. In teoria, si dovrebbe avere lo stesso numero di clienti in ogni bin, ma un numero elevato di valori a pari merito possono influire sulla distribuzione nei bin. Per la gestione dei casi pari merito esistono due alternative:

- **Assegna pari merito allo stesso bin.** Questo metodo assegna sempre valori a pari merito allo stesso bin, indipendentemente dall'impatto sulla distribuzione. Ciò consente di utilizzare un metodo di categorizzazione coerente: Se due clienti hanno lo stesso valore per la data dell'ultimo acquisto, ad essi verrà sempre assegnato lo stesso punteggio per la data dell'ultimo acquisto. In un caso molto particolare, tuttavia, si potrebbero avere 1.000 clienti, 500 dei quali che effettuano l'acquisto più recente nella stessa data. In una classificazione a 5 bin, il 50% dei clienti ricevono quindi un punteggio di 5 per la data dell'ultimo acquisto, anziché il 20% desiderato.

Si noti che il metodo di categorizzazione nidificato “uniformità” in qualche modo è più complicato per i punteggi basati su frequenza e valore monetario, poiché per la frequenza i punteggi vengono assegnati nell'ambito dei bin di punteggio per la data dell'ultimo acquisto, mentre i punteggi per il valore monetario vengono assegnati nell'ambito dei bin di punteggio per la frequenza. Quindi, due clienti con lo stesso valore di frequenza possono non avere lo stesso punteggio per la frequenza se non hanno anche lo stesso punteggio per la data dell'ultimo acquisto, a prescindere dal modo in cui sono gestiti i valori a pari merito.

- **Assegna in modo casuale i pari merito.** Assicura una distribuzione uniforme mediante l'assegnazione di un fattore di varianza casuale molto piccolo prima della classificazione. In questo modo, per l'assegnazione dei valori ai bin classificati non vi sono valori a pari merito. Questo processo non ha effetto sui valori originali. Viene utilizzato solo per distinguere i casi pari merito. Se da un lato il risultato è una distribuzione uniforme nei bin (circa lo stesso numero di clienti in ogni bin), si possono avere punteggi completamente diversi per i clienti che presentano punteggi simili o identici per la data dell'ultimo acquisto, la frequenza e/o il valore monetario, in particolare se il numero totale di clienti è relativamente piccolo e/o il numero di casi pari merito è relativamente alto.

Tabella 2-1

Confronto tra Assegna pari merito allo stesso bin e Assegna in modo casuale i pari merito

ID	Acquisto più recente (data ultimo acquisto)	Classificazione data ultimo acquisto	
		Assegna pari merito allo stesso bin	Assegna in modo casuale i pari merito
1	10/29/2006	5	5
2	10/28/2006	4	4
3	10/28/2006	4	4
4	10/28/2006	4	5
5	10/28/2006	4	3
6	9/21/2006	3	3
7	9/21/2006	3	2
8	8/13/2006	2	2
9	8/13/2006	2	1
10	6/20/2006	1	1

- In questo esempio, l'assegnazione di pari merito allo stesso bin produce una distribuzione non uniforme: 5 (10%), 4 (40%), 3 (20%), 2 (20%), 1 (10%).
- L'assegnazione casuale dei pari merito produce un 20% in ogni bin, ma per ottenere questo risultato i quattro casi con valore di data 28/10/2006 vengono assegnati a 3 bin differenti e anche i 2 casi con valore di data 13/08/2006 vengono assegnati a bin differenti.

Notare che il modo in cui i pari merito vengono assegnati a bin differenti è completamente casuale (entro i limiti in base ai quali il risultato finale deve essere un numero uguale di casi in ogni bin). Se è stato calcolato un secondo gruppo di punteggi utilizzando lo stesso metodo, la classificazione per un caso particolare con un valore a pari merito potrebbe cambiare. Ad esempio, le classificazioni per la data dell'ultimo acquisto di 5 e 3 rispettivamente per i casi 4 e 5 potrebbero essere scambiate la seconda volta.

## ***Salvataggio dei punteggi RFM dai dati della transazione***

L'opzione Punteggi RFM dai dati della transazione crea sempre un nuovo insieme di dati aggregato con una riga per ogni cliente. Utilizzare la scheda Salva per specificare i punteggi e le altre variabili da salvare e la posizione in cui salvarli.

Figura 2-7  
Dati transazioni, scheda Salva

Analisi RFM dei dati delle transazioni

Variabili    Categorizzazione    **Salva**    Output

Variabili:

Salva	Contenuto	Nome della variabile salvata
<input checked="" type="checkbox"/>	Data della transazione più recente	Data_più_recente
<input checked="" type="checkbox"/>	Numero di transazioni	Conteggio_transazioni
<input checked="" type="checkbox"/>	Importo	Importo
<input checked="" type="checkbox"/>	Punteggio data ultimo acquisto	Recency_score
<input checked="" type="checkbox"/>	Punteggio frequenza	Frequency_score
<input checked="" type="checkbox"/>	Punteggio valore monetario	Monetary_score
<input checked="" type="checkbox"/>	Punteggio RFM	RFM_score

Posizione

Crea nuovo insieme di dati    Nome:

Scrivi nuovo file di dati   

### Variabili

Le variabili ID che identificano in modo univoco ciascun cliente vengono automaticamente salvate nel nuovo insieme di dati. Le seguenti variabili aggiuntive possono essere salvate nel nuovo insieme di dati:

- **Data della transazione più recente per ciascun cliente.**
- **Numero di transazioni.** Numero totale delle righe delle transazioni per ciascun cliente.
- **Importo.** Importo riassuntivo per ogni cliente basato sul metodo di riepilogo selezionato nella scheda Variabili.
- **Punteggio data ultimo acquisto.** Il punteggio assegnato a ciascun cliente in base alla data della transazione più recente. I punteggi più elevati indicano date più recenti.
- **Punteggio frequenza.** Il punteggio assegnato a ciascun cliente in base al numero totale di transazioni. I punteggi più elevati indicano un maggior numero di transazioni.
- **Punteggio valore monetario.** Il punteggio assegnato a ciascun cliente in base alla misura di riepilogo selezionata per il valore monetario. I punteggi più elevati indicano un valore più elevato per la misura di riepilogo del valore monetario.
- **Punteggio RFM.** I tre singoli punteggi combinati in un unico valore: *(data dell'ultimo acquisto x 100) + (frequenza x 10) + valore monetario.*

Per impostazione predefinita, tutte le variabili disponibili vengono incluse nel nuovo insieme di dati, quindi deselezionare quelle che non si desidera includere. Se necessario, è possibile specificare i propri nomi di variabile. I nomi delle variabili devono essere conformi alle regole standard dei nomi delle variabili.

### Posizione

L'opzione Punteggi RFM dai dati della transazione crea sempre un nuovo insieme di dati aggregato con una riga per ogni cliente. È possibile creare un nuovo insieme di dati nella sessione corrente o salvare i dati dei punteggi RFM in un file di dati esterno. I nomi degli insiemi di dati devono essere conformi alle regole standard dei nomi delle variabili (questa restrizione non è valida per i nomi dei file di dati esterni).

## Salvataggio dei punteggi RFM dai dati dei clienti

Per i dati dei clienti, è possibile aggiungere le variabili dei punteggi RFM all'insieme di dati attivo o creare un nuovo insieme di dati contenente le variabili dei punteggi selezionate. Utilizzare la scheda Salva per specificare le variabili dei punteggi da salvare e la posizione in cui salvarle.

Figura 2-8  
Dati clienti, scheda Salva

The screenshot shows a dialog box titled "Analisi RFM dei dati dei clienti" with a close button (X) in the top right corner. The dialog has four tabs: "Variabili", "Categorizzazione", "Salva", and "Output". The "Salva" tab is selected.

Under "Nomi della variabile salvata", there are two radio buttons: "Genera automaticamente nomi univoci" (selected) and "Nomi personalizzati".

Under "Variabili:", there is a table with three columns: "Salva", "Contenuto", and "Nome della variabile salvata".

Salva	Contenuto	Nome della variabile salvata
<input checked="" type="checkbox"/>	Punteggio data ultimo acquisto	Recency_score
<input checked="" type="checkbox"/>	Punteggio frequenza	Frequency_score
<input checked="" type="checkbox"/>	Punteggio valore monetario	Monetary_score
<input checked="" type="checkbox"/>	Punteggio RFM	RFM_score

Under "Posizione", there are three radio buttons: "Scrivi in insieme di dati attivo" (selected), "Crea nuovo insieme di dati", and "Scrivi nuovo file di dati". The "Crea nuovo insieme di dati" option has a "Nome:" text box next to it. The "Scrivi nuovo file di dati" option has a "Sfoglia..." button next to it.

At the bottom of the dialog, there are five buttons: "OK", "Incolla", "Reimposta", "Annulla", and "Aiuto".

**Nomi delle variabili salvate**

- **Genera automaticamente nomi univoci.** Quando si aggiungono variabili di punteggi all'insieme di dati attivo, garantisce l'univocità dei nomi delle nuove variabili. Ciò è particolarmente utile se si desidera aggiungere più gruppi di punteggi RFM diversi (basati su criteri differenti) all'insieme di dati attivo.
- **Nomi personalizzati.** Consente di assegnare i propri nomi di variabile alle variabili dei punteggi. I nomi delle variabili devono essere conformi alle regole standard dei nomi delle variabili.

**Variabili**

Selezionare le variabili dei punteggi che si desidera salvare:

- **Punteggio data ultimo acquisto.** Il punteggio assegnato a ciascun cliente in base al valore della variabile Data o intervallo di transazione selezionata nella scheda Variabili. I punteggi più elevati vengono assegnati alle date più recenti o ai valori di intervallo più bassi.
- **Punteggio frequenza.** Il punteggio assegnato a ciascun cliente in base alla variabile Numero di transazioni selezionata nella scheda Variabili. I punteggi più elevati sono assegnati ai valori più elevati.
- **Punteggio valore monetario.** Il punteggio assegnato a ciascun cliente in base alla variabile Importo selezionata nella scheda Variabili. I punteggi più elevati sono assegnati ai valori più elevati.
- **Punteggio RFM.** I tre singoli punteggi combinati in un unico valore:  $(data\ ultimo\ acquisto * 100) + (frequenza * 10) + valore\ monetario$ .

**Posizione**

Per i dati dei clienti, esistono tre alternative per il salvataggio dei nuovi punteggi RFM:

- **Insieme di dati attivo.** Le variabili dei punteggi RFM selezionate vengono aggiunte all'insieme di dati attivo.
- **Nuovo insieme di dati.** Le variabili dei punteggi RFM selezionate e le variabili ID che identificano ogni cliente (caso) in modo univoco verranno scritte in un nuovo insieme di dati nella sessione corrente. I nomi degli insiemi di dati devono essere conformi alle regole standard dei nomi delle variabili. Questa opzione è disponibile solo se si selezionano una o più variabili Identificatore cliente nella scheda Variabili.
- **File.** I punteggi RFM selezionati e le variabili ID che identificano in modo univoco ogni cliente (caso) verranno salvati in un file di dati esterno. Questa opzione è disponibile solo se si selezionano una o più variabili Identificatore cliente nella scheda Variabili.

## Output RFM

Figura 2-9  
Scheda Output RFM



### **Dati categorizzati**

I grafici e le tabelle relativi ai dati categorizzati si basano sui punteggi calcolati per la data dell'ultimo acquisto, la frequenza e il valore monetario.

**Mappa termica del valore monetario medio per data ultimo acquisto e frequenza.** La mappa termica della distribuzione monetaria media visualizza il valore monetario medio relativo alle categorie definite dai punteggi sull'ultima data di acquisto e la frequenza. Le aree più scure indicano un valore monetario medio più alto.

**Grafico dei conteggi dei bin.** Il grafico dei conteggi dei bin visualizza la distribuzione nei bin relativa al metodo di categorizzazione selezionato. Ogni barra rappresenta il numero di casi a cui verranno assegnati i singoli punteggi RFM combinati.

- Per quanto sia auspicabile ottenere una distribuzione piuttosto omogenea con tutte, o quasi, le barre della stessa altezza, quando si utilizza il metodo di categorizzazione predefinito che assegna i valori pari merito allo stesso bin, verrà rilevata una certa varianza.
- La presenza di fluttuazioni estreme nella distribuzione dei bin e/o un numero elevato di bin vuoti potrebbe indicare la necessità di provare un altro metodo di categorizzazione (meno bin e/o assegnazione casuale dei valori pari merito) oppure di riconsiderare l'effettiva idoneità dell'analisi RFM.

**Tabella dei conteggi dei bin.** Visualizza le stesse informazioni presenti nel grafico dei conteggi dei bin, espresse però sotto forma di tabella, con i conteggi visualizzati in ogni cella.

***Dati non categorizzati***

I grafici e le tabelle relativi ai dati non categorizzati si basano sulle variabili originali usate per creare i punteggi relativi alla data dell'ultimo acquisto, alla frequenza e al valore monetario.

**Istogrammi.** Gli istogrammi visualizzano la distribuzione relativa dei valori per le tre variabili utilizzate per calcolare i punteggi sulla data dell'ultimo acquisto, la frequenza e il valore monetario. Questi istogrammi, spesso, indicano distribuzioni in qualche modo asimmetriche invece che distribuzioni normali o simmetriche.

L'asse orizzontale di ogni istogramma visualizza sempre i valori più bassi a sinistra e quelli via via più elevati verso destra. Nel caso della data dell'ultimo acquisto, tuttavia, l'interpretazione del grafico dipende dal tipo di misurazione utilizzata: data o intervallo di tempo. Nel caso delle date, le barre a sinistra rappresentano i valori meno recenti (una data meno recente ha un valore meno elevato rispetto a una data più recente). Nel caso degli intervalli di tempo, le barre a sinistra rappresentano i valori più recenti (più limitato l'intervallo di tempo, più recente la transazione).

**Grafici a dispersione di coppie di variabili.** Questi grafici a dispersione visualizzano le relazioni fra le tre variabili utilizzate per calcolare i punteggi sulla data dell'ultimo acquisto, la frequenza e il valore monetario.

Spesso è possibile rilevare significativi raggruppamenti lineari di punti sulla scala della frequenza, poiché la frequenza spesso rappresenta un intervallo relativamente limitato di valori discreti. Ad esempio, se il numero totale delle transazioni non supera 15, esistono solo 15 valori di frequenza possibili (a meno che non si contino le transazioni frazionarie) mentre potrebbero esserci centinaia di valori relativamente alla data dell'ultimo acquisto e migliaia di valori monetari.

L'interpretazione dell'asse della data dell'ultimo acquisto dipende dal tipo di misurazione utilizzata: data o intervallo di tempo. Nel caso delle date, i punti più vicini all'origine rappresentano le date meno recenti. Nel caso degli intervalli di tempo, i punti più vicini all'origine rappresentano i valori più recenti.



# ***Analisi cluster***

L'analisi cluster è uno strumento esplorativo progettato per rivelare i raggruppamenti naturali (o cluster) all'interno dei dati. Ad esempio, è in grado di identificare gruppi diversi di clienti in base a varie caratteristiche demografiche e di acquisto.

**Esempio.** I produttori di articoli al dettaglio e prodotti per i consumatori applicano regolarmente tecniche di raggruppamento ai dati relativi alle abitudini di acquisto dei propri clienti, al sesso, all'età, al livello di reddito e così via. In questo modo adattano le strategie di sviluppo del prodotto e di mercato ad ogni gruppo di consumatori al fine di aumentare le vendite e accrescere la fedeltà alla marca.

## ***Considerazioni sui dati dell'analisi cluster***












**Dati.** Questa procedura può essere utilizzata sia con i campi continui sia con i campi categoriali. Ogni record (riga) rappresenta un cliente da raggruppare e i campi (variabili) corrispondono agli attributi in base ai quali viene eseguito il raggruppamento.

**Ordine dei record.** È utile notare che i risultati possono dipendere dall'ordine dei record. Per ridurre al minimo gli effetti dell'ordine, può essere opportuno disporre i record in ordine casuale. Può essere utile eseguire l'analisi più volte, con i record disposti in ordini casuali diversi, per verificare la stabilità di una soluzione specifica.

**Livello di misurazione.** L'assegnazione del livello di misurazione corretto è importante perché influisce sul calcolo dei risultati.

- **Nominale.** Una variabile può essere considerata nominale quando i relativi valori rappresentano categorie prive di ordinamento intrinseco, per esempio l'ufficio di una società. Tra gli esempi di variabili nominali troviamo la regione, il codice postale e la religione.
- **Ordinale.** Una variabile può essere considerata ordinale quando i relativi valori rappresentano categorie con qualche ordinamento intrinseco, per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, i punteggi di atteggiamento corrispondenti a gradi di soddisfazione o fiducia e i punteggi di preferenza.
- **Continuo.** Una variabile può essere considerata di scala (continua) quando i relativi valori rappresentano categorie ordinate con una metrica significativa, tale che i confronti fra le distanze dei relativi valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

Un'icona accanto a ciascun campo indica il livello di misurazione corrente.

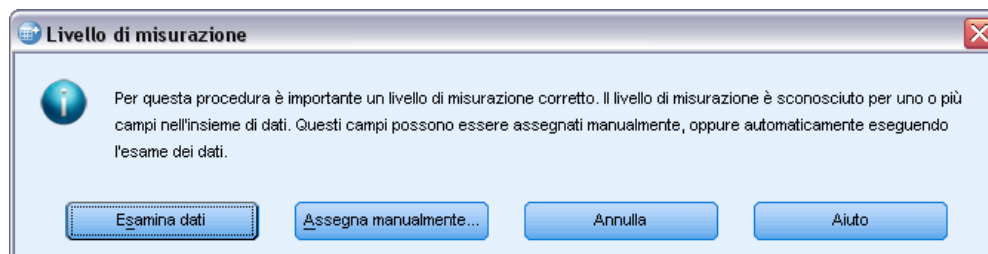
Livello di misurazione	Tipo di dati			
	Numerico	Stringa	Data	Ora
Scala (continuo)		n/d		
Ordinale				
Nominale				

È possibile modificare il livello di misurazione nella Visualizzazione variabili dell'Editor dei dati ; in alternativa, è possibile utilizzare la finestra di dialogo Definisci proprietà variabili per indicare un livello di misurazione appropriato per ogni campo.

### **Campi con livello di misurazione sconosciuto**

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) dell'insieme di dati è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Figura 3-1  
Avviso Livello di misurazione



- **Esamina dati.** Legge i dati dell'insieme di dati attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con insiemi di dati di grandi dimensioni, questa operazione può richiedere del tempo.
- **Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Visualizzazione variabili dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

### **Per ottenere l'analisi cluster**

Dai menu, scegliere:

Direct marketing > Scegli tecnica

- Selezionare Segmenta i miei contatti in cluster.

Figura 3-2

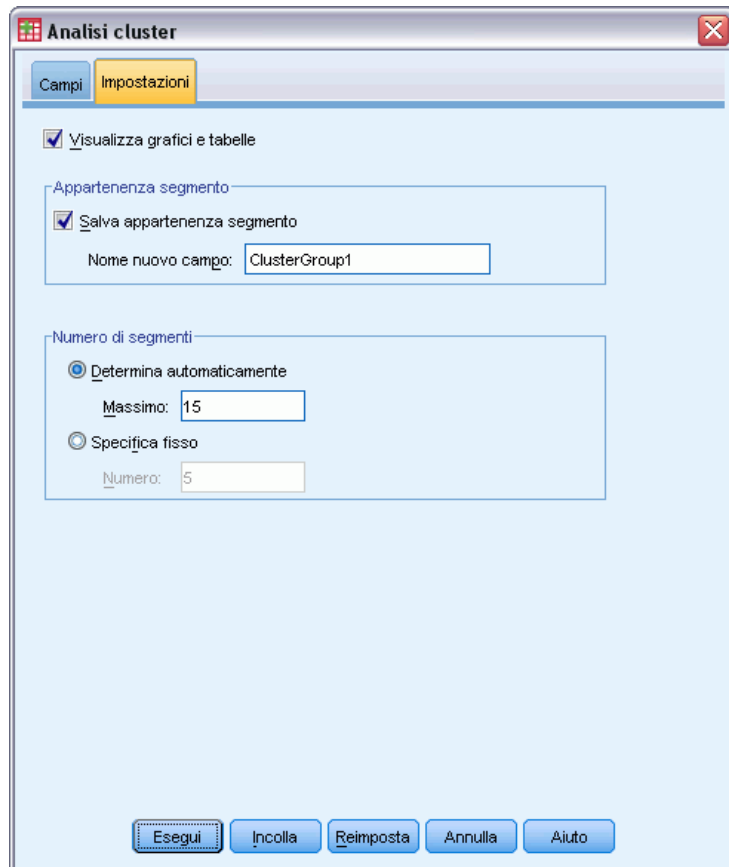
Scheda Campi di Analisi cluster



- Selezionare i campi categoriali (nominali, ordinali) e continui (scala) da utilizzare per creare i segmenti.
- Fare clic su Esegui per eseguire la procedura.

## Impostazioni

Figura 3-3  
Scheda Impostazioni di Analisi cluster



La scheda Impostazioni consente di attivare o disattivare la visualizzazione di grafici e tabelle che descrivono i segmenti, salvare un nuovo campo nell'insieme di dati per identificare il segmento (cluster) per ogni record dell'insieme, nonché specificare il numero di segmenti da includere nella soluzione cluster.

**Visualizza grafici e tabelle.** Visualizza i grafici e le tabelle che descrivono i segmenti.

**Appartenenza segmento.** Salva un nuovo campo (variabile) che identifica il segmento di appartenenza di ogni record.

- I nomi dei campi devono essere conformi alle regole di denominazione di IBM® SPSS® Statistics.
- Il nome del campo di appartenenza segmento non può essere uguale al nome di un altro campo già esistente nell'insieme di dati. Se si esegue questa procedura più di una volta sullo stesso insieme di dati sarà necessario specificare ogni volta un nome diverso.

- **Numero di segmenti.** Controlla la modalità di determinazione del numero di segmenti.
- **Determina automaticamente.** La procedura determina automaticamente il numero di segmenti ottimale, fino al massimo specificato.

**Specifica fisso.** La procedura genera il numero di segmenti specificato.

## *Profili dei potenziali clienti*

Questa tecnica utilizza i risultati di una campagna precedente o di prova per creare profili descrittivi. È possibile utilizzare i profili per definire gruppi specifici di contatti nelle campagne successive. Il campo Risposta indica i contatti che hanno risposto alla campagna precedente o di prova. L'elenco Profili contiene le caratteristiche da utilizzare per creare il profilo.

**Esempio.** In base ai risultati di un mailing di prova, la divisione Direct marketing di una società desidera generare i profili dei tipi di clienti con le maggiori probabilità di rispondere positivamente a un'offerta a seconda dei dati demografici.

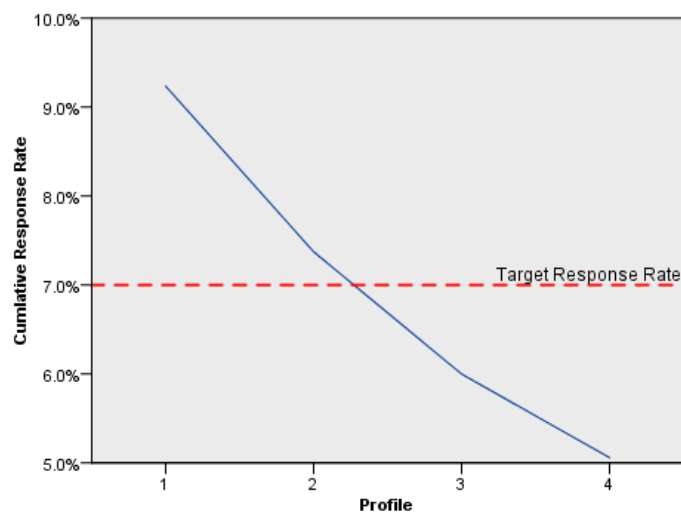
### ***Output***

Output comprende una tabella che riporta la descrizione di ogni gruppo di profili e mostra i tassi di risposta (percentuale di risposte positive), i tassi di risposta cumulativi e un grafico di questi ultimi. Se si include un tasso di risposta obiettivo minimo, la tabella sarà codificata mediante colori per mostrare i profili che soddisfano il tasso di risposta cumulativo minimo, e il grafico includerà una linea di riferimento in corrispondenza del valore del tasso di risposta minimo specificato.

Figura 4-1  
Tabella e grafico del tasso di risposta

Tasso di risposta				
Numero	Profilo			
	Descrizione	Dimensione gruppo	Tasso di risposta	Tasso di risposta cumulativo
1	Regione = "Ovest","Sud","Est" Sesso = "Femmina" Coniugato = "No"	379	9.2%	9.2%
2	Regione = "Ovest","Sud","Est" Sesso = "Femmina" Coniugato = "Si"	299	5.0%	7.4%
3	Regione = "Ovest","Sud","Est" Sesso = "Maschio"	722	4.7%	6.0%
4	Regione = "Nord"	517	2.5%	5.1%

Verde: soddisfa il tasso di risposta obiettivo.  
Rosso: non soddisfa il tasso di risposta obiettivo.



### **Profili dei potenziali clienti: considerazioni sui dati**

**Campo Risposta.** Il campo Risposta deve essere nominale o ordinale. Può essere di tipo stringa o numerico. Se contiene un valore che indica il numero o l'importo degli acquisti, sarà necessario creare un nuovo campo in cui tutte le risposte positive siano rappresentate da un solo valore. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale a pag. 24.](#)

**Valore risposta positiva.** Il valore della risposta positiva identifica i clienti che hanno risposto positivamente (ad esempio, che hanno effettuato un acquisto). Si presume che tutti gli altri valori non mancanti relativi alle risposte indichino una risposta negativa. Se sono state definite etichette dei valori per il campo Risposta, tali etichette sono visualizzate nell'elenco a discesa.

**Crea profili con.** Questi campi possono essere nominali, ordinali o continui (scala). Possono essere di tipo stringa o numerici.

**Livello di misurazione.** L'assegnazione del livello di misurazione corretto è importante perché influisce sul calcolo dei risultati.

- **Nominale.** Una variabile può essere considerata nominale quando i relativi valori rappresentano categorie prive di ordinamento intrinseco, per esempio l'ufficio di una società, Tra gli esempi di variabili nominali troviamo la regione, il codice postale e la religione.
- **Ordinale.** Una variabile può essere considerata ordinale quando i relativi valori rappresentano categorie con qualche ordinamento intrinseco, per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, i punteggi di attecchimento corrispondenti a gradi di soddisfazione o fiducia e i punteggi di preferenza.
- **Continuo.** Una variabile può essere considerata di scala (continua) quando i relativi valori rappresentano categorie ordinate con una metrica significativa, tale che i confronti fra le distanze dei relativi valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

Un'icona accanto a ciascun campo indica il livello di misurazione corrente.

Livello di misurazione	Tipo di dati			
	Numerico	Stringa	Data	Ora
Scala (continuo)		n/d		
Ordinale				
Nominale				

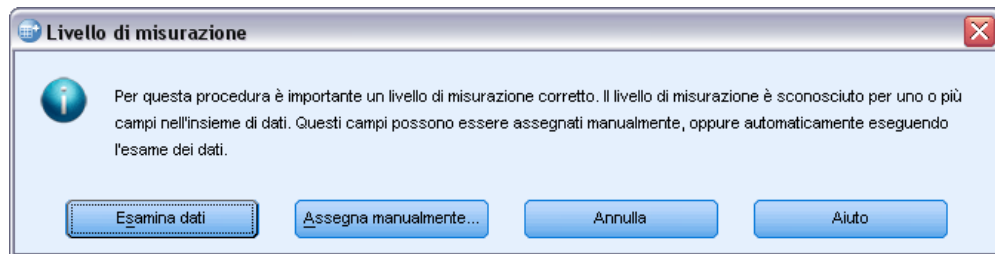
È possibile modificare il livello di misurazione nella Visualizzazione variabili dell'Editor dei dati ; in alternativa, è possibile utilizzare la finestra di dialogo Definisci proprietà variabili per indicare un livello di misurazione appropriato per ogni campo.

### **Campi con livello di misurazione sconosciuto**

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) dell'insieme di dati è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Figura 4-2

Avviso Livello di misurazione



- **Esamina dati.** Legge i dati dell'insieme di dati attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con insiemi di dati di grandi dimensioni, questa operazione può richiedere del tempo.



- **Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Visualizzazione variabili dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

### **Per ottenere i profili dei potenziali clienti**

Dai menu, scegliere:

Direct marketing > Scegli tecnica

- Selezionare Genera profili dei contatti che hanno risposto a un'offerta.

Figura 4-3

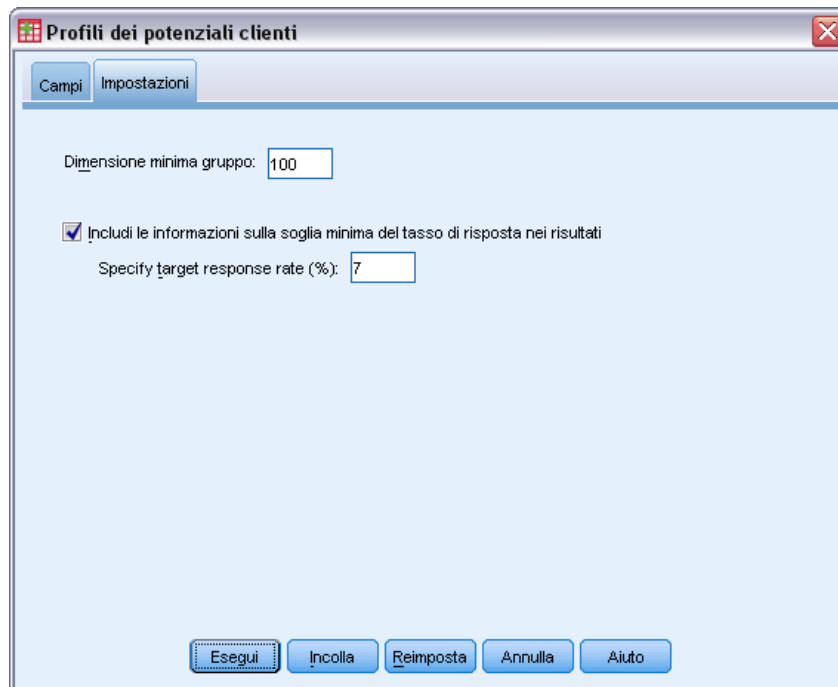
*Profili dei potenziali clienti: scheda Campi*



- Selezionare il campo che identifica i contatti che hanno risposto all'offerta. Il campo deve essere nominale o ordinale.
- Immettere il valore che indica una risposta positiva. Se per alcuni valori sono state definite delle etichette, è possibile selezionare l'etichetta del valore dall'elenco a discesa per visualizzare il valore corrispondente.
- Selezionare i campi che si desidera utilizzare per creare i profili.
- Fare clic su Esegui per eseguire la procedura.

## Impostazioni

Figura 4-4  
Profili dei potenziali clienti: scheda Impostazioni



La scheda Impostazioni consente di controllare le dimensioni minime del gruppo di profili e di includere nell'output una soglia minima per il tasso di risposta.

**Dimensione minima gruppo profili.** Ogni profilo rappresenta le caratteristiche comuni di un gruppo di contatti dell'insieme di dati (ad esempio, donne con meno di 40 anni che abitano nelle regioni occidentali). Per impostazione predefinita, la dimensione minima del gruppo di profili è 100. Dimensioni inferiori possono rivelare un maggior numero di gruppi, ma dimensioni superiori possono fornire risultati più affidabili. Il valore deve essere un numero intero positivo.

**Includi le informazioni sulla soglia minima del tasso di risposta nei risultati.** I risultati comprendono una tabella che mostra i tassi di risposta (percentuale di risposte positive), i tassi di risposta cumulativi e un grafico di questi ultimi. Se si inserisce un tasso di risposta obiettivo minimo, la tabella sarà codificata mediante colori per mostrare i profili che soddisfano il tasso di risposta cumulativo minimo, e il grafico includerà una linea di riferimento in corrispondenza del valore del tasso di risposta minimo specificato. Il valore deve essere maggiore di 0 e minore di 100.

## Creazione di un campo Risposta categoriale

Il campo risposta deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro

valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero.

- Se le risposte negative vengono registrate come 0 (non vuoto, che viene considerato valore mancante), il calcolo si può effettuare con la seguente formula:

$$\text{NuovoNome} = \text{VecchioNome} > 0$$

dove *NuovoNome* è il nome del nuovo campo e *VecchioNome* è il nome del campo originale. Questa è un'espressione logica che assegna il valore 1 a tutti i valori non mancanti superiori a 0, e il valore 0 a tutti i valori non mancanti inferiori o uguali a 0.

- Se per le risposte negative non viene registrato alcun valore, questi valori vengono considerati mancanti e la formula è leggermente più complicata:

$$\text{NuovoNome} = \text{NOT}(\text{MISSING}(\text{VecchioNome}))$$

In questa espressione logica, a tutti i valori di risposta non mancanti viene assegnato valore 1 e a tutti i valori di risposta mancanti viene assegnato valore 0.

- Se non è possibile distinguere i valori di risposta negativi (0) dai valori mancanti, non è possibile calcolare un valore di risposta preciso. Se i valori mancanti effettivi sono relativamente pochi, questo può non avere un effetto significativo sui tassi di risposta calcolati. Se, invece, i valori mancanti sono numerosi (ad esempio quando i dati delle risposte vengono registrati solo per un piccolo campione di verifica dell'insieme di dati totale), i tassi di risposta calcolati saranno privi di significato perché saranno notevolmente più bassi rispetto ai tassi di risposta effettivi.

### **Per creare un campo Risposta categoriale**

- ▶ Dai menu, scegliere:  
Trasforma > Calcola variabile
- ▶ Per Variabile di destinazione, immettere il nome del nuovo campo (variabile).
- ▶ Se le risposte negative vengono registrate come 0, immettere *VecchioNome*>0 in Espressione numerica, dove *VecchioNome* è il nome del campo originale.
- ▶ Se le risposte negative vengono registrate come mancanti (vuote), immettere *NOT(MISSING(VecchioNome))* in Espressione numerica, dove *VecchioNome* è il nome del campo originale.

## ***Tassi di risposta codice postale***

Questa tecnica utilizza i risultati di una campagna precedente per calcolare i tassi di risposta relativi al codice postale. Tali tassi possono essere utilizzati per identificare codici postali specifici nelle campagne successive. Il campo Risposta indica chi ha risposto alla campagna precedente. Il Campo codice postale identifica il campo che contiene i codici postali.

**Esempio.** In base ai risultati di un precedente mailing, la divisione Direct marketing di una società genera dei tassi di risposta in base ai codici postali. Secondo vari criteri, quali il tasso minimo di risposta accettabile e/o il numero massimo di contatti da includere nel mailing, la divisione potrà in seguito definire codici postali specifici.

### ***Output***

L'output di questa procedura è costituito da un nuovo insieme di dati contenente i tassi di risposta in base al codice postale, nonché una tabella e un grafico che riassumono i risultati per rango di decile (primo 10%, primo 20%, ecc.). La tabella può essere codificata mediante colori in base a un tasso di risposta cumulativo minimo specificato dall'utente o a un numero massimo di contatti.

Figura 5-1  
Insieme di dati con tassi di risposta in base al codice postale

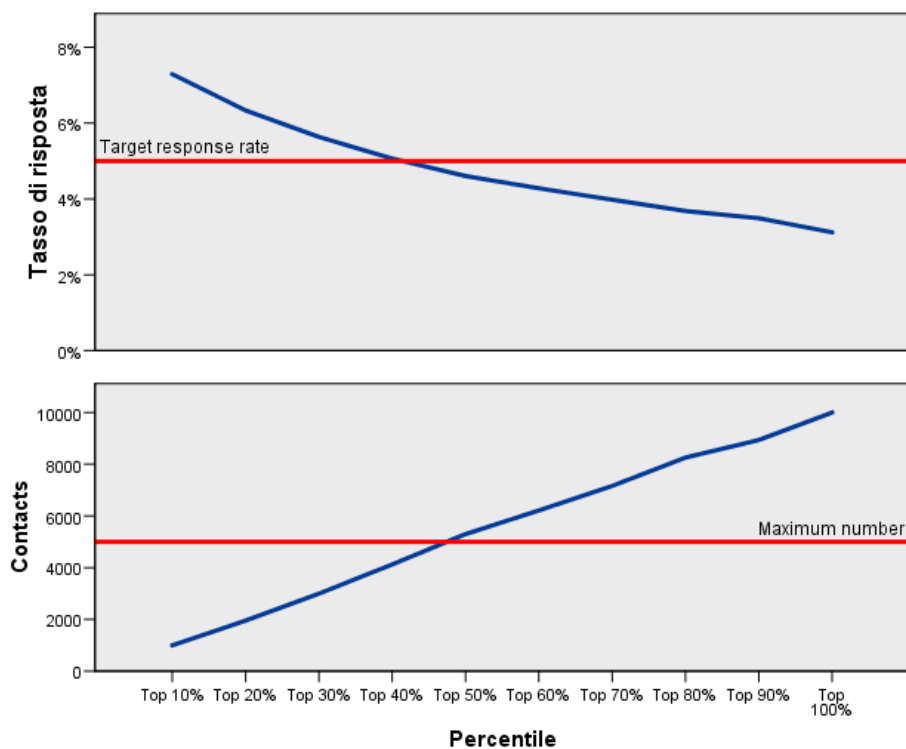
The screenshot shows the PASW Statistics Data Editor interface. The main window displays a data table with 21 rows and 7 columns. The columns are labeled: CodicePostale, ResponseRate, Responses, Contacts, Index, and Rank. The data is sorted by Rank in ascending order. The first row (Rank 1) has a ResponseRate of 10.00%, 4.00 Responses, 40 Contacts, and an Index of 3.60. The last row (Rank 2) has a ResponseRate of 5.36%, 3.00 Responses, 56 Contacts, and an Index of 2.84. The interface includes a menu bar (File, Modifica, Visualizza, Dati, Trasforma, Analizza, Applicazioni, Grafici, Utilità, Finestra, Aiuto) and a toolbar with various icons. The status bar at the bottom indicates 'PASW Statistics Il processore è pronto'.

	CodicePostale	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.00%	4.00	40	3.60	1
2	098	8.82%	6.00	68	5.47	1
3	740	7.76%	9.00	116	8.30	1
4	100	7.69%	7.00	91	6.46	1
5	110	7.69%	5.00	65	4.62	1
6	954	7.55%	4.00	53	3.70	1
7	108	7.32%	6.00	82	5.56	1
8	107	7.04%	5.00	71	4.65	1
9	090	6.90%	4.00	58	3.72	1
10	966	6.90%	4.00	58	3.72	1
11	760	6.72%	8.00	119	7.46	1
12	113	6.25%	5.00	80	4.69	1
13	927	6.00%	3.00	50	2.82	1
14	969	6.00%	3.00	50	2.82	1
15	972	5.88%	3.00	51	2.82	2
16	123	5.71%	2.00	35	1.89	2
17	950	5.66%	3.00	53	2.83	2
18	602	5.56%	6.00	108	5.67	2
19	617	5.56%	5.00	90	4.72	2
20	934	5.56%	3.00	54	2.83	2
21	946	5.36%	3.00	56	2.84	2

Figura 5-2  
Tabella e grafico riepilogativi

Tasso di risposta				
Percentile	Tasso di risposta	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

Green Red Caption



Il nuovo insieme di dati contiene i campi seguenti:

- **Codice postale.** Se i gruppi di codici postali si basano solo su una parte del valore completo, questo è il valore di quella parte del codice postale. L'etichetta della riga di intestazione di questa colonna del file Excel è il nome del campo codice postale dell'insieme di dati originale.
- **Tasso di risposta.** La percentuale di risposte positive in ogni codice postale.
- **Risposte.** Il numero di risposte positive in ogni codice postale.

- **Contatti.** Il numero totale di contatti per ogni codice postale che contiene un valore non mancante per il campo risposta.
- **Indice.** La risposta “pesata” in base alla formula  $N \times P \times (1-P)$ , dove  $N$  è il numero di contatti e  $P$  è il tasso di risposta espresso come proporzione.
- **Rango.** Rango di decile (primo 10%, primo 20%, ecc.) dei tassi di risposta in base al codice postale cumulativi in ordine decrescente.

#### ***Tassi di risposta codice postale: considerazioni sui dati***

**Campo Risposta.** Il campo Risposta può essere di tipo stringa o numerico. Se contiene un valore che indica il numero o il valore monetario degli acquisti, sarà necessario creare un nuovo campo in cui tutte le risposte positive siano rappresentate da un solo valore. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale a pag. 33.](#)

**Valore risposta positiva.** Il valore della risposta positiva identifica i clienti che hanno risposto positivamente (ad esempio, che hanno effettuato un acquisto). Si presume che tutti gli altri valori non mancanti relativi alle risposte indichino una risposta negativa. Se sono state definite etichette dei valori per il campo Risposta, tali etichette sono visualizzate nell'elenco a discesa.

**Campo codice postale.** Il campo codice postale può essere di tipo stringa o numerico.

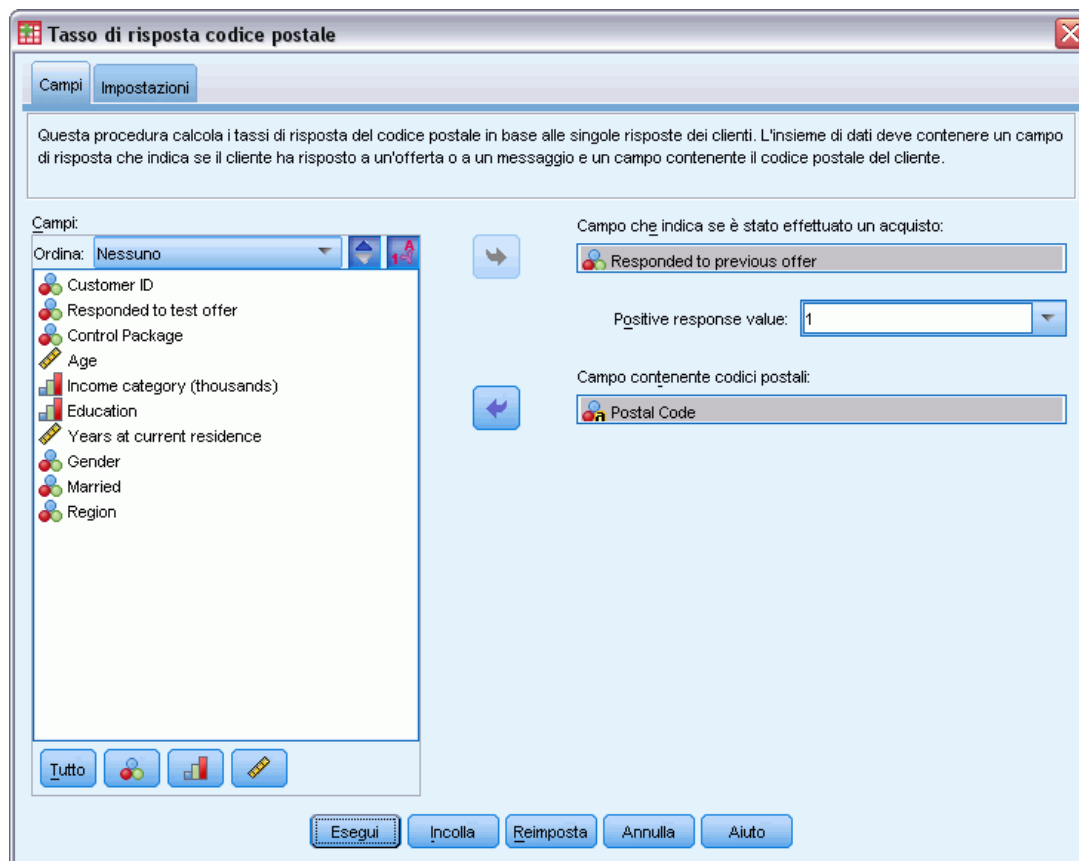
#### ***Per ottenere i tassi di risposta in base al codice postale***

Dai menu, scegliere:

Direct marketing > Scegli tecnica

- ▶ Selezionare Identifica i codici postali con la risposta migliore.

Figura 5-3  
Tassi di risposta codice postale: scheda Campi



- ▶ Selezionare il campo che identifica i contatti che hanno risposto all'offerta.
- ▶ Immettere il valore che indica una risposta positiva. Se per alcuni valori sono state definite delle etichette, è possibile selezionare l'etichetta del valore dall'elenco a discesa per visualizzare il valore corrispondente.
- ▶ Selezionare il campo che contiene il codice postale.
- ▶ Fare clic su Esegui per eseguire la procedura.

In alternativa, è possibile anche:

- Generare tassi di risposta in base ai primi  $n$  caratteri o cifre del codice postale anziché in base al valore completo
- Salvare automaticamente i risultati in un file di Excel
- Controllare le opzioni di visualizzazione dell'output



## Impostazioni

Figura 5-4  
Tassi di risposta codice postale: scheda Impostazioni

### **Raggruppa codici postali in base a**

Determina le modalità di raggruppamento dei record per il calcolo dei tassi di risposta. Per impostazione predefinita viene utilizzato il codice postale per esteso, e tutti i record con lo stesso codice postale vengono raggruppati insieme per calcolare il tasso di risposta del gruppo. In alternativa è possibile raggruppare i record in base a solo una parte del codice postale composta dalle prime  $n$  cifre o caratteri. Ad esempio, è possibile raggruppare i record in base ai primi 5 caratteri di un codice postale di 10 caratteri o in base alle prime tre cifre di un codice postale di 5 cifre. L'insieme di dati di output conterrà un record per ogni gruppo di codici postali. Se si immette un valore, deve essere un numero intero positivo.

### **Formato codice postale numerico**

Se il campo codice postale è numerico e si desidera raggruppare i codici postali in base alle prime  $n$  cifre anziché in base all'intero valore, è necessario specificare il numero di cifre del valore originale. Il numero di cifre è il *massimo* numero possibile di cifre del codice postale. Per

esempio, se il campo codice postale contiene un insieme eterogeneo di codici postali di 5 e 9 cifre, è necessario specificare 9 come numero di cifre.

Nota: a seconda del formato di visualizzazione, alcuni codici postali di 5 cifre possono risultare composti da sole 4 cifre, ma lo zero iniziale è sottointeso.

### **Output**

Oltre al nuovo insieme di dati contenente i tassi di risposta in base al codice postale, è possibile visualizzare una tabella e un grafico che riassumono i risultati per rango di decile (primo 10%, primo 20%, ecc.). La tabella mostra i tassi di risposta, i tassi di risposta cumulativi, il numero di record e il numero cumulativo di record di ogni decile. Il grafico mostra i tassi di risposta cumulativi e il numero cumulativo di record di ogni decile.

**Tasso minimo di risposta accettabile.** Se si inserisce un tasso di risposta obiettivo minimo o una formula di pareggio, la tabella sarà codificata mediante colori per mostrare i decili che soddisfano il tasso di risposta cumulativo minimo, e il grafico includerà una linea di riferimento in corrispondenza del valore del tasso di risposta minimo specificato.

- **Tasso di risposta obiettivo.** Tasso di risposta espresso sotto forma di percentuale (percentuale di risposte positive in ogni gruppo di codici postali). Il valore deve essere maggiore di 0 e minore di 100.
- **Calcola tasso di pareggio dalla formula.** Calcola il tasso di risposta cumulativo minimo in base alla formula:  $(\text{Costo di invio di un pacchetto} / \text{Reddito netto per risposta}) \times 100$ . Entrambi i valori devono essere numeri positivi. Il risultato deve essere un valore maggiore di 0 e minore di 100. Ad esempio, se il costo per l'invio di un pacchetto è di 0,75 € e il reddito netto per risposta è di 56 €, il tasso di risposta minimo è:  $(0,75/56) \times 100 = 1,34\%$ .

**Numero massimo di contatti.** Se si specifica un numero massimo di contatti, la tabella sarà codificata mediante colori per mostrare i decili che non superano il numero massimo cumulativo di contatti (record) e il grafico riporterà una linea di riferimento in corrispondenza di quel valore.

- **Percentuale di contatti.** Massimo espresso come percentuale. Ad esempio, è possibile individuare i decili con i tassi di risposta più elevati che non contengono più del 50% dei contatti totali. Il valore deve essere maggiore di 0 e minore di 100.
- **Numero di contatti.** Massimo espresso come numero di contatti. Ad esempio, se non si prevede di inviare più di 10.000 pacchetti, si può impostare il valore su 10.000. Il valore deve essere un numero intero positivo (senza simboli di raggruppamento).

Se si specifica sia un tasso minimo di risposta accettabile che un numero massimo di contatti, la codifica mediante colori della tabella sarà basata sulla condizione che viene soddisfatta per prima.

**Esporta in Excel**

Questa procedura crea automaticamente un nuovo insieme di dati che contiene i tassi di risposta in base al codice postale. Ogni record (riga) dell'insieme di dati rappresenta un codice postale. Le stesse informazioni si possono salvare automaticamente anche in un file di Excel. Il file viene salvato in formato Excel 97-2003.

**Creazione di un campo Risposta categoriale**

Il campo risposta deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero.

- Se le risposte negative vengono registrate come 0 (non vuoto, che viene considerato valore mancante), il calcolo si può effettuare con la seguente formula:

$$\text{NuovoNome} = \text{VecchioNome} > 0$$

dove *NuovoNome* è il nome del nuovo campo e *VecchioNome* è il nome del campo originale. Questa è un'espressione logica che assegna il valore 1 a tutti i valori non mancanti superiori a 0, e il valore 0 a tutti i valori non mancanti inferiori o uguali a 0.

- Se per le risposte negative non viene registrato alcun valore, questi valori vengono considerati mancanti e la formula è leggermente più complicata:

$$\text{NuovoNome} = \text{NOT}(\text{MISSING}(\text{VecchioNome}))$$

In questa espressione logica, a tutti i valori di risposta non mancanti viene assegnato valore 1 e a tutti i valori di risposta mancanti viene assegnato valore 0.

- Se non è possibile distinguere i valori di risposta negativi (0) dai valori mancanti, non è possibile calcolare un valore di risposta preciso. Se i valori mancanti effettivi sono relativamente pochi, questo può non avere un effetto significativo sui tassi di risposta calcolati. Se, invece, i valori mancanti sono numerosi (ad esempio quando i dati delle risposte vengono registrati solo per un piccolo campione di verifica dell'insieme di dati totale), i tassi di risposta calcolati saranno privi di significato perché saranno notevolmente più bassi rispetto ai tassi di risposta effettivi.

**Per creare un campo Risposta categoriale**

- ▶ Dai menu, scegliere:  
Trasforma > Calcola variabile
- ▶ Per Variabile di destinazione, immettere il nome del nuovo campo (variabile).
- ▶ Se le risposte negative vengono registrate come 0, immettere *VecchioNome*>0 in Espressione numerica, dove *VecchioNome* è il nome del campo originale.
- ▶ Se le risposte negative vengono registrate come mancanti (vuote), immettere NOT(MISSING(*VecchioNome*)) in Espressione numerica, dove *VecchioNome* è il nome del campo originale.

## ***Propensione all'acquisto***

La Propensione all'acquisto utilizza i risultati di un mailing di prova o di campagne precedenti per generare punteggi. I punteggi indicano i contatti che risponderanno con maggiore probabilità. Il campo Risposta indica chi ha risposto all'invio postale di prova o alla campagna precedente. I campi di propensione sono le caratteristiche da utilizzare per prevedere la probabilità che i contatti con caratteristiche simili rispondano.

Questa tecnica utilizza la regressione logistica binaria per creare un modello predittivo. Il processo di creazione e applicazione di un modello predittivo è costituito da due passaggi fondamentali:

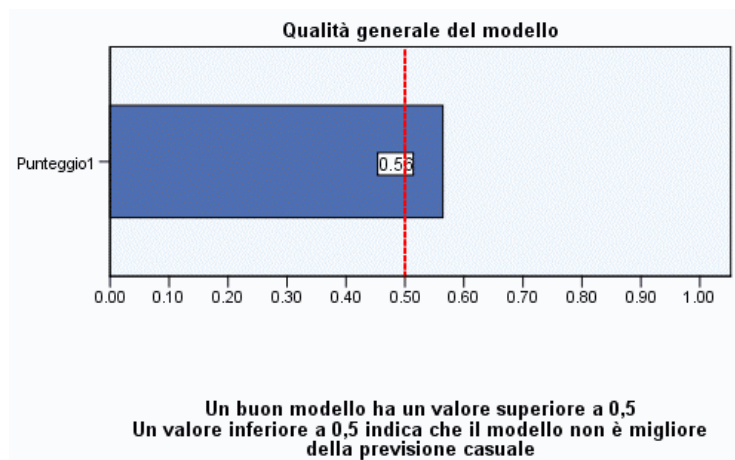
- ▶ Creare il modello e salvare il file del modello. Il modello viene creato utilizzando un insieme di dati per il quale è noto il risultato di interesse (spesso denominato **obiettivo**). Ad esempio, se si desidera creare un modello in grado di generare una previsione di chi risponderà a una campagna di marketing per posta, è necessario iniziare con un insieme di dati che contiene già informazioni su chi ha risposto e chi non ha risposto. Ciò potrebbe essere costituito ad esempio dai risultati di un mailing di prova a un piccolo gruppo di clienti o da informazioni sulle risposte date a una campagna simile in passato.
- ▶ Applicare il modello a diversi insiemi di dati (per i quali non è noto il risultato di interesse) per prevedere i risultati.

**Esempio.** La divisione Direct marketing di una società utilizza i risultati di un mailing di prova per assegnare dei punteggi di propensione al resto dei contatti presenti nel database, identificando i contatti maggiormente propensi a rispondere e a effettuare un acquisto secondo varie caratteristiche demografiche.

### ***Output***

Questa procedura crea automaticamente nell'insieme di dati un nuovo campo contenente i punteggi di propensione per i dati di test e un file di modello XML utilizzabile per calcolare il punteggio di altri insiemi di dati. L'output diagnostico facoltativo comprende un grafico della qualità generale del modello e una tabella di classificazione che confronta le risposte previste con quelle effettive.

Figura 6-1  
Grafico della qualità generale del modello



### ***Propensione all'acquisto: considerazioni sui dati***

**Campo Risposta.** Il campo Risposta può essere di tipo stringa o numerico. Se contiene un valore che indica il numero o il valore monetario degli acquisti, sarà necessario creare un nuovo campo in cui tutte le risposte positive siano rappresentate da un solo valore. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale a pag. 41.](#)

**Valore risposta positiva.** Il valore della risposta positiva identifica i clienti che hanno risposto positivamente (ad esempio, che hanno effettuato un acquisto). Si presume che tutti gli altri valori non mancanti relativi alle risposte indichino una risposta negativa. Se sono state definite etichette dei valori per il campo Risposta, tali etichette sono visualizzate nell'elenco a discesa.

**Prevedi propensione con.** I campi utilizzati per prevedere la propensione possono essere di tipo stringa o numerici e possono inoltre essere nominali, ordinali o continui (scala): è però importante assegnare il livello di misurazione corretto a tutti i campi predittore.

**Livello di misurazione.** L'assegnazione del livello di misurazione corretto è importante perché influisce sul calcolo dei risultati.

- **Nominale.** Una variabile può essere considerata nominale quando i relativi valori rappresentano categorie prive di ordinamento intrinseco, per esempio l'ufficio di una società. Tra gli esempi di variabili nominali troviamo la regione, il codice postale e la religione.
- **Ordinale.** Una variabile può essere considerata ordinale quando i relativi valori rappresentano categorie con qualche ordinamento intrinseco, per esempio i gradi di soddisfazione per un servizio, da molto insoddisfatto a molto soddisfatto, i punteggi di atteggiamento corrispondenti a gradi di soddisfazione o fiducia e i punteggi di preferenza.
- **Continuo.** Una variabile può essere considerata di scala (continua) quando i relativi valori rappresentano categorie ordinate con una metrica significativa, tale che i confronti fra le distanze dei relativi valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

Un'icona accanto a ciascun campo indica il livello di misurazione corrente.

Livello di misurazione	Tipo di dati			
	Numerico	Stringa	Data	Ora
Scala (continuo)		n/d		
Ordinale				
Nominale				

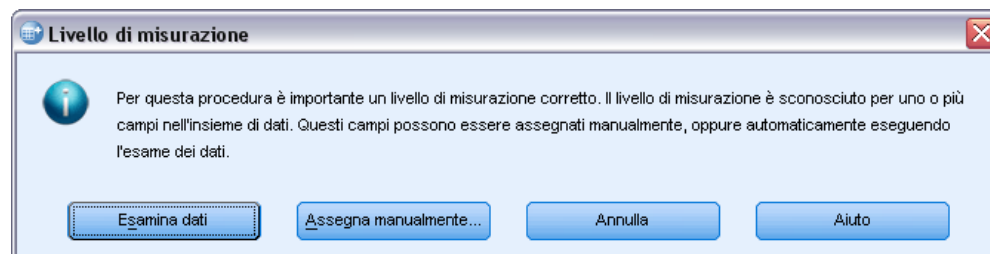
È possibile modificare il livello di misurazione nella Visualizzazione variabili dell'Editor dei dati ; in alternativa, è possibile utilizzare la finestra di dialogo Definisci proprietà variabili per indicare un livello di misurazione appropriato per ogni campo.

### **Campi con livello di misurazione sconosciuto**

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) dell'insieme di dati è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Figura 6-2

*Avviso Livello di misurazione*



- **Esamina dati.** Legge i dati dell'insieme di dati attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con insiemi di dati di grandi dimensioni, questa operazione può richiedere del tempo.
- **Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Visualizzazione variabili dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

### Per ottenere i punteggi di propensione all'acquisto

Dai menu, scegliere:

Direct Marketing > Scegli tecnica

- Selezionare Seleziona contatti con maggiore probabilità di acquisto.

Figura 6-3

Propensione all'acquisto: scheda Campi

**Propensione all'acquisto**

**Campi** | Impostazioni

Questa tecnica utilizza i risultati di un invio postale di prova o di una campagna precedente per generare i punteggi. I punteggi indicano i contatti che risponderanno con maggiore probabilità. Il campo Risposta indica chi ha risposto all'invio postale di prova o alla campagna precedente. I campi di propensione sono le caratteristiche da utilizzare per prevedere la probabilità che i contatti con caratteristiche simili rispondano.

**Campi:**

Ordina: Nessuno

- ID cliente
- Ha risposto all'offerta precedente
- Pacchetto di controllo
- Codice postale
- Sequenza

**Campo risposta:**

Ha risposto all'offerta di prova

Valore risposta positiva:

**Prevedi propensione con:**

- Età
- Categoria di reddito (migliaia)
- Istruzione
- Anni alla residenza attuale
- Sesso
- Coniugato
- Figli
- Regione

**Save Model**

Export model information to XML file

File name:  **Sfogli**

**Esegui** **Incolla** **Reimposta** **Annulla** **Aiuto**

- Selezionare il campo che identifica i contatti che hanno risposto all'offerta.
- Immettere il valore che indica una risposta positiva. Se per alcuni valori sono state definite delle etichette, è possibile selezionare l'etichetta del valore dall'elenco a discesa per visualizzare il valore corrispondente.

- ▶ Selezionare i campi che si desidera utilizzare per prevedere la propensione.  
Per salvare un file di modello XML per calcolare il punteggio di altri file di dati:
- ▶ Selezionare Esporta informazioni modello in file XML.
- ▶ Digitare il percorso di una directory e un nome file o fare clic su Sfoglia per passare alla cartella dove si desidera salvare il file di modello XML.
- ▶ Fare clic su Esegui per eseguire la procedura.  
Per utilizzare un file di modello per calcolare il punteggio di altri insiemi di dati:
- ▶ Aprire l'insieme di dati per il quale si desidera calcolare il punteggio.
- ▶ Utilizzare la procedura guidata di calcolo del punteggio per applicare il modello all'insieme di dati. Dai menu, scegliere:  
Utilità > Procedura guidata di calcolo del punteggio.



## Impostazioni

Figura 6-4  
Propensione all'acquisto: scheda Impostazioni

**Propensione all'acquisto**

**Campi** **Impostazioni**

**Convalida del modello**

È possibile convalidare il modello utilizzato per generare i punteggi. Per convalidare il modello, è necessario suddividere i dati in partizioni. La partizione di training viene utilizzata per formare, o costruire, il modello. La partizione di test viene utilizzata per convalidare il modello. Se si desidera convalidare il modello, questa tecnica assegna automaticamente i record alle partizioni.

Convalida il modello

Dimensioni partizione campione di training (%):

Imposta seme per replicare i risultati

Seme:

**Output diagnostico**

Qualità generale del modello

Tabella classificazioni

Probabilità minima:

**Nome ed etichetta del campo risposta ricodificata**

Questa tecnica ricodifica automaticamente il campo risposta in un nuovo campo nel quale 1 rappresenta le risposte positive e 0 rappresenta le risposte negative.

Nome nuovo campo:

Etichetta nuovo campo:

**Salva punteggi**

Questa tecnica utilizza i risultati di un invio postale di prova o di una campagna precedente per generare i punteggi. I punteggi vengono automaticamente salvati per consentirne l'uso. Gli altri controlli della scheda consentono di controllare ulteriormente i dati salvati.

Nuovo nome campo per i punteggi:

### Convalida del modello

La convalida del modello crea dei gruppi di training e di test a scopo diagnostico. Se si seleziona la tabella di classificazione della sezione Output diagnostico, la tabella viene suddivisa in una sezione di training (selezionata) e una di verifica (non selezionata) a scopo di confronto. Non selezionare la convalida del modello se non si seleziona anche la tabella di classificazione. I punteggi si basano sul modello generato dal campione di training, che conterrà sempre un numero di record inferiore al numero totale dei record disponibili. Ad esempio, le dimensioni predefinite del campione di training sono pari al 50%, e un modello creato in base alla metà dei record disponibili può non essere altrettanto affidabile di un modello creato in base a tutti i record disponibili.

- **Dimensioni partizione campione di training (%).** Specificare la percentuale di record da assegnare al campione di training. Il resto dei record con valori non mancanti per il campo Risposta viene assegnato al campione di verifica. Il valore deve essere maggiore di 0 e minore di 100.
- **Imposta seme per replicare i risultati.** Dal momento che i record vengono assegnati in modo casuale ai campioni di training e di verifica, ogni volta che si esegue la procedura si possono ottenere risultati diversi a meno di non specificare sempre lo stesso valore di seme del numero casuale di partenza.

### ***Output diagnostico***

**Qualità generale del modello.** Visualizza un grafico a barre della qualità generale del modello, espressa come valore compreso fra 0 e 1. Un buon modello deve avere un valore superiore a 0,5.

**Tabella classificazioni.** Mostra una tabella che confronta le risposte positive e negative previste alle risposte positive e negative effettive. La percentuale di precisione generale può fornire un'indicazione del buon funzionamento del modello, ma la percentuale di risposte positive corrette prevista può essere un dato più interessante.

- **Probabilità minima.** Assegna i record con un punteggio maggiore del valore specificato alla categoria di risposta positiva prevista nella tabella di classificazione. I punteggi generati dalla procedura rappresentano la probabilità che il contatto dia una risposta positiva (ad esempio, che effettui un acquisto). Come regola generale è bene specificare un valore vicino a quello del tasso di risposta obiettivo minimo, espresso come proporzione. Ad esempio, se quello che interessa è un tasso di risposta di almeno il 5%, specificare 0,05. Il valore deve essere maggiore di 0 e minore di 1.

### ***Nome ed etichetta del campo risposta ricodificata***

Questa procedura ricodifica automaticamente il campo Risposta in un nuovo campo nel quale 1 rappresenta le risposte positive e 0 rappresenta le risposte negative. L'analisi viene eseguita sul campo ricodificato. È possibile ignorare il nome e l'etichetta predefiniti e specificarne di diversi. I nomi devono essere conformi alle regole di denominazione di IBM® SPSS® Statistics.

### ***Salva punteggi***

Un nuovo campo contenente i punteggi di propensione viene salvato automaticamente nell'insieme di dati originale. I punteggi rappresentano la probabilità di una risposta positiva, espressa come proporzione.

- I nomi dei campi devono essere conformi alle regole di denominazione di SPSS Statistics.
- Il nome del campo non può essere uguale al nome di un altro campo già esistente nell'insieme di dati. Se si esegue questa procedura più di una volta sullo stesso insieme di dati sarà necessario specificare ogni volta un nome diverso.

## Creazione di un campo Risposta categoriale

Il campo risposta deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero.

- Se le risposte negative vengono registrate come 0 (non vuoto, che viene considerato valore mancante), il calcolo si può effettuare con la seguente formula:

$$\text{NuovoNome} = \text{VecchioNome} > 0$$

dove *NuovoNome* è il nome del nuovo campo e *VecchioNome* è il nome del campo originale. Questa è un'espressione logica che assegna il valore 1 a tutti i valori non mancanti superiori a 0, e il valore 0 a tutti i valori non mancanti inferiori o uguali a 0.

- Se per le risposte negative non viene registrato alcun valore, questi valori vengono considerati mancanti e la formula è leggermente più complicata:

$$\text{NuovoNome} = \text{NOT}(\text{MISSING}(\text{VecchioNome}))$$

In questa espressione logica, a tutti i valori di risposta non mancanti viene assegnato valore 1 e a tutti i valori di risposta mancanti viene assegnato valore 0.

- Se non è possibile distinguere i valori di risposta negativi (0) dai valori mancanti, non è possibile calcolare un valore di risposta preciso. Se i valori mancanti effettivi sono relativamente pochi, questo può non avere un effetto significativo sui tassi di risposta calcolati. Se, invece, i valori mancanti sono numerosi (ad esempio quando i dati delle risposte vengono registrati solo per un piccolo campione di verifica dell'insieme di dati totale), i tassi di risposta calcolati saranno privi di significato perché saranno notevolmente più bassi rispetto ai tassi di risposta effettivi.

### Per creare un campo Risposta categoriale

- ▶ Dai menu, scegliere:  
Trasforma > Calcola variabile
- ▶ Per Variabile di destinazione, immettere il nome del nuovo campo (variabile).
- ▶ Se le risposte negative vengono registrate come 0, immettere *VecchioNome*>0 in Espressione numerica, dove *VecchioNome* è il nome del campo originale.
- ▶ Se le risposte negative vengono registrate come mancanti (vuote), immettere *NOT(MISSING(VecchioNome))* in Espressione numerica, dove *VecchioNome* è il nome del campo originale.

## ***Test pacchetto di controllo***

Questa tecnica confronta le campagne di marketing per rilevare eventuali differenze significative nell'efficacia di diversi pacchetti o offerte. L'efficacia di una campagna viene misurata in base alle risposte. Il Campo campagna identifica le diverse campagne, ad esempio Offerta A e Offerta B. Il Campo risposta indica se un contatto ha risposto alla campagna. Selezionare Importo acquisto quando la risposta viene registrata come importo di acquisto, ad esempio "99,99". Selezionare Replica se la risposta indica semplicemente se il contatto ha risposto positivamente o meno, ad esempio "Sì" o "No".

**Esempio.** La divisione Direct marketing di una società desidera verificare se il nuovo design di una confezione genererà un numero maggiore di risposte positive rispetto alla confezione esistente. Invia quindi un mailing di prova per determinare se la nuova confezione generi un aumento significativo del tasso di risposta positiva. Il mailing di prova consiste di un gruppo di controllo che riceve la confezione esistente e un gruppo di prova che riceve la confezione con il nuovo design. I risultati dei due gruppi vengono confrontati per rilevare una eventuale differenza significativa.

### ***Output***

Output include una tabella che mostra conteggi e percentuali di risposte positive e negative per ogni gruppo definito dal Campo campagna e una tabella che identifica i gruppi che differiscono in modo significativo dagli altri.

Figura 7-1  
*Output di Test pacchetto di controllo*

		Pacchetto di controllo			
		Controllo		Test	
		Conteggio	% casi colonna	Conteggio	% casi colonna
Efficacia (1=Si 0=No)	0	875	96.2%	945	93.8%
	1	35	3.8%	62	6.2%

È presente una differenza statisticamente significativa tra Controllo e Test.

### ***Considerazioni e assunzioni relative ai dati del test pacchetto di controllo***

**Campo campagna.** Il Campo campagna deve essere categoriale (nominale o ordinale).

**Campo risposta efficacia.** Se si seleziona Importo acquisto per il campo Efficacia, il campo deve essere numerico e il livello di misurazione deve essere continuo (scala).

Se non è possibile distinguere i valori di risposta negativi (per l'importo di acquisto un valore di 0) dai valori mancanti, non è possibile calcolare un tasso di risposta preciso. Se i valori mancanti effettivi sono relativamente pochi, questo può non avere un effetto significativo sui tassi di risposta calcolati. Se, invece, i valori mancanti sono numerosi (ad esempio quando i dati delle risposte vengono registrati solo per un piccolo campione di verifica dell'insieme di dati totale), i tassi di risposta calcolati saranno privi di significato perché saranno notevolmente più bassi rispetto ai tassi di risposta effettivi.

**Assunzioni.** Questa procedura presuppone che i contatti siano stati assegnati a ciascun gruppo della campagna in modo casuale. In altre parole, l'assegnazione ai gruppi non viene determinata da alcuna specifica caratteristica demografica, di acquisto o di altro tipo e tutti i contatti hanno la stessa probabilità di assegnazione a qualsiasi gruppo.

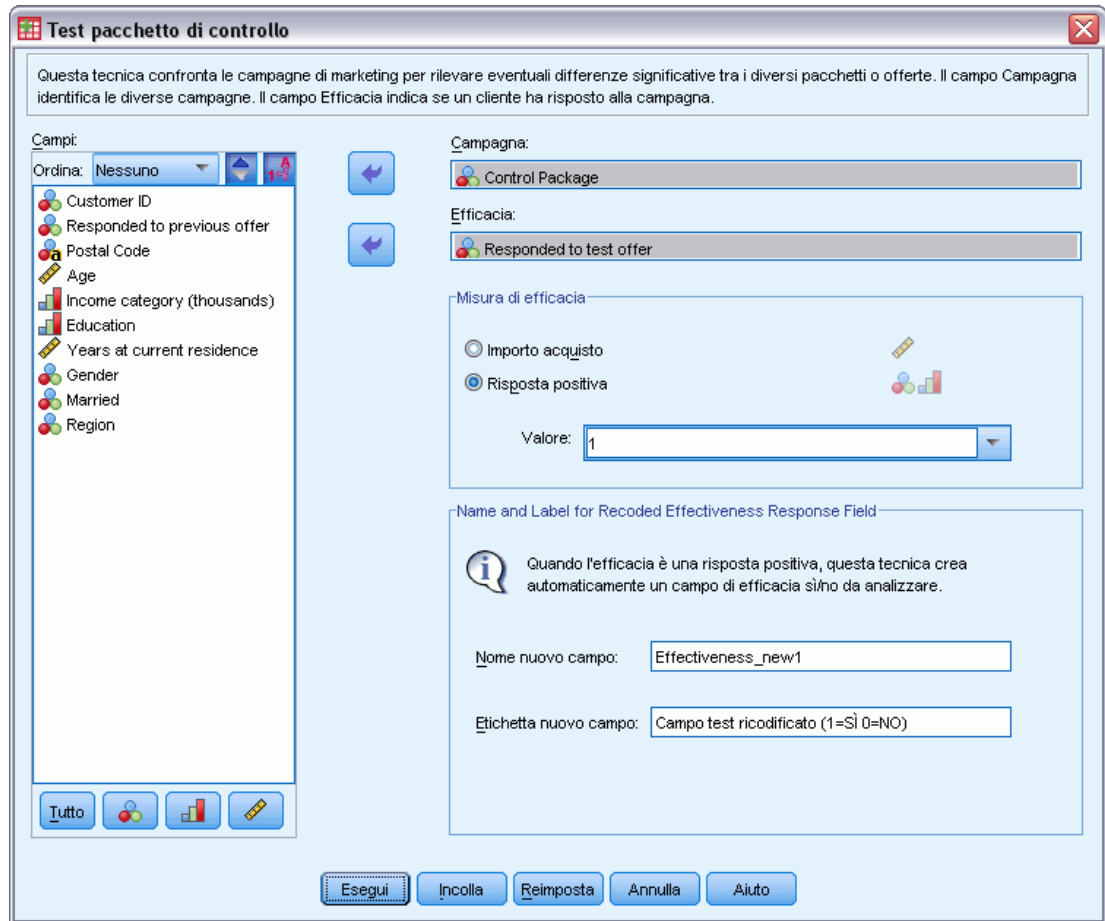
***Per ottenere un test pacchetto di controllo***

Dai menu, scegliere:

Direct marketing > Scegli tecnica

- ▶ Selezionare Confronta efficacia delle campagne.

Figura 7-2  
Finestra di dialogo Test pacchetto di controllo



- Selezionare il campo che identifica il gruppo a cui appartiene ogni contatto (ad esempio offerta A, offerta B e così via). Questo campo deve essere nominale o ordinale.
- Selezionare il campo che indica l'efficacia della risposta.

Se corrisponde a un importo di acquisto, il campo della risposta deve essere numerico.

Se il campo della risposta indica semplicemente se il contatto ha risposto positivamente o meno, ad esempio "Sì" o "No", selezionare Replica e immettere il valore che rappresenta una risposta positiva. Se per alcuni valori sono state definite delle etichette, è possibile selezionare l'etichetta del valore dall'elenco a discesa per visualizzare il valore corrispondente.

Viene automaticamente creato un nuovo campo, nel quale 1 rappresenta le risposte positive e 0 rappresenta le risposte negative. L'analisi viene eseguita sul nuovo campo. È possibile ignorare il nome e l'etichetta predefiniti e specificarne di diversi. I nomi devono essere conformi alle regole di denominazione di IBM® SPSS® Statistics.

- Fare clic su Esegui per eseguire la procedura.

# ***Parte II: Esempi***

# Analisi RFM dei dati delle transazioni

In un file di dati delle transazioni, ogni riga rappresenta una transazione separata, non un cliente separato, e possono essere presenti più righe delle transazioni per ciascun cliente. Questo esempio utilizza il file di dati *rfm\_transactions.sav*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A a pag. 100](#).

## Dati della transazione

Le variabili presenti nell'insieme di dati devono contenere le seguenti informazioni:

- Una variabile o combinazione di variabili che identifica ciascun caso (cliente).
- Una variabile con la data di ciascuna transazione.
- Una variabile con il valore monetario di ciascuna transazione.

Figura 8-1

*Dati della transazione RFM*

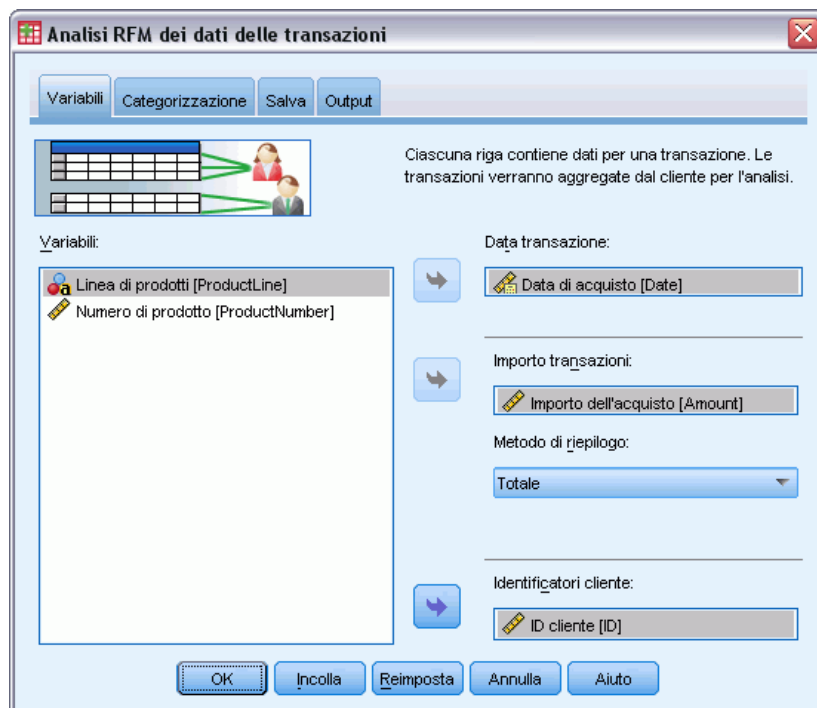
ID	Date	Amount
1	08/04/2005	129
1	10/25/2004	50
1	07/24/2004	118
1	07/24/2004	136
1	09/04/2006	52
2	09/23/2005	183
2	11/05/2004	24
2	11/10/2005	66
2	12/03/2004	77
3	06/04/2005	102
3	05/15/2005	131

## Esecuzione dell'analisi

- ▶ Per calcolare i punteggi RFM, dai menu scegliere:  
Direct marketing > Scegli tecnica
- ▶ Selezionare Aiutami a individuare i contatti migliori (Analisi RFM) e fare clic su Continua.
- ▶ Fare clic su Dati transazioni, quindi fare clic su Continua.



Figura 8-2  
RFM dei dati della transazione, scheda Variabili



- ▶ Fare clic su Ripristina per annullare qualsiasi impostazione precedente.
- ▶ Per Data transazione, selezionare *Data di acquisto [Data]*.
- ▶ Per Importo transazioni, selezionare *Importo dell'acquisto [Importo]*.
- ▶ Per Metodo di riepilogo, selezionare Totale.
- ▶ Per Identificatori cliente, selezionare *ID cliente [ID]*.
- ▶ Quindi fare clic sulla scheda Output.

Figura 8-3  
Analisi RFM dei dati delle transazioni, scheda Output



- ▶ Selezionare Grafico dei conteggi dei bin.
- ▶ Quindi fare clic su OK per eseguire la procedura.

## Valutazione dei risultati

Quando si calcolano i punteggi RFM dai dati della transazione, viene creato un nuovo insieme di dati che comprende i nuovi punteggi RFM.

Figura 8-4  
Insieme di dati RFM dai dati della transazione

ID	Data_più_rec ente	Conteggio_t ransazioni	Importo	Recency_ score	Frequency_ score	Monetary_ score	RFM_score
1	04-Sep-2006	5	485,00	4	3	4	434
2	10-Nov-2005	4	350,00	2	2	2	222
3	04-Jun-2005	2	233,00	1	2	4	124
4	18-Aug-2006	7	936,00	4	4	5	445
5	07-Jul-2006	3	359,00	4	1	5	415
6	16-Jul-2006	3	249,00	4	1	4	414
7	15-Feb-2006	7	1089,00	2	5	5	255

Per impostazione predefinita, l'insieme di dati comprende le informazioni seguenti su ogni cliente:

- Variabili ID cliente
- Data della transazione più recente

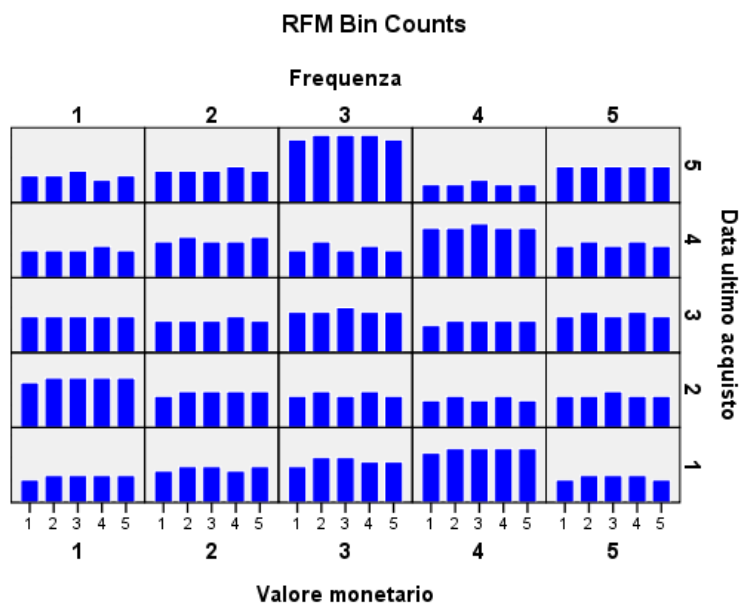
- Numero totale di transazioni
- Importo riassuntivo della transazione (l'impostazione predefinita è il totale)
- Punteggi RFM data ultimo acquisto, frequenza, valore monetario e combinato

Il nuovo insieme di dati contiene solo una riga (record) per ogni cliente. I dati originali della transazione sono stati aggregati tramite i valori delle variabili identificatrici cliente. Le variabili identificatrici sono sempre incluse nel nuovo insieme di dati, altrimenti non sarebbe possibile fare corrispondere i punteggi RFM ai clienti.

Il punteggio RFM combinato per ogni cliente è semplicemente la concatenazione dei tre punteggi individuali, calcolati nel seguente modo: (data dell'ultimo acquisto x 100) + (frequenza x 10) + valore monetario.

Il grafico dei conteggi dei bin visualizzato nel Viewer mostra il numero di clienti in ogni categoria RFM.

Figura 8-5  
Grafico dei conteggi dei bin



Utilizzando il metodo predefinito di cinque categorie di punteggi per ciascuno dei tre componenti RFM, il risultato è 125 categorie possibili di punteggio RFM. Ogni barra del grafico rappresenta il numero di clienti di ogni categoria RFM.

Idealmente conviene avere una distribuzione di clienti relativamente uniforme su tutte le categorie di punteggi RFM. In realtà, si otterrà in genere una certa quantità di variazione, come quella risultante in questo esempio. Se sono presenti molte categorie vuote, conviene modificare il metodo di categorizzazione.

Esistono diverse strategie per gestire le distribuzioni non uniformi dei punteggi RFM, tra cui:

- Utilizzare la categorizzazione nidificata al posto di quella indipendente.

- Ridurre il numero di categorie di punteggi possibili (bin).
- In presenza di molti valori a pari merito, assegnare in modo casuale i casi con gli stessi punteggi a categorie diverse.

Per ulteriori informazioni, vedere l'argomento [Categorizzazione RFM in il capitolo 2 a pag. 6.](#)

## Unire i dati sui punteggi ai dati sui clienti

Ora che si è ottenuto un insieme di dati contenente i punteggi RFM, è necessario fare corrispondere tali punteggi ai clienti. Si potrebbero unire nuovamente i punteggi al file di dati delle transazioni, ma generalmente è preferibile unire i dati dei punteggi a un file di dati che, come l'insieme dei dati sui punteggi RFM, contenga una riga (record) per ogni cliente, oltre a informazioni quali il nome e l'indirizzo del cliente.

**Figura 8-6**

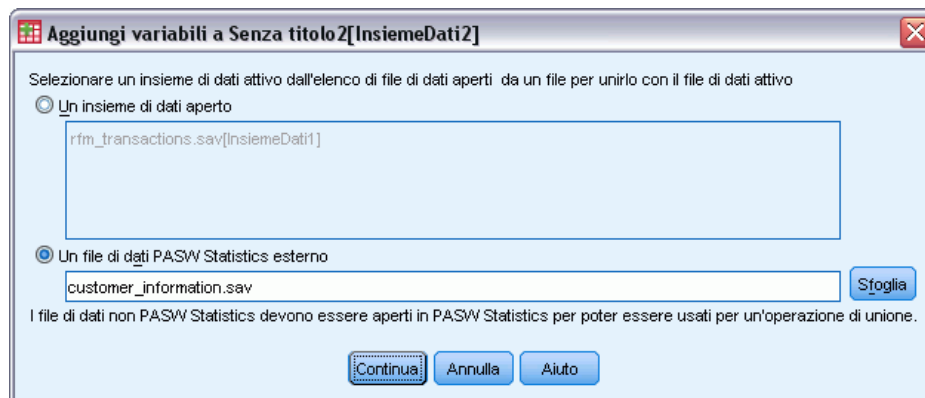
*Insieme di dati sui punteggi RFM nella Visualizzazione variabili*

Nome	Tipo	Larghezza	Decimali	Etichetta	Valori
ID	Numerico	5	0		Nessuno/a
Data_più_recente	Data	10	0	Data della transazio...	Nessuno/a
Conteggio_transazioni	Numerico	7	0	Numero di transazioni	Nessuno/a
Importo	Numerico	8	2	Importo	Nessuno/a
Recency_score	Numerico	3	0	Punteggio data ulti...	Nessuno/a
Frequency_score	Numerico	3	0	Punteggio frequenza	Nessuno/a
Monetary_score	Numerico	3	0	Punteggio valore mo...	Nessuno/a
RFM_score	Numerico	3	0	Punteggio RFM	Nessuno/a

- ▶ Rendere attivo l'insieme di dati contenente i punteggi RFM (fare clic in qualsiasi punto della finestra dell'Editor dei dati contenente l'insieme di dati).
- ▶ Dai menu, scegliere:  
Dati > Unisci file > Aggiungi variabili

**Figura 8-7**

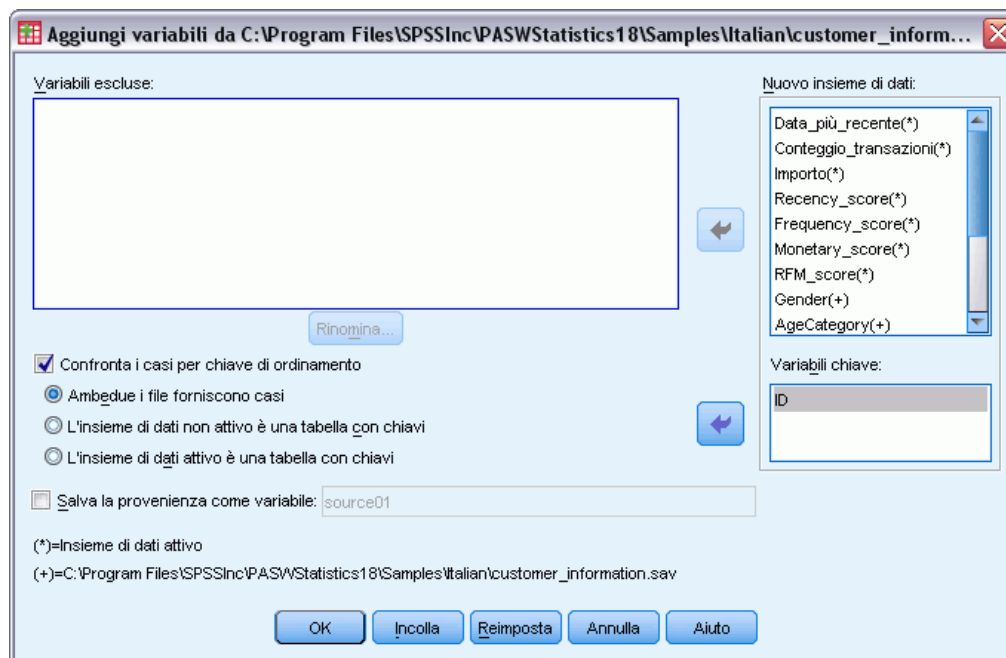
*Finestra di dialogo Aggiungi variabili, selezione dei file*



- ▶ Selezionare Un file di dati esterno.

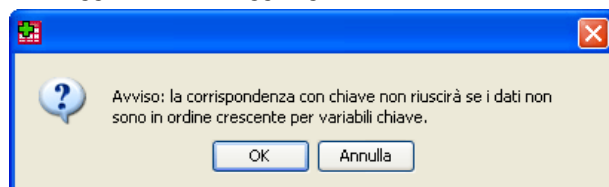
- ▶ Utilizzare il pulsante Sfoglia per andare nella cartella *Campioni* e selezionare *customer\_information.sav*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A a pag. 100](#).
- ▶ Fare clic su Continua.

Figura 8-8

Finestra di dialogo *Aggiungi variabili*, selezione delle variabili

- ▶ Selezionare *Confronta i casi per chiave di ordinamento*.
- ▶ Selezionare *Ambedue i file forniscono casi*.
- ▶ Selezionare *ID* per la lista *Variabili chiave*.
- ▶ Fare clic su *OK*.

Figura 8-9

Messaggio di avviso *Aggiungi variabili*

Si noti il messaggio che avvisa che entrambi i file devono essere in ordine crescente per variabili chiave. In questo esempio entrambi i file sono già in ordine crescente per variabile chiave, ovvero la variabile identificatrice cliente selezionata durante il calcolo dei punteggi RFM. Quando si calcolano i punteggi RFM dai dati della transazione, il nuovo insieme di dati viene ordinato automaticamente in ordine crescente per variabili identificatrici cliente. Se si cambia l'ordine crescente dell'insieme dei dati sui punteggi o se il file di dati al quale si desidera unire l'insieme

dei dati sui punteggi non è in ordine crescente, è necessario per prima cosa ordinare entrambi i file in ordine crescente per variabili identificatrici cliente.

- Fare clic su OK per unire i due insieme di dati.

L'insieme di dati contenente i punteggi RFM contiene ora anche il nome, l'indirizzo e altre informazioni su ogni cliente.

**Figura 8-10**  
*Insiemi di dati uniti*

Nome	Tipo	Larghezza	Decimali	Etichetta	Valori
ID	Numerico	5	0	Customer ID	Nessuno/a
Data_più_recente	Data	10	0	Data della transazione più ...	Nessuno/a
Conteggio_transazioni	Numerico	7	0	Numero di transazioni	Nessuno/a
Importo	Numerico	8	2	Importo	Nessuno/a
Recency_score	Numerico	3	0	Punteggio data ultimo acq...	Nessuno/a
Frequency_score	Numerico	3	0	Punteggio frequenza	Nessuno/a
Monetary_score	Numerico	3	0	Punteggio valore monetario	Nessuno/a
RFM_score	Numerico	3	0	Punteggio RFM	Nessuno/a
Gender	Numerico	4	0		{0, Female}...
AgeCategory	Numerico	7	0	Age Category	{1, <25}...
Name	Stringa	4	0		Nessuno/a
Address	Stringa	14	0		Nessuno/a
City	Stringa	11	0		Nessuno/a
State_Province	Stringa	7	0	State/Province	Nessuno/a
PostalCode	Stringa	1	0	Postal Code	Nessuno/a
Country	Stringa	1	0		Nessuno/a

# Analisi cluster

L'analisi cluster è uno strumento esplorativo progettato per rivelare i raggruppamenti naturali (o cluster) all'interno dei dati. Ad esempio, è in grado di identificare gruppi diversi di clienti in base a varie caratteristiche demografiche e di acquisto.

Per esempio, la divisione Direct marketing di una società desidera identificare i raggruppamenti demografici all'interno del database clienti per determinare più facilmente le strategie delle campagne di marketing e sviluppare nuovi prodotti.

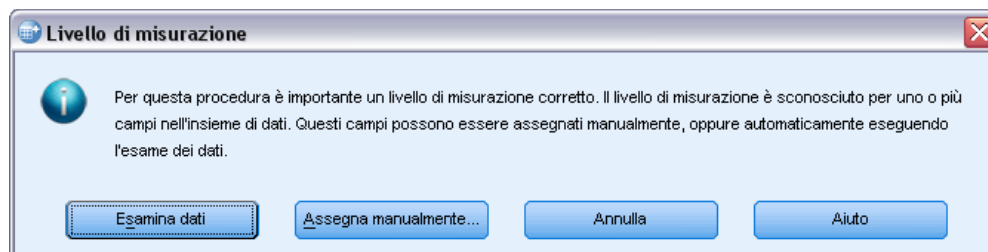
Tali informazioni vengono raccolte nel file *dmdata.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)

## Esecuzione dell'analisi

- ▶ Per eseguire un'analisi cluster, dai menu scegliere:  
Direct marketing > Scegli tecnica
- ▶ Selezionare Segmenta i miei contatti in cluster e fare clic su Continua.

L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) dell'insieme di dati è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Figura 9-1  
Avviso Livello di misurazione



- **Esamina dati.** Legge i dati dell'insieme di dati attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con insiemi di dati di grandi dimensioni, questa operazione può richiedere del tempo.
- **Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Visualizzazione variabili dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

In questo file di esempio non vi sono campi con livello di misurazione sconosciuto e tutti i campi hanno il livello di misurazione corretto; quindi, l'avviso relativo al livello di misurazione non dovrebbe essere visualizzato.

Figura 9-2  
Scheda Campi di Analisi cluster

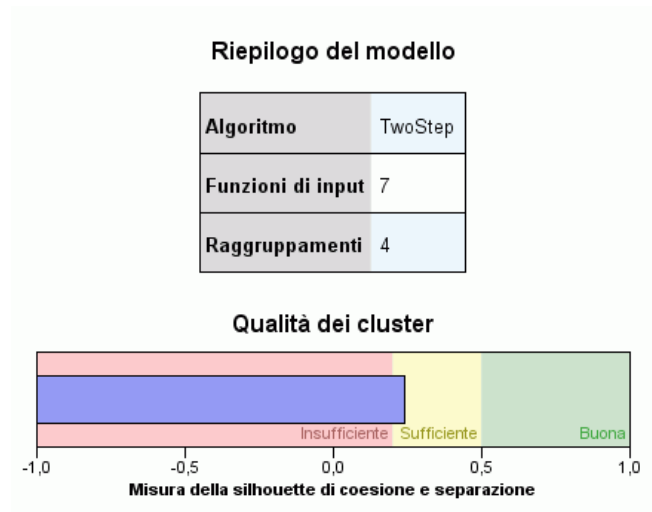


- ▶ Selezionare i campi seguenti per creare i segmenti: *Età*, *Categoria di reddito*, *Istruzione*, *Anni alla residenza attuale*, *Sesso*, *Coniugato/a* e *Figli*.
- ▶ Fare clic su *Esegui* per eseguire la procedura.



## Output

Figura 9-3  
Riepilogo del modello di cluster



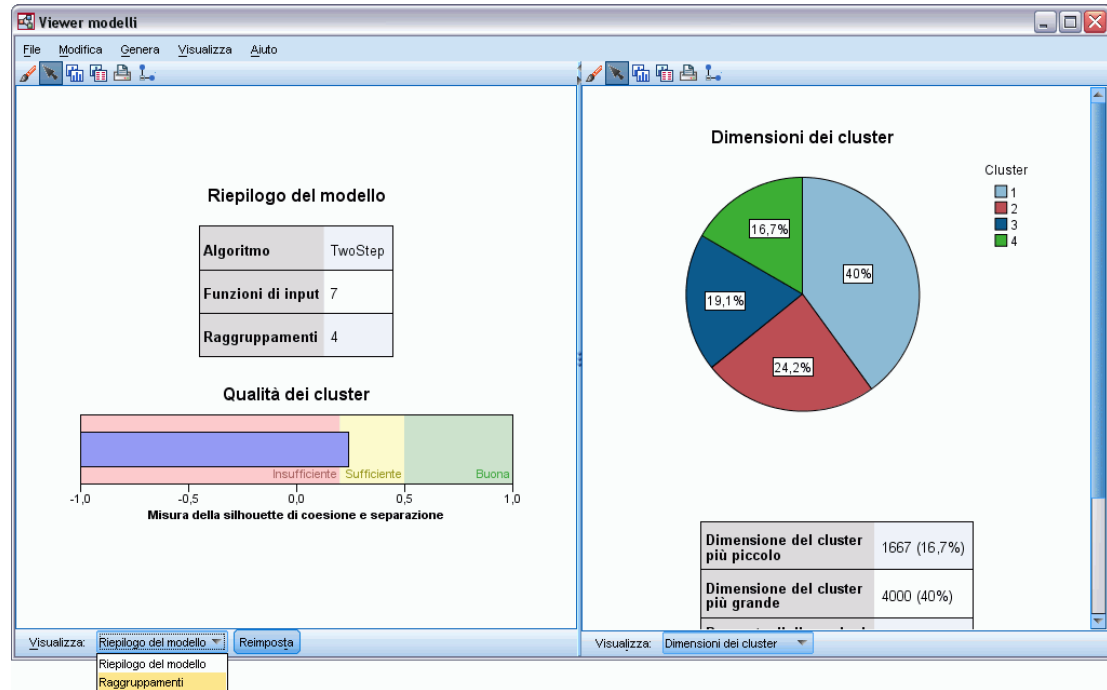
I risultati vengono visualizzati nel Viewer dei modelli di cluster.

- Il riepilogo del modello mostra che sono stati individuati quattro cluster in base alle sette funzioni di input (campi) selezionate.
- Il grafico sulla qualità dei cluster indica che la qualità generale del modello si trova a metà dell'intervallo "Adatto".

- Fare doppio clic sull'output del Viewer dei modelli di cluster per attivare il Viewer modelli.

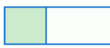



Figura 9-4

Viewer modelli di cluster attivato



- Dall'elenco a discesa Visualizza nella parte inferiore della finestra del Viewer dei modelli di cluster, selezionare Cluster.

Figura 9-5  
Visualizzazione Cluster

Cluster	1	2	3	4
Etichetta				
Descrizione				
Dimensione	 40,0% (4000)	 24,2% (2424)	 19,1% (1909)	 16,7% (1667)
Funzioni	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

La visualizzazione Cluster mostra le informazioni relative agli attributi dei singoli cluster.

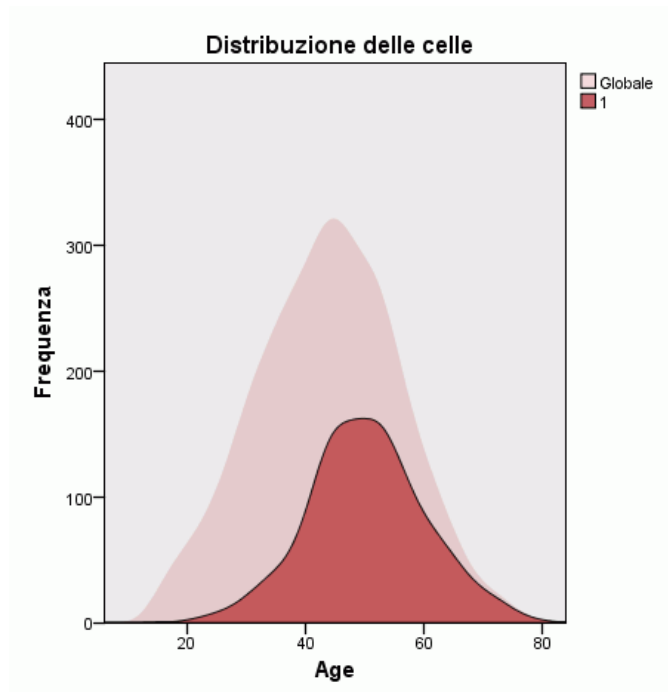
- Per i campi continui (scala) viene visualizzato il valore medio (media).
- Per i campi categoriali (nominali, ordinali) viene visualizzata la moda. La moda è la categoria con il maggior numero di record. In questo esempio, ogni record corrisponde a un cliente.
- Per impostazione predefinita, i campi sono visualizzati in ordine di importanza globale per il modello. In questo esempio, *Età* è il campo con la massima importanza globale. I campi si possono ordinare anche per importanza all'interno del cluster o per ordine alfabetico.

Se si seleziona (si fa clic su) qualsiasi cella nella visualizzazione Cluster è possibile visualizzare un grafico che riassume i valori di quel campo per un determinato cluster.

- Per esempio, selezionare la cella *Età* del cluster 1.

Figura 9-6

*Istogramma dell'età per il cluster 1*

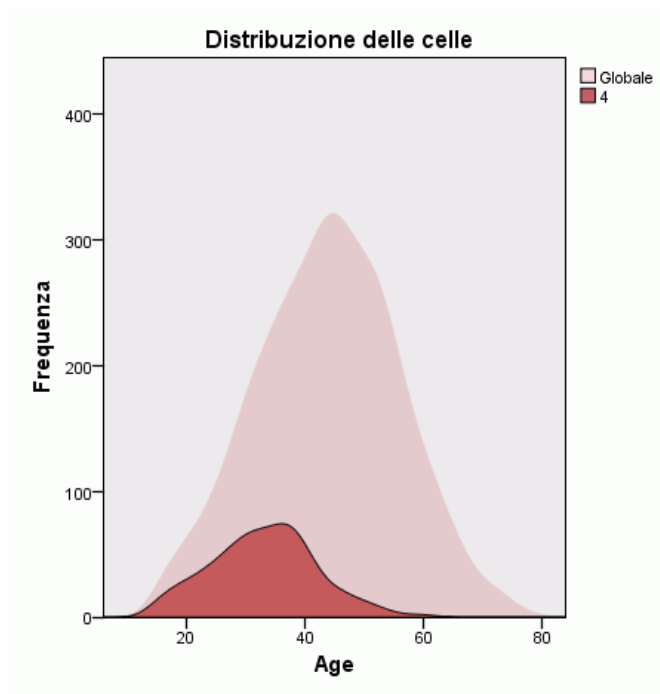


Per i campi continui viene visualizzato un istogramma che mostra sia la distribuzione dei valori all'interno del cluster, sia la distribuzione globale dei valori per il campo. L'istogramma indica che i clienti del cluster 1 tendono a essere leggermente più vecchi.

- Selezionare la cella *Età* per il cluster 4 nella visualizzazione Cluster.

Figura 9-7

*Istogramma dell'età per il cluster 4*

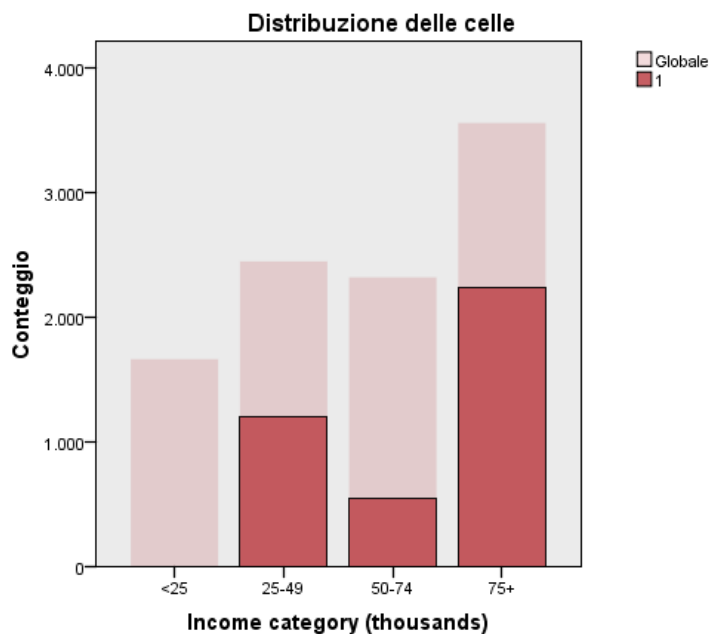


Al contrario del cluster 1, i clienti del cluster 4 tendono a essere più giovani rispetto alla media globale.

- Selezionare la cella *Categoria di reddito* per il cluster 1 nella visualizzazione Cluster.

Figura 9-8

Grafico a barre delle categorie di reddito per il cluster 1

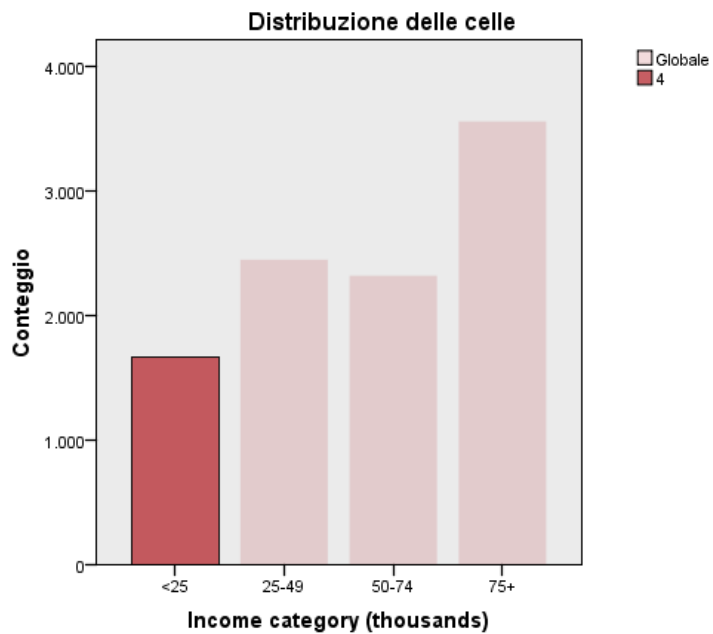


Per i campi categoriali viene visualizzato un grafico a barre. La caratteristica saliente del grafico a barre della categoria di reddito per questo cluster è la totale assenza di clienti nella categoria di reddito più basso.

- Selezionare la cella *Categoria di reddito* per il cluster 4 nella visualizzazione Cluster.

Figura 9-9

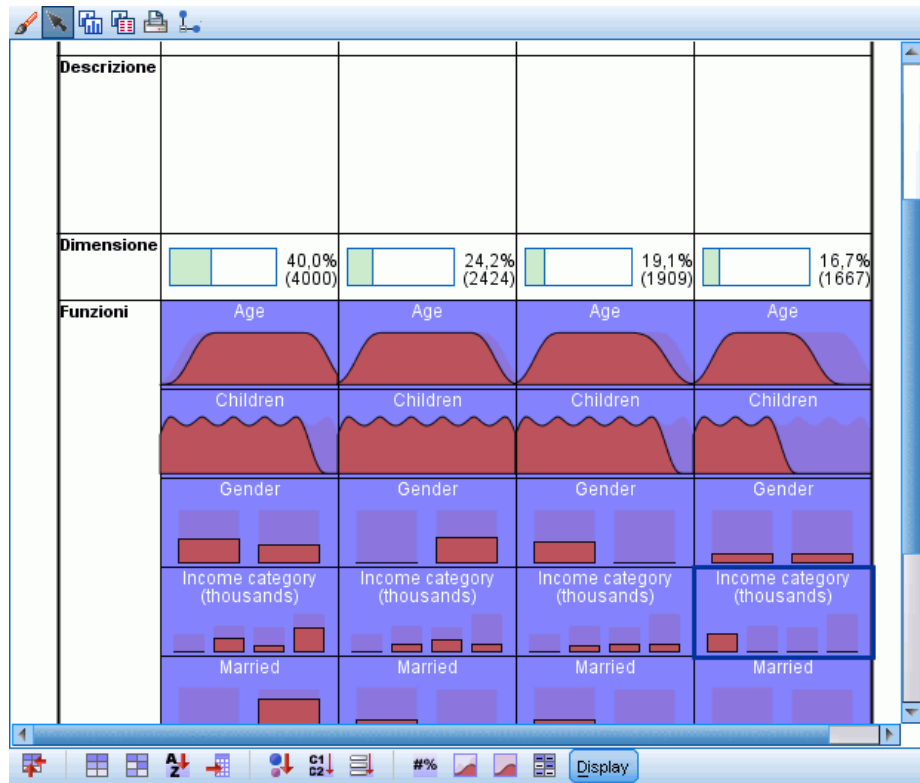
Grafico a barre delle categorie di reddito per il cluster 4



Al contrario del cluster 1, tutti i clienti del cluster 4 si trovano nella categoria di reddito più basso.

È possibile anche modificare la visualizzazione Cluster per visualizzare i grafici nelle celle e confrontare così più rapidamente le distribuzioni dei valori tra i cluster passando a un'altra visualizzazione mediante la barra degli strumenti nella parte inferiore della finestra del Viewer modelli.

Figura 9-10  
Grafici visualizzati nel cluster



Se si osserva la visualizzazione Cluster e i dati aggiuntivi forniti nei grafici di ogni cella, si notano alcune differenze sostanziali tra i cluster:





- I clienti del cluster 1 sono tendenzialmente individui più vecchi, sposati, con figli e con redditi più elevati.
- I clienti del cluster 2 sono tendenzialmente madri single non giovanissime con redditi modesti.
- I clienti del cluster 3 sono tendenzialmente uomini single più giovani senza figli.
- I clienti del cluster 4 sono tendenzialmente donne single più giovani, senza figli e con redditi più bassi.



Le celle Descrizione della visualizzazione Cluster sono campi di testo modificabili che consentono di inserire descrizioni di ogni cluster.

Figura 9-11

Visualizzazione Cluster con descrizioni dei cluster

Cluster	1	2	3	4
<b>Etichetta</b>				
<b>Descrizione</b>	Older, married, have children, higher income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
<b>Dimensione</b>	 40.0% (4000)	 24.2% (2424)	 19.1% (1909)	 16.7% (1667)
<b>Funzioni</b>	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

## Selezione dei record in base ai cluster

Esistono due modi per selezionare i record in base all'appartenenza a un cluster:

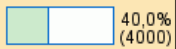
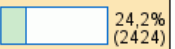

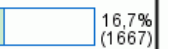
- Creare interattivamente una condizione di filtro nel Viewer dei modelli di cluster.
- Utilizzare i valori del campo cluster generato dalla procedura per specificare condizioni di filtro o di selezione.

## Creazione di un filtro nel Viewer dei modelli di cluster

Per creare una condizione di filtro che seleziona i record di determinati cluster nel Viewer dei modelli di cluster:

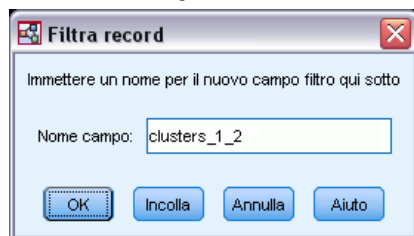
- ▶ Attivare (con un doppio clic del mouse) il Viewer dei modelli di cluster.
- ▶ Dall'elenco a discesa Visualizza nella parte inferiore della finestra del Viewer dei modelli di cluster, selezionare Cluster.
- ▶ Fare clic sul numero del cluster desiderato nella parte superiore della visualizzazione Cluster. Per selezionare più cluster, premere Ctrl mentre si fa clic su ogni numero di cluster desiderato.

Figura 9-12  
Cluster selezionati nella visualizzazione Cluster

Cluster	1	2	3	4
Etichetta				
Descrizione	Older, married, have children, high income	Older single mothers, moderate income	Younger single men, no children	Younger single women, no children, low income
Dimensione	 40,0% (4000)	 24,2% (2424)	 19,1% (1909)	 16,7% (1667)
Funzioni	Age 50,30	Age 44,07	Age 39,05	Age 33,09
	Children 1,58	Children 1,29	Children 0,39	Children 0,12
	Gender Male (57,0%)	Gender Female (100,0%)	Gender Male (100,0%)	Gender Female (50,9%)
	Income category (thousands) 75+ (56,1%)	Income category (thousands) 50-74 (47,2%)	Income category (thousands) 75+ (34,8%)	Income category (thousands) <25 (100,0%)
	Married Yes (100,0%)	Married No (78,5%)	Married No (100,0%)	Married No (78,5%)
	Education Post-graduate (20,5%)	Education Post-graduate (20,5%)	Education College (21,1%)	Education Post-graduate (20,6%)
	Years at current residence 9,47	Years at current residence 9,51	Years at current residence 9,47	Years at current residence 9,42

- ▶ Dai menu del Viewer dei modelli di cluster scegliere:  
Genera > Filtra record

Figura 9-13  
Finestra di dialogo Filtra record



- Inserire un nome per il campo filtro e fare clic su OK. I nomi devono essere conformi alle regole di denominazione di IBM® SPSS® Statistics.

Figura 9-14  
Record filtrati nell'Editor dei dati

	ID	Married	Children	Region	ClusterGroup1	clusters_1_2
<del>14</del>	03623	No	0	West	3	.00
<del>15</del>	01353	No	0	West	3	.00
<del>16</del>	07055	No	0	West	3	.00
17	04455	No	0	West	2	1.00
18	07210	No	1	West	2	1.00
<del>19</del>	08054	No	0	West	4	.00
<del>20</del>	06937	No	0	West	4	.00
<del>21</del>	06512	No	0	West	4	.00
<del>22</del>	08315	No	0	West	4	.00
23	09676	No	3	West	2	1.00
<del>24</del>	09636	No	0	West	4	.00
25	08579	No	1	West	2	1.00
26	01480	No	1	West	2	1.00

In questo modo si crea un nuovo campo nell'insieme di dati e si filtrano i record dell'insieme di dati in base ai valori di tale campo.

- I record con valore 1 per il campo filtro vengono inclusi nelle analisi, nei grafici e nei rapporti successivi.
- I record con valore 0 per il campo filtro vengono esclusi.
- I record esclusi non vengono cancellati dall'insieme di dati; vengono invece conservati con un indicatore dello stato del filtro, visualizzato come barra diagonale sul numero del record nell'Editor dei dati.

## Selezione dei record in base ai valori del campo Cluster

Per impostazione predefinita, l'Analisi cluster crea un nuovo campo che identifica il gruppo cluster per ogni record. Il nome predefinito di questo campo è *ClusterGroupn*, dove *n* è un numero intero che costituisce un nome campo univoco.

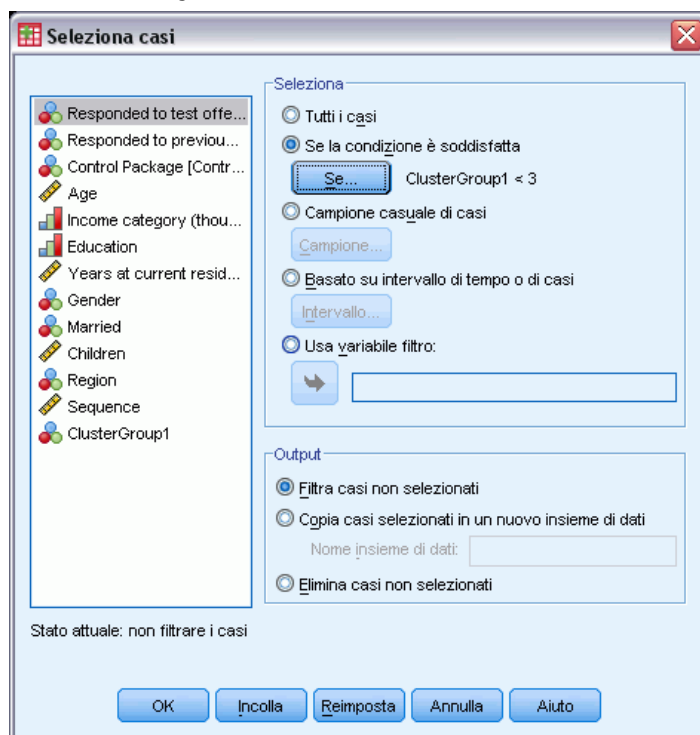
Figura 9-15  
Campo cluster aggiunto a un insieme di dati

	ID	Gender	Married	Children	Region	ClusterGroup1
1	01359	Female	No	0	West	4
2	06262	Female	No	1	West	2
3	08031	Male	No	0	West	3
4	01971	Male	No	0	West	4
5	09689	Male	No	0	West	3
6	06108	Male	No	1	West	3
7	09853	Male	No	0	West	3
8	06802	Male	No	0	West	4
9	07597	Male	No	0	West	3
10	03692	Male	No	1	West	3
11	00071	Male	No	0	West	4
12	00769	Male	No	0	West	3

Per utilizzare i valori del campo cluster per selezionare i record di determinati cluster:

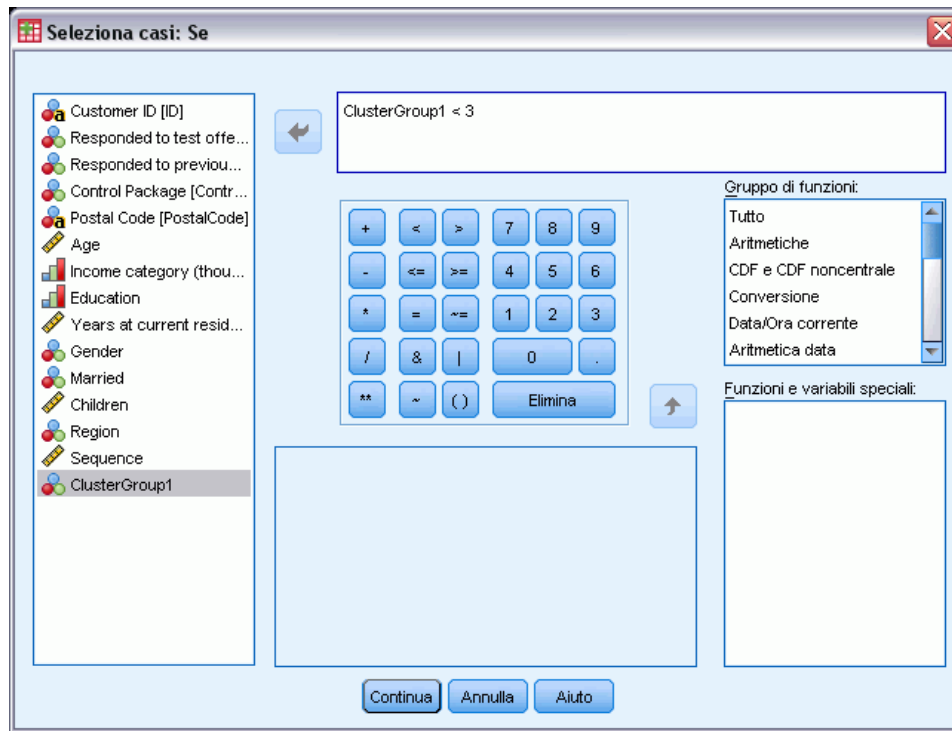
- Dai menu, scegliere:  
Dati > Seleziona casi

Figura 9-16  
Finestra di dialogo *Seleziona casi*



- Nella finestra di dialogo *Seleziona casi*, selezionare *Se la condizione è soddisfatta* e fare clic su *Se*.

Figura 9-17  
 Seleziona casi: Finestra di dialogo Se



- Inserire la condizione di selezione.

Per esempio,  $\text{ClusterGroup1} < 3$  selezionerà tutti i record dei cluster 1 e 2 escludendo i record del cluster 3 e di quelli successivi.

- Fare clic su Continua.

Nella finestra di dialogo Seleziona casi sono disponibili varie opzioni che consentono di decidere come trattare i record selezionati e non selezionati:

**Rimuovi casi non selezionati.** Crea un nuovo campo che specifica una condizione di filtro. I record esclusi non vengono cancellati dall'insieme di dati; vengono invece conservati con un indicatore dello stato del filtro, visualizzato come barra diagonale sul numero del record nell'Editor dei dati. Questa operazione equivale a selezionare in modo interattivo i cluster nel Viewer dei modelli di cluster.

**Copia casi selezionati in un nuovo insieme di dati.** Crea nella sessione corrente un nuovo insieme di dati che contiene solo i record che soddisfano la condizione di filtro. Questa operazione non ha alcun effetto sull'insieme di dati originale.

**Elimina casi non selezionati.** I record non selezionati vengono cancellati dall'insieme di dati. I record cancellati si possono recuperare solo uscendo dal file senza salvare le modifiche e quindi riaprendo il file. Se si salvano le modifiche al file di dati, l'eliminazione dei casi sarà definitiva.

La finestra di dialogo Seleziona casi dispone anche di un'opzione che consente di utilizzare una variabile esistente come variabile (campo) di filtro. Se si crea una condizione di filtro in modo interattivo nel Viewer dei modelli di cluster e si salva il campo filtro generato nell'insieme di dati, durante le sessioni successive sarà possibile utilizzare quel campo per filtrare i record.

## ***Riepilogo***

L'analisi cluster è un utile strumento esplorativo progettato per rivelare i raggruppamenti naturali (o cluster) all'interno dei dati. Le informazioni ricavate da questi cluster si possono utilizzare per determinare le strategie delle campagne di marketing e per sviluppare nuovi prodotti. È possibile selezionare i record in base all'appartenenza a un cluster per eseguire ulteriori analisi o per campagne di marketing mirate.

## ***Profili dei potenziali clienti***

Profili dei potenziali clienti utilizza i risultati di una campagna precedente o di prova per creare profili descrittivi. È possibile utilizzare i profili per definire gruppi specifici di contatti nelle campagne successive. Ad esempio, in base ai risultati di un mailing di prova, la divisione Direct marketing di una società desidera generare i profili dei tipi di persone con le maggiori probabilità di rispondere positivamente a un determinato tipo di offerta, a seconda dei dati demografici. In base ai risultati ottenuti, può quindi determinare i tipi di liste di distribuzione da utilizzare per offerte simili.

Ad esempio, la divisione Direct marketing di una società invia un mailing di prova a circa il 20% del totale del database di clienti. I risultati del mailing di prova vengono registrati in un file di dati che contiene anche le caratteristiche demografiche di ogni cliente, inclusi età, sesso, stato civile e area geografica. I risultati vengono registrati come semplici sì/no, indicando i clienti che hanno risposto (effettuato un acquisto) nel mailing di prova e quelli che non lo hanno fatto.

Tali informazioni vengono raccolte nel file *dmdata.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)

### ***Considerazioni sui dati***

Il campo risposta deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale in il capitolo 4 a pag. 24.](#)

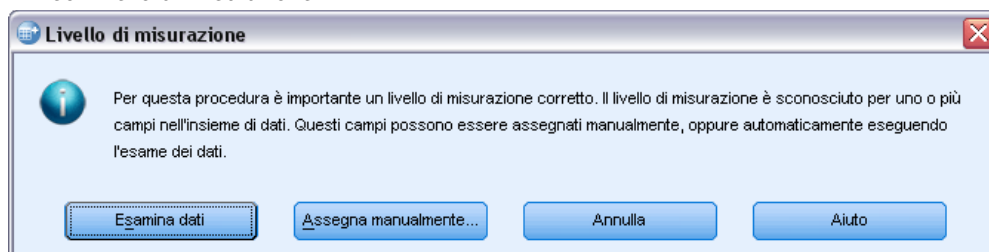
### ***Esecuzione dell'analisi***

- ▶ Per eseguire un'analisi Profili dei potenziali clienti, dai menu scegliere:  
Direct marketing > Scegli tecnica
  
- ▶ Selezionare Genera profili dei contatti che hanno risposto a un'offerta e fare clic su Continua.



L'avviso Livello di misurazione viene visualizzato quando il livello di misurazione di una o più variabili (campi) dell'insieme di dati è sconosciuto. Poiché influisce sul calcolo dei risultati di questa procedura, il livello di misurazione deve essere definito per tutte le variabili.

Figura 10-1  
Avviso Livello di misurazione



- **Esamina dati.** Legge i dati dell'insieme di dati attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con insiemi di dati di grandi dimensioni, questa operazione può richiedere del tempo.
- **Assegna manualmente.** Apre una finestra di dialogo che elenca tutti i campi con livello di misurazione sconosciuto, mediante la quale è possibile assegnare un livello di misurazione a questi campi. Il livello di misurazione si può assegnare anche nella Visualizzazione variabili dell'Editor dei dati.

Dal momento che il livello di misurazione è importante per questa procedura, è possibile accedere alla finestra di dialogo per la sua esecuzione solo quando per tutti i campi è stato definito un livello di misurazione.

In questo file di esempio non vi sono campi con livello di misurazione sconosciuto e tutti i campi hanno il livello di misurazione corretto; quindi, l'avviso relativo al livello di misurazione non dovrebbe essere visualizzato.

Figura 10-2  
Profili dei potenziali clienti: scheda Campi



- ▶ Per Campo risposta, selezionare *Ha risposto all'offerta di prova*.
- ▶ Per Valore risposta positiva, selezionare *Sì* dall'elenco a discesa. Nel campo di testo viene visualizzato il valore 1 perché "Sì" è in realtà un'etichetta associata a un valore registrato di 1. Se per il valore della risposta positiva non è definita un'etichetta, è possibile immettere il valore desiderato direttamente nel campo di testo.
- ▶ Per Crea profili con, selezionare *Età, Categoria di reddito, Istruzione, Anni alla residenza attuale, Sesso, Coniugato/a, Regione e Figli*.
- ▶ Fare clic sulla scheda Impostazioni.

Figura 10-3  
 Profili dei potenziali clienti: scheda Impostazioni

Profili dei potenziali clienti

Campi Impostazioni

Dimensione minima gruppo: 100

Includi le informazioni sulla soglia minima del tasso di risposta nei risultati

Specify target response rate (%): 7

Esegui Incolla Reimposta Annulla Aiuto

- ▶ Selezionare Includi le informazioni sulla soglia minima del tasso di risposta nei risultati.
- ▶ Come tasso di risposta obiettivo, immettere 7.
- ▶ Quindi fare clic su Esegui per eseguire la procedura.

## Output

Figura 10-4  
 Tabella dei tassi di risposta

Numero	Descrizione	Profilo		
		Dimensione gruppo	Tasso di risposta	Tasso di risposta cumulativo
1	Region = "West","South","East" Gender = "Female" Married = "No"	379	9.2%	9.2%
2	Region = "West","South","East" Gender = "Female" Married = "Yes"	299	5.0%	7.4%
3	Region = "West","South","East" Gender = "Male"	722	4.7%	6.0%
4	Region = "North"	517	2.5%	5.1%

Verde: soddisfa il tasso di risposta obiettivo.  
 Rosso: non soddisfa il tasso di risposta obiettivo.

La tabella dei tassi di risposta visualizza informazioni per ogni gruppo di profili identificato dalla procedura.

- I profili vengono visualizzati in ordine o tasso di risposta decrescente.
- Il tasso di risposta è la percentuale di clienti che hanno risposto positivamente (hanno effettuato un acquisto).
- Il tasso di risposta cumulativo è costituito dal tasso di risposta combinato nel gruppo di profili corrente e in tutti i gruppi precedenti. Poiché i profili vengono visualizzati in ordine decrescente di tasso di risposta, ciò vuol dire che il tasso di risposta cumulativo è il tasso di risposta combinato per il gruppo di profili corrente sommato a tutti i gruppi di profili con un tasso di risposta più elevato.
- La descrizione dei profili include solo le caratteristiche dei campi che forniscono un contributo significativo al modello. In questo esempio, regione, sesso e stato civile sono inclusi nel modello. I campi restanti, età, reddito, istruzione e anni alla residenza attuale, non sono inclusi perché non contribuiscono in modo significativo al modello.
- L'area verde della tabella rappresenta l'insieme di profili con un tasso di risposta cumulativo maggiore o uguale al tasso di risposta obiettivo specificato, che in questo esempio è 7%.
- L'area rossa della tabella rappresenta l'insieme di profili con un tasso di risposta cumulativo minore del tasso di risposta obiettivo specificato.
- Il tasso di risposta cumulativo nell'ultima riga della tabella è il tasso di risposta globale o medio per tutti i clienti inclusi nel mailing di prova, poiché è il tasso di risposta per tutti i gruppi di profili.

I risultati visualizzati nella tabella indicano che, se si definiscono come target donne nelle regioni Ovest, Sud ed Est, si dovrebbe ottenere un tasso di risposta leggermente più alto del tasso di risposta obiettivo.

Si noti tuttavia una differenza sostanziale tra i tassi di risposta delle donne non coniugate (9,2%) e quelle coniugate (5%) di queste regioni. Sebbene il tasso di risposta cumulativo di entrambi i gruppi sia maggiore del tasso di risposta obiettivo, il tasso di risposta del secondo gruppo considerato isolatamente è in realtà più basso del tasso di risposta obiettivo, circostanza che suggerisce l'opportunità di individuare altre caratteristiche che potrebbero migliorare il modello.

### Output intelligente

Figura 10-5  
Output intelligente

Tasso di risposta la tabella visualizza informazioni per ogni gruppo di profili identificato dalla procedura. La descrizione del profilo include le caratteristiche solo di quei campi che forniscono un contributo significativo al modello. I campi che non forniscono alcun contributo significativo non sono inclusi.

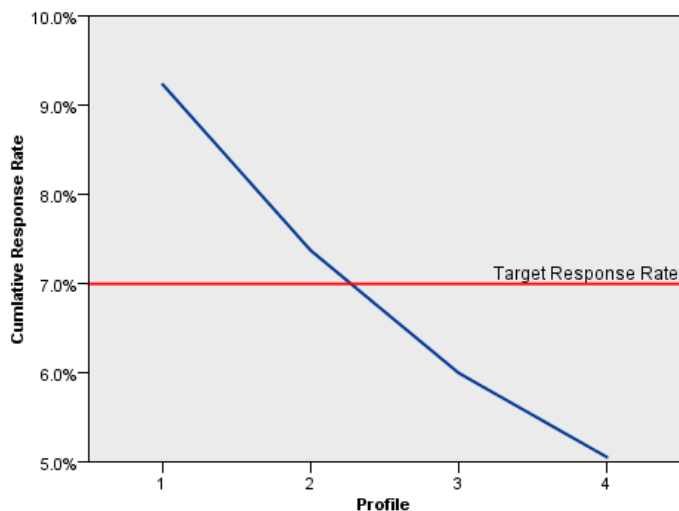
I profili vengono visualizzati in ordine di tasso di risposta decrescente. Il tasso di risposta indica la percentuale di clienti che ha risposto positivamente (ha effettuato un acquisto).

Il tasso di risposta cumulativo è costituito dal tasso di risposta combinato del gruppo di profili corrente e di tutti i gruppi di profili precedenti. Poiché i profili vengono visualizzati in ordine di tasso di risposta decrescente, il tasso di risposta cumulativo corrisponde al tasso di risposta combinato del gruppo di profili corrente a cui si aggiungono tutti i gruppi di profili con tasso di risposta più alto.

Il tasso di risposta obiettivo specificato è 7.00%. Le righe verdi hanno un tasso di risposta cumulativo maggiore di 7.00%, mentre le righe rosse hanno un tasso di risposta cumulativo minore di 7.00%. Sebbene alcuni gruppi di profili nell'area verde possano avere tassi di risposta individuali minori di 7.00%, il tasso di risposta cumulativo a quel punto è ancora maggiore di 7.00%.

La tabella è accompagnata da un “output intelligente” che fornisce informazioni generali su come interpretare la tabella e informazioni specifiche sui risultati in essa contenuti.

Figura 10-6  
Grafico del tasso di risposta cumulativo



Il grafico del tasso di risposta cumulativo costituisce una rappresentazione visiva dei tassi di risposta cumulativi visualizzati nella tabella. Poiché i profili vengono riportati in ordine decrescente di tasso di risposta, la linea del tasso di risposta cumulativo si abbassa per ogni profilo successivo. Proprio come la tabella, il grafico mostra che il tasso di risposta cumulativo si abbassa al di sotto del tasso di risposta obiettivo tra il gruppo di profili 2 e il gruppo di profili 3.

## **Riepilogo**

Per questo specifico mailing di prova sono stati identificati 4 gruppi di profili. I risultati indicano che le sole caratteristiche demografiche significative che sembrano correlate alla risposta o alla mancanza di risposta all'offerta sono il sesso, la regione e lo stato civile. Il gruppo con il tasso di risposta più alto è costituito da donne non coniugate, residenti nelle regioni Sud, Est e Ovest. Dopo questo gruppo i tassi di risposta si riducono rapidamente, anche se l'inclusione di donne coniugate delle stesse regioni produce comunque un tasso di risposta cumulativo maggiore del tasso di risposta obiettivo.

## ***Tassi di risposta codice postale***

Questa tecnica utilizza i risultati di una campagna precedente per calcolare i tassi di risposta relativi al codice postale. Tali tassi possono essere utilizzati per identificare codici postali specifici nelle campagne successive.

Ad esempio, in base ai risultati di un precedente mailing, la divisione Direct marketing di una società genera dei tassi di risposta in base ai codici postali. Secondo vari criteri, quali il tasso minimo di risposta accettabile e/o il numero massimo di contatti da includere nel mailing, la divisione potrà in seguito definire codici postali specifici.

Tali informazioni vengono raccolte nel file *dmdata.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)

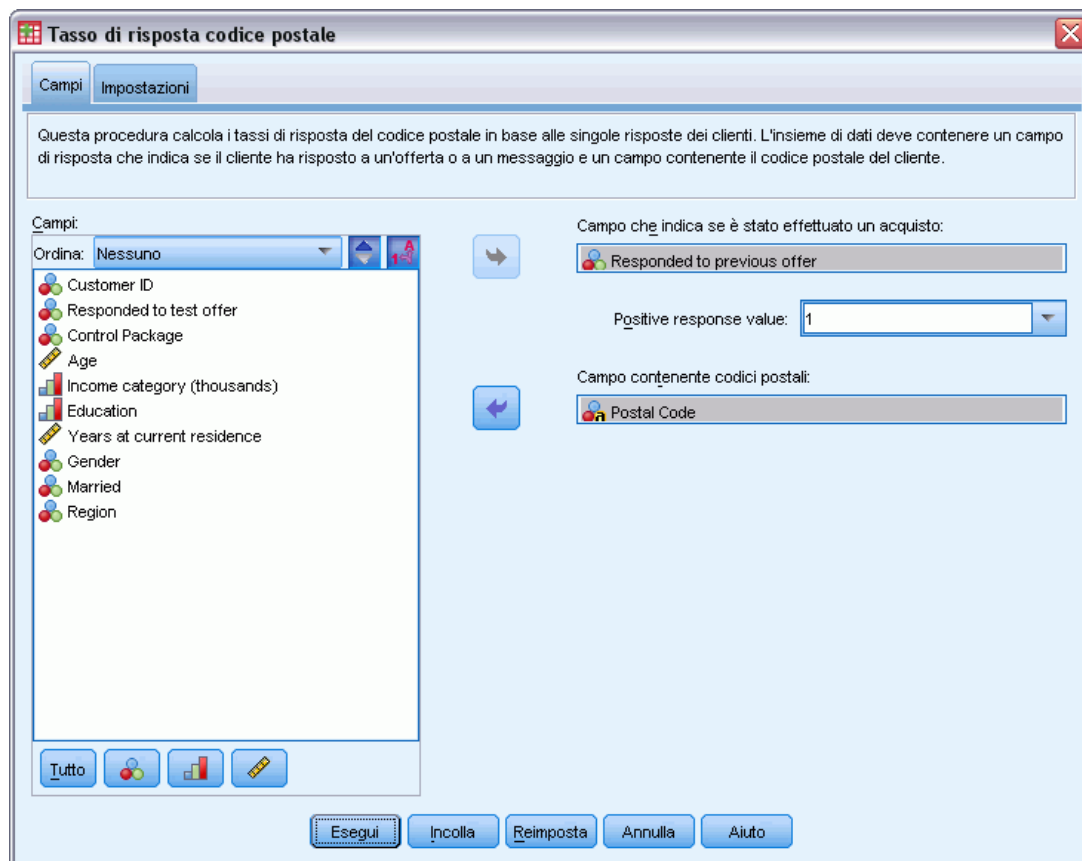
### ***Considerazioni sui dati***

Il campo risposta deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale in il capitolo 5 a pag. 33.](#)

### ***Esecuzione dell'analisi***

- ▶ Per calcolare i tassi di risposta relativi al codice postale, dai menu scegliere:  
Direct marketing > Scegli tecnica
  
- ▶ Selezionare Identifica i codici postali con la risposta migliore e fare clic su Continua.

Figura 11-1  
Tassi di risposta codice postale: scheda Campi



- ▶ Per Campo risposta, selezionare *Ha risposto all'offerta precedente*.
- ▶ Per Valore risposta positiva, selezionare *Sì* dall'elenco a discesa. Nel campo di testo viene visualizzato il valore 1 perché "Sì" è in realtà un'etichetta associata a un valore registrato di 1. Se per il valore della risposta positiva non è definita un'etichetta, è possibile immettere il valore desiderato direttamente nel campo di testo.
- ▶ Per Campo codice postale, selezionare *Codice postale*.
- ▶ Fare clic sulla scheda Impostazioni.



Figura 11-2  
Tassi di risposta codice postale: scheda Impostazioni

- ▶ Nel gruppo Raggruppa codici postali in base a, selezionare Prime 3 cifre o caratteri. In questo modo verranno calcolati i tassi di risposta combinati per tutti i contatti il cui codice postale inizia con le stesse 3 cifre o caratteri. Ad esempio, le prime 3 cifre di un codice postale statunitense corrispondono a un'area geografica comune più estesa dell'area definita dal codice completo a 5 cifre.
- ▶ Nel gruppo Output, selezionare Tasso di risposta e analisi di capacità.
- ▶ Selezionare Tasso di risposta obiettivo e immettere un valore di 5.
- ▶ Selezionare Numero di contatti e immettere un valore di 5000.
- ▶ Quindi fare clic su Esegui per eseguire la procedura.

## Output

Figura 11-3  
Nuovo insieme di dati con tassi di risposta in base al codice postale

	CodicePostale	ResponseRate	Responses	Contacts	Index	Rank
1	932	10.00%	4.00	40	3.60	1
2	098	8.82%	6.00	68	5.47	1
3	740	7.76%	9.00	116	8.30	1
4	100	7.69%	7.00	91	6.46	1
5	110	7.69%	5.00	65	4.62	1
6	954	7.55%	4.00	53	3.70	1
7	108	7.32%	6.00	82	5.56	1
8	107	7.04%	5.00	71	4.65	1
9	090	6.90%	4.00	58	3.72	1
10	966	6.90%	4.00	58	3.72	1
11	760	6.72%	8.00	119	7.46	1
12	113	6.25%	5.00	80	4.69	1
13	927	6.00%	3.00	50	2.82	1
14	969	6.00%	3.00	50	2.82	1
15	972	5.88%	3.00	51	2.82	2
16	123	5.71%	2.00	35	1.89	2
17	950	5.66%	3.00	53	2.83	2
18	602	5.56%	6.00	108	5.67	2
19	617	5.56%	5.00	90	4.72	2
20	934	5.56%	3.00	54	2.83	2
21	946	5.36%	3.00	56	2.84	2

Viene automaticamente creato un nuovo insieme di dati, che contiene un unico record (riga) per ogni codice postale. In questo esempio, ogni riga contiene informazioni di riepilogo per tutti i codici postali che iniziano con le stesse 3 cifre o caratteri.

Oltre al campo che contiene il codice postale, il nuovo insieme di dati include i campi seguenti:

- **Tasso di risposta.** La percentuale di risposte positive in ogni codice postale. I record vengono automaticamente disposti in ordine decrescente di tassi di risposta, in modo che i codici postali con i tassi di risposta più alti vengano visualizzati all'inizio dell'insieme di dati.
- **Risposte.** Il numero di risposte positive in ogni codice postale.
- **Contatti.** Il numero totale di contatti per ogni codice postale che contiene un valore non mancante per il campo risposta.

- **Indice.** La risposta “pesata” in base alla formula  $N \times P \times (1-P)$ , dove  $N$  è il numero di contatti e  $P$  è il tasso di risposta espresso come proporzione. Nel caso di due codici postali con lo stesso tasso di risposta, questa formula assegnerà un valore di indice più alto al codice postale con il numero più elevato di contatti.
- **Rango.** Rango di decile (primo 10%, primo 20%, ecc.) dei tassi di risposta in base al codice postale cumulativi in ordine decrescente.

Poiché nella scheda Impostazioni della finestra di dialogo Tassi di risposta codice postale è stata selezionata l'opzione Tasso di risposta e analisi di capacità, nel Viewer vengono visualizzati una tabella e un grafico di riepilogo dei tassi di risposta.

Figura 11-4

Tabella dei tassi di risposta

Percentile	Tasso di risposta	Contacts	Cumulative Response Rate	Total Contacts
Top 10%	7.3	1001	7.3	1001
Top 20%	5.3	956	6.3	1957
Top 30%	4.3	1042	5.6	2999
Top 40%	3.5	1127	5.1	4126
Top 50%	3.0	1173	4.6	5299
Top 60%	2.4	914	4.3	6213
Top 70%	2.0	948	4.0	7161
Top 80%	1.7	1095	3.7	8256
Top 90%	1.2	680	3.5	8936
Top 100%	.0	1064	3.1	10000

Green Red Caption

La tabella riepiloga i risultati per rango di decile in ordine decrescente (primo 10%, primo 20%, ecc.).

- Il tasso di risposta cumulativo è costituito dalla percentuale combinata di risposte positive nella riga corrente e in tutte le righe precedenti. Poiché i risultati vengono visualizzati in ordine decrescente di tassi di risposta, si tratta del tasso di risposta combinato del decile corrente e di tutti i decili con un tasso di risposta più elevato.
- La tabella può essere codificata mediante colori in base ai valori immessi per il tasso di risposta obiettivo e il numero massimo di contatti. Le righe con un tasso di risposta cumulativo maggiore o uguale al 5% e un massimo di 5.000 contatti cumulativi sono di colore verde. La codifica mediante colori si basa sul valore soglia raggiunto per primo. In questo esempio, entrambi i valori soglia vengono raggiunti nello stesso decile.

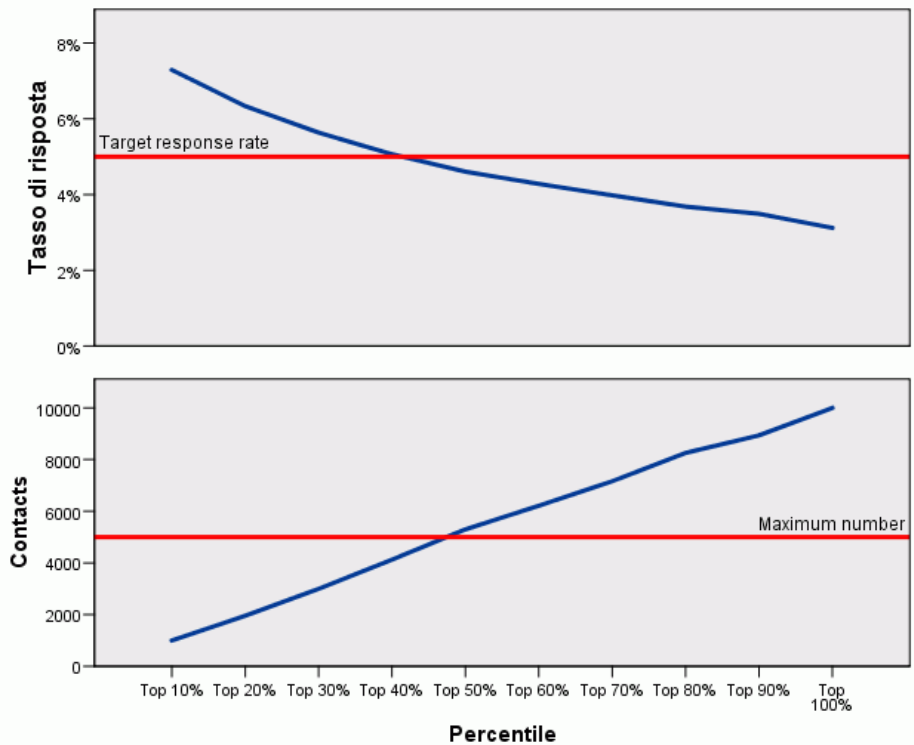
**Figura 11-5**  
*Output intelligente per la tabella del tasso di risposta*

La tabella dei tassi di risposta riepiloga i risultati per rango di decile in ordine decrescente (primo 10%, primo 20%, ecc.). Il tasso di risposta cumulativo è costituito dalla percentuale combinata di risposte positive nella riga corrente e in tutte le righe precedenti. Poiché i risultati vengono visualizzati in ordine decrescente di tassi di risposta, si tratta del tasso di risposta combinato del decile corrente e di tutti i decili con un tasso di risposta più elevato. Poiché il rango di decile è incluso nel nuovo insieme di dati, è possibile individuare facilmente i codici postali che soddisfano un determinato tasso di risposta cumulativo. Il campo nel nuovo insieme di dati che identifica il rango di decile è denominato Rango, dove 1=Primo 10%, 2=Primo 20%, ecc.

Il tasso di risposta minimo specificato è 5.00%. Il numero massimo di contatti specificato è 5000. La codifica mediante colori della tabella si basa sul valore soglia raggiunto per primo. Entrambi i valori soglia vengono raggiunti nella stessa categoria. Le righe verdi hanno un tasso di risposta cumulativo maggiore o uguale al tasso di risposta minimo specificato e un numero di contatti cumulativo minore o uguale al numero di contatti massimo specificato. Le righe rosse hanno un tasso di risposta cumulativo minore del tasso di risposta minimo specificato e un numero di contatti cumulativo maggiore del numero di contatti massimo specificato.

La tabella è accompagnata da una descrizione generale che illustra come leggere la tabella. Se è stato specificato un tasso di risposta minimo o un numero massimo di contatti, la tabella comprende anche una sezione che spiega la relazione dei risultati con i valori soglia specificati.

**Figura 11-6**  
*Grafico del tasso di risposta cumulativo*



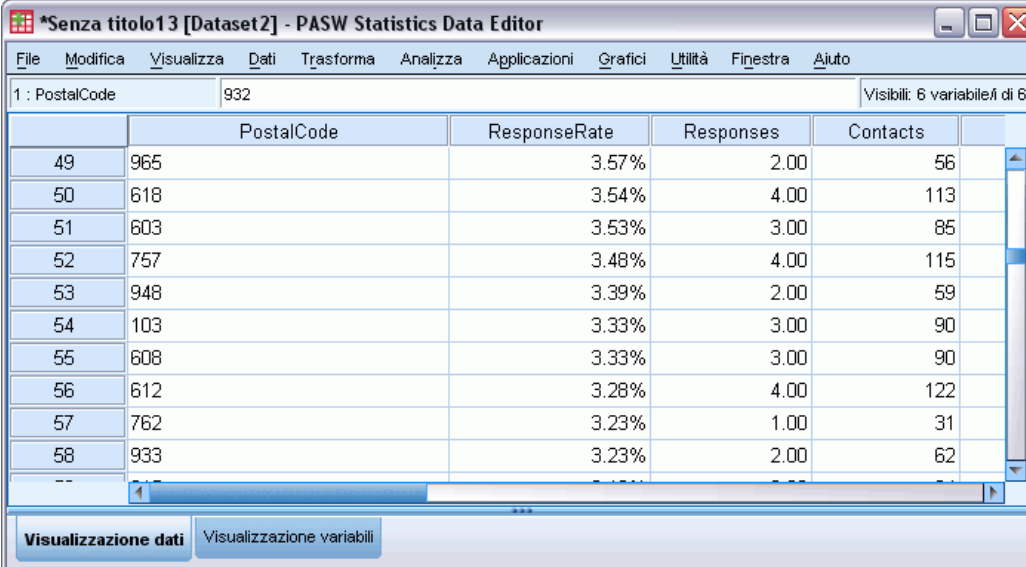
Il grafico del tasso di risposta cumulativo e del numero di contatti cumulativo in ogni decile è una rappresentazione visiva delle stesse informazioni visualizzate nella tabella dei tassi di risposta. La soglia per il tasso di risposta cumulativo minimo e il numero di contatti cumulativo massimo viene raggiunta in qualche punto tra il 40esimo e il 50esimo percentile.

- Poiché il grafico visualizza i tassi di risposta cumulativi in ordine decrescente di rango di decile di tasso di risposta, la linea del tasso di risposta cumulativo si abbassa per ogni decile successivo.
- Poiché la linea del numero di contatti rappresenta il numero di contatti cumulativo, tale linea si alza sempre.

Le informazioni contenute nella tabella e nel grafico indicano che, se si desidera ottenere un tasso di risposta di almeno il 5% senza includere più di 5.000 contatti nella campagna, è necessario concentrarsi sui codici postali nei primi 4 decili. Poiché il rango di decile è incluso nel nuovo insieme di dati, è possibile individuare facilmente i codici postali che soddisfano il requisito del primo 40%.

Figura 11-7

Nuovo insieme di dati



	PostalCode	ResponseRate	Responses	Contacts
49	965	3.57%	2.00	56
50	618	3.54%	4.00	113
51	603	3.53%	3.00	85
52	757	3.48%	4.00	115
53	948	3.39%	2.00	59
54	103	3.33%	3.00	90
55	608	3.33%	3.00	90
56	612	3.28%	4.00	122
57	762	3.23%	1.00	31
58	933	3.23%	2.00	62

Nota: *Rango* viene registrato come valore intero compreso tra 1 e 10. Il campo ha etichette dei valori definite, dove 1= primo 10%, 2= primo 20%, ecc. Nella Visualizzazione dati dell'Editor dei dati verranno visualizzati gli effettivi valori del rango o le etichette dei valori, in base alle impostazioni di visualizzazione.

## Riepilogo

La procedura Tassi di risposta codice postale utilizza i risultati di una campagna precedente per calcolare i tassi di risposta relativi al codice postale. Tali tassi possono essere utilizzati per identificare codici postali specifici nelle campagne successive. La procedura crea automaticamente un nuovo insieme di dati che contiene i tassi di risposta per ogni codice postale. In base alle informazioni contenute nella tabella e nel grafico dei tassi di risposta e alle informazioni sul rango di decile nel nuovo insieme di dati, è possibile individuare l'insieme di codici postali che corrisponde a uno specifico tasso di risposta cumulativo minimo e/o uno specifico numero di contatti cumulativo massimo.



## ***Propensione all'acquisto***

Propensione all'acquisto utilizza i risultati di un mailing di prova o di una campagna precedente per generare punteggi di propensione. I punteggi indicano i contatti che risponderanno con maggiore probabilità, in base a varie caratteristiche selezionate.

Questa tecnica utilizza la regressione logistica binaria per creare un modello predittivo. Il processo di creazione e applicazione di un modello predittivo è costituito da due passaggi fondamentali:

- ▶ Creare il modello e salvare il file del modello. Il modello viene creato utilizzando un insieme di dati per il quale è noto il risultato di interesse (spesso denominato **obiettivo**). Ad esempio, se si desidera creare un modello in grado di generare una previsione di chi risponderà a una campagna di marketing per posta, è necessario iniziare con un insieme di dati che contiene già informazioni su chi ha risposto e chi non ha risposto. Ciò potrebbe essere costituito ad esempio dai risultati di un mailing di prova a un piccolo gruppo di clienti o da informazioni sulle risposte date a una campagna simile in passato.
- ▶ Applicare il modello a diversi insiemi di dati (per i quali non è noto il risultato di interesse) per prevedere i risultati.

In questo esempio, vengono utilizzati due file di dati: *dmdata2.sav* viene utilizzato per costruire il modello e in seguito questo modello viene applicato a *dmdata3.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)

### ***Considerazioni sui dati***

Il campo risposta (obiettivo di interesse) deve essere categoriale, cioè con un valore che rappresenta tutte le risposte positive. Si presume che tutti gli altri valori non mancanti indichino una risposta negativa. Se il campo risposta rappresenta un valore continuo (scala) quale il numero di acquisti o il loro valore monetario, è necessario creare un nuovo campo che assegna un valore di risposta positiva unico a tutti i valori di risposta diversi da zero. [Per ulteriori informazioni, vedere l'argomento Creazione di un campo Risposta categoriale in il capitolo 6 a pag. 41.](#)

### ***Costruzione di un modello predittivo***

- ▶ Aprire il file di dati *dmdata2.sav*.

Questo file contiene varie caratteristiche demografiche delle persone che hanno ricevuto il mailing di prova e contiene anche informazioni sulla loro risposta al mailing. Queste informazioni sono registrate nel campo (variabile) *Ha risposto*. Il valore 1 indica che il contatto ha risposto al mailing, mentre il valore 0 indica che il contatto non ha risposto.

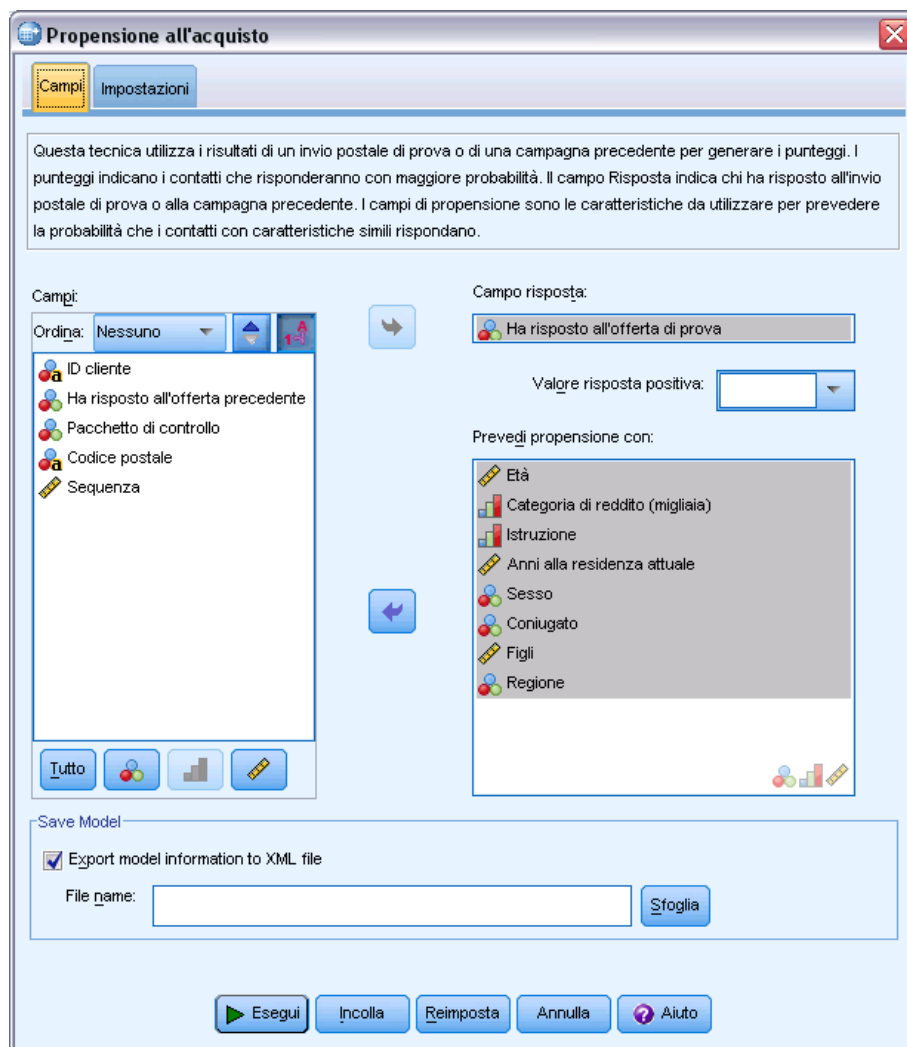
**Figura 12-1**  
Contenuto di un file di dati nell'Editor dei dati

ID	HaRisposto	Indietro	PacchettiDiControllo	CodicePostale	Età	Reddito	Istruzione	Nfam	Sesso
06262	0	0	1	96600	67	3	5	10	1
03692	0	0	0	95510	53	3	4	9	0
01480	0	0	1	92590	56	3	2	11	1
06118	0	0	1	92670	55	2	3	8	1
07378	0	0	1	92690	56	2	1	6	1
08467	0	0	0	93410	50	1	1	8	1
09621	0	0	1	93480	67	4	2	9	1
03029	0	0	0	93490	47	1	2	10	1
02660	0	0	1	93490	58	4	5	11	1

- ▶ Dai menu, scegliere:  
Direct Marketing > Scegli tecnica
- ▶ Selezionare Seleziona contatti con maggiore probabilità di acquisto e fare clic su Continua.



Figura 12-2  
Propensione all'acquisto: scheda Campi



- ▶ Per Campo risposta, selezionare *Ha risposto all'offerta di prova*.
- ▶ Per Valore risposta positiva, selezionare *Sì* dall'elenco a discesa. Nel campo di testo viene visualizzato il valore 1 perché "Sì" è in realtà un'etichetta associata a un valore registrato di 1. Se per il valore della risposta positiva non è definita un'etichetta, è possibile immettere il valore desiderato direttamente nel campo di testo.
- ▶ Per Prevedi propensione con, selezionare *Età*, *Categoria di reddito*, *Istruzione*, *Anni alla residenza attuale*, *Sesso*, *Coniugato/a*, *Regione* e *Figli*.
- ▶ Selezionare *Esporta informazioni modello in file XML*.
- ▶ Fare clic su *Sfogli* per accedere alla posizione in cui salvare il file e immettere un nome per il file.
- ▶ Nella finestra di dialogo *Propensione all'acquisto*, fare clic sulla scheda *Impostazioni*.

Figura 12-3  
Propensione all'acquisto: scheda Impostazioni

**Propensione all'acquisto**

Campi Impostazioni

**Convalida del modello**

È possibile convalidare il modello utilizzato per generare i punteggi. Per convalidare il modello, è necessario suddividere i dati in partizioni. La partizione di training viene utilizzata per formare, o costruire, il modello. La partizione di test viene utilizzata per convalidare il modello. Se si desidera convalidare il modello, questa tecnica assegna automaticamente i record alle partizioni.

Convalida il modello

Dimensioni partizione campione di training (%): 50

Imposta seme per replicare i risultati

Seme: 2000000

**Output diagnostico**

Qualità generale del modello

Tabella classificazioni

Probabilità minima: 0.05

**Nome ed etichetta del campo risposta ricodificata**

Questa tecnica ricodifica automaticamente il campo risposta in un nuovo campo nel quale 1 rappresenta le risposte positive e 0 rappresenta le risposte negative.

Nome nuovo campo: Response\_recoded1

Etichetta nuovo campo: Risposta ricodificata(1=Si, 0=No)

**Salva punteggi**

Questa tecnica utilizza i risultati di un invio postale di prova o di una campagna precedente per generare i punteggi. I punteggi vengono automaticamente salvati per consentirne l'uso. Gli altri controlli della scheda consentono di controllare ulteriormente i dati salvati.

Nuovo nome campo per i punteggi: Score1

Esegui Incolla Reimposta Annulla Aiuto

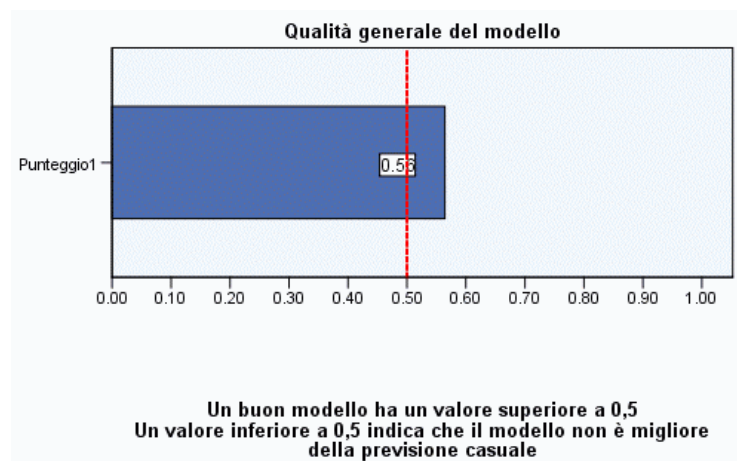
- ▶ Nel gruppo Convalida del modello, selezionare Convalida modello e Imposta seme per replicare i risultati.
- ▶ Utilizzare le dimensioni partizione campione di training predefinite del 50% e il valore seme predefinito di 2000000.
- ▶ Nel gruppo Output diagnostico, selezionare Qualità generale del modello e Tabella classificazioni.
- ▶ Per Probabilità minima, immettere 0,05. Come regola generale è bene specificare un valore vicino a quello del tasso di risposta obiettivo minimo, espresso come proporzione. Il valore 0,05 rappresenta un tasso di risposta del 5%.
- ▶ Fare clic su **Esegui** per eseguire la procedura e generare il modello.

## Valutazione del modello

La funzione Propensione all'acquisto produce un grafico della qualità generale del modello e una tabella classificazioni, che possono essere utilizzati per valutare il modello.

Il grafico della qualità generale del modello fornisce una rapida indicazione visiva della qualità del modello. Come regola generale, la qualità generale del modello deve essere superiore a 0,5.

Figura 12-4  
Grafico della qualità generale del modello



Per verificare che il modello sia adeguato per il calcolo del punteggio, è necessario anche esaminare la tabella classificazioni.

Figura 12-5  
Tabella classificazioni

		Tabella Classificazione					
		Previsto					
		Campione di training			Campione di verifica		
Osservato		Risposta ricodificata(1=Si, 0=No)		Percentuale corretta	Risposta ricodificata(1=Si, 0=No)		Percentuale corretta
		No	Sì		No	Sì	
Risposta ricodificata (1=Si, 0=No)	No	651	249	72.33	653	267	70.98
	Sì	19	20	51.28	36	22	37.93
Percentuale globale		2.84	7.43	71.46	5.22	7.61	69.02

La tabella classificazioni confronta i valori attesi ed effettivi del campo obiettivo. La percentuale di precisione generale può fornire un'indicazione del buon funzionamento del modello, ma la percentuale di risposte positive corrette prevista può essere un dato più interessante se lo scopo è quello di costruire un modello che identifichi il gruppo di contatti che probabilmente genereranno un tasso di risposta positiva uguale o superiore al tasso di risposta positiva minimo specificato.

In questo esempio, la tabella classificazioni è divisa in un **campione di training** e un **campione di verifica**. Il campione di training viene utilizzato per costruire il modello. Il modello viene quindi applicato al campione di verifica per controllare il corretto funzionamento del modello.

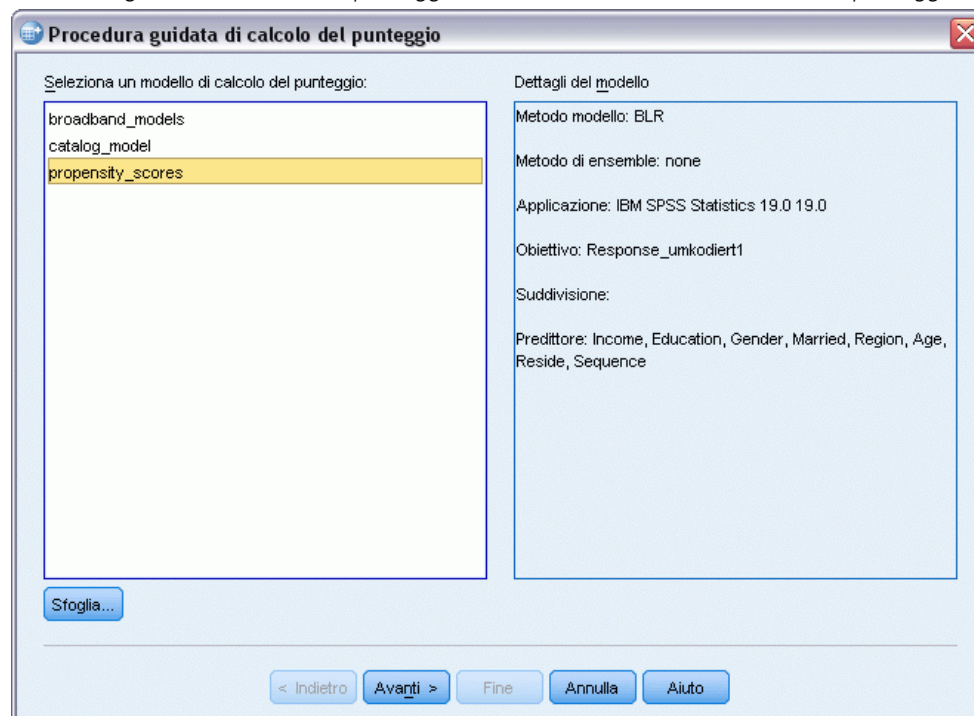
Il tasso di risposta minimo specificato è 0,05 o 5%. La tabella classificazioni mostra che il tasso di classificazioni corrette per risposte positive è 7,43% nel campione di training e 7,61% nel campione di verifica. Poiché il tasso di risposta del campione di verifica è superiore al 5%, questo modello dovrebbe essere in grado di identificare un gruppo di contatti che probabilmente genereranno una risposta superiore al 5%.

## Applicazione del modello

- ▶ Aprire il file di dati *dmdata3.sav*. Questo file di dati contiene informazioni demografiche e altre informazioni su tutti i contatti che non sono stati inclusi nel mailing di prova. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)
- ▶ Aprire la Procedura guidata di calcolo del punteggio. A tale scopo, dai menu scegliere: Utilità > Procedura guidata di calcolo del punteggio

Figura 12-6

*Procedura guidata di calcolo del punteggio, Seleziona un modello di calcolo del punteggio*



- ▶ Fare clic su Sfoglia per accedere alla posizione in cui è stato salvato il file XML del modello e fare clic su Seleziona nella finestra di dialogo Sfoglia.

Tutti i file con estensione .xml o .zip vengono visualizzati nella procedura guidata di calcolo del punteggio. Le estensioni non vengono visualizzate. Se il file selezionato viene riconosciuto come un file di modello valido, viene visualizzata una descrizione del modello.

- ▶ Selezionare il file XML del modello creato e fare clic su Avanti.

Figura 12-7  
Procedura guidata di calcolo del punteggio, Fai corrispondere i campi del modello

Nome modello: propensity\_scores  
Tipo di modello: Regressione logistica binaria

Fai corrispondere i campi del modello all'insieme di dati:

Campi dell'insieme di dati	Campi del modello	Ruolo	Misura	Tipo
IncomeCategory	Income	Predittore	Nominale	Numerico
Education	Education	Predittore	Nominale	Numerico
Gender	Gender	Predittore	Nominale	Numerico
Married	Married	Predittore	Nominale	Numerico
Region	Region	Predittore	Nominale	Numerico
Age	Age	Predittore	Continuo	Numerico
Reside	Reside	Predittore	Continuo	Numerico
Sequence	Sequence	Predittore	Continuo	Numerico

Valori mancanti

Usa sostituzione valori  
 Usa mancanti di sistema

< Indietro Avanti > Fine Annulla Aiuto

Per calcolare il punteggio dell'insieme di dati attivo, è necessario che l'insieme di dati contenga campi (variabili) corrispondenti a tutti i predittori del modello. Se il modello include anche campi di suddivisione, è necessario che l'insieme di dati contenga anche campi corrispondenti a tutti i campi di suddivisione del modello.

- Per impostazione predefinita, tutti i campi dell'insieme di dati attivo con lo stesso nome e tipo dei campi del modello vengono fatti corrispondere automaticamente.
- Utilizzare l'elenco a discesa per far corrispondere i campi dell'insieme di dati ai campi del modello. Per far corrispondere i campi, il tipo di dati di ogni campo deve essere lo stesso nel modello e nell'insieme di dati.
- Sarà possibile continuare con la procedura guidata o il calcolo del punteggio dell'insieme di dati attivo solo se tutti i predittori (ed eventuali campi di suddivisione) del modello corrispondono ai campi dell'insieme di dati attivo.

L'insieme di dati attivo non contiene un campo denominato *Reddito*. Per questo motivo, la cella contenuta nella colonna Campi dell'insieme di dati che corrisponde al campo del modello *Reddito* è vuota inizialmente. È necessario selezionare un campo nel file di dati attivo che sia equivalente a quel campo del modello.

- Dall'elenco a discesa nella colonna Campi dell'insieme di dati, nella cella vuota della riga del campo del modello *Reddito*, selezionare *Categoria di reddito*.

*Nota:* oltre al nome e al tipo di campo, è necessario assicurarsi che i valori effettivi dei dati nell'insieme di dati per il quale viene calcolato il punteggio siano registrati come i valori dei dati nell'insieme di dati utilizzato per creare il modello. Ad esempio, se il modello è stato costruito con un campo *Reddito* nel quale il reddito è suddiviso in quattro categorie e *Categoria reddito* nell'insieme di dati attivo presenta il reddito suddiviso in sei categorie o in quattro categorie diverse, questi campi non corrispondono realmente tra di loro e i risultati del calcolo del punteggio non saranno affidabili.

Fare clic su **Avanti** per procedere al passaggio successivo della procedura guidata di calcolo del punteggio.

Figura 12-8

*Procedura guidata di calcolo del punteggio: Seleziona funzioni di calcolo del punteggio*

Nome modello: propensity\_scores  
 Tipo di modello: BLR

Seleziona funzioni di calcolo del punteggio  
 Ciascuna funzione selezionata creerà un nuovo campo nell'insieme di dati.

	Funzione	Nomi campi	Valore
<input type="checkbox"/>	Valore atteso	PredictedValue	
<input type="checkbox"/>	Probabilità della categoria prevista	PredictedProbability	
<input checked="" type="checkbox"/>	Probabilità della categoria selezionata	SelectedProbability	1
<input type="checkbox"/>	Confidenza	Confidence	

< Indietro   Avanti >   Fine   Annulla   Aiuto

Le funzioni di punteggio corrispondono ai tipi di “punteggio” disponibili per il modello selezionato. Le funzioni di punteggio disponibili variano a seconda del modello. Per il modello logistico binario utilizzato in questo esempio, le funzioni disponibili sono valore atteso, probabilità del valore atteso, probabilità di un valore selezionato e confidenza.

In questo esempio, l'interesse è rivolto alla probabilità prevista di una risposta positiva al mailing; pertanto occorre ottenere la probabilità di un valore selezionato.

- ▶ Selezionare Probabilità della categoria selezionata.
- ▶ Nella colonna Valore, selezionare 1 dall'elenco a discesa. L'elenco dei possibili valori per l'obiettivo viene definito nel modello, in base ai valori obiettivo contenuti nel file di dati utilizzato per costruire il modello.

- ▶ Deselezionare tutte le altre funzioni di punteggio.
- ▶ Se lo si desidera, è possibile assegnare un nome più descrittivo al nuovo campo che conterrà i valori di punteggio nell'insieme di dati attivo. Ad esempio, *Probabilità\_di\_risposta*.
- ▶ Fare clic su Fine per applicare il modello all'insieme di dati attivo.

Il nuovo campo che contiene la probabilità di una risposta positiva viene aggiunto alla fine dell'insieme di dati.

**Figura 12-9**

*Insieme di dati con nuovo campo di probabilità*

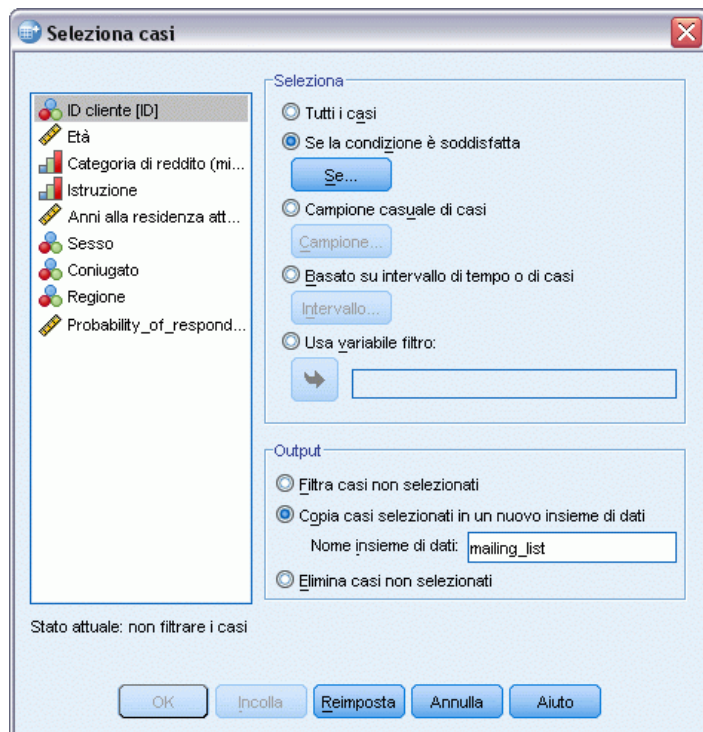
Nfam	Sesso	Coniugato	Regione	Probability_of_responding
7	1	0	4	.04
9	0	0	4	.03
12	0	0	4	.03
8	0	0	4	.04
13	0	0	4	.07
10	0	0	4	.04
12	0	0	4	.03
15	0	0	4	.05
10	0	0	4	.05
14	0	0	4	.02
5	0	0	4	.12

È possibile quindi utilizzare questo campo per selezionare il sottoinsieme di contatti che probabilmente genereranno un tasso di risposta positiva di un determinato livello o sopra di esso. Ad esempio, è possibile creare un nuovo insieme di dati che contiene il sottoinsieme di casi che probabilmente genereranno un tasso di risposta positiva pari almeno al 5%.



- Dai menu, scegliere:  
Dati > Seleziona casi

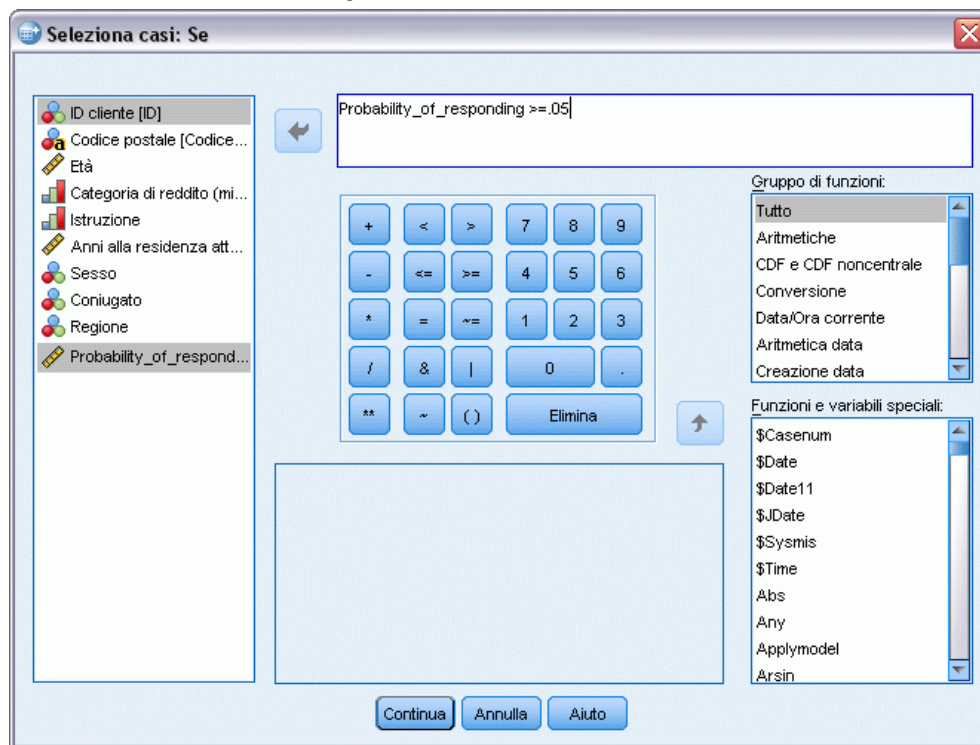
Figura 12-10  
Finestra di dialogo Seleziona casi



- Nella finestra di dialogo Seleziona casi, selezionare Se la condizione è soddisfatta e fare clic su Se.



Figura 12-11  
Seleziona casi: Finestra di dialogo Se



- In Seleziona casi: Finestra di dialogo Se, immettere la seguente espressione:

Probabilità\_di\_risposta >=.05

*Nota:* se è stato utilizzato un nome diverso per il campo che contiene i valori di probabilità, immettere questo nome invece di Probabilità\_di\_risposta. Il nome predefinito del campo è *SelectedProbability*.

- Fare clic su Continua.
- Nella finestra di dialogo Seleziona casi, selezionare Copia casi selezionati in un nuovo insieme di dati e immettere un nome per il nuovo insieme di dati. I nomi degli insiemi di dati devono essere conformi alle regole dei nomi dei campi (variabili).
- Fare clic su OK per creare l'insieme di dati con i contatti selezionati.

Il nuovo insieme di dati contiene solo i contatti con una probabilità prevista di risposta positiva pari almeno al 5%.

Figura 12-12

*Nuovo insieme di dati con i contatti selezionati*

Nfam	Sesso	Coniugato	Regione	Probability_of_responding
13	0	0	4	.07
15	0	0	4	.05
10	0	0	4	.05
5	0	0	4	.12
7	0	0	4	.08
10	0	0	4	.10
15	1	0	4	.05
11	1	0	4	.08
9	1	0	4	.08
9	1	0	4	.05

## **Riepilogo**

Propensione all'acquisto utilizza i risultati di un mailing di prova o di campagne precedenti per generare punteggi di propensione. I punteggi indicano i contatti che risponderanno con maggiore probabilità, in base a varie caratteristiche selezionate. Questa tecnica costruisce un modello predittivo che può essere applicato all'insieme di dati per ottenere punteggi di propensione.

## ***Test pacchetto di controllo***

Questa tecnica confronta le campagne di marketing per rilevare eventuali differenze significative nell'efficacia di diversi pacchetti o offerte. L'efficacia di una campagna viene misurata in base alle risposte.

Per esempio, la divisione Direct marketing di una società desidera verificare se il nuovo design di una confezione genererà un numero maggiore di risposte positive rispetto alla confezione esistente. Invia quindi un mailing di prova per determinare se la nuova confezione generi un aumento significativo del tasso di risposta positiva. Il mailing di prova consiste di un gruppo di controllo che riceve la confezione esistente e un gruppo di prova che riceve la confezione con il nuovo design. I risultati dei due gruppi vengono confrontati per rilevare una eventuale differenza significativa.

Tali informazioni vengono raccolte nel file *dmdata.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A a pag. 100.](#)

### ***Esecuzione dell'analisi***

- ▶ Per eseguire un test pacchetto di controllo, dai menu scegliere:  
Direct marketing > Scegli tecnica
  
- ▶ Selezionare Confronta efficacia delle campagne (Test pacchetto di controllo) e fare clic su Continua.

Figura 13-1  
Test pacchetto di controllo, scheda Campi

Questa tecnica confronta le campagne di marketing per rilevare eventuali differenze significative tra i diversi pacchetti o offerte. Il campo Campagna identifica le diverse campagne. Il campo Efficacia indica se un cliente ha risposto alla campagna.

Campi:  
Ordina: Nessuno

- Customer ID
- Responded to previous offer
- Postal Code
- Age
- Income category (thousands)
- Education
- Years at current residence
- Gender
- Married
- Region

Campagna:  
Control Package

Efficacia:  
Risposta positiva

Misura di efficacia

Importo acquisto

Risposta positiva

Valore: 1

Nome nuovo campo: Effectiveness\_new1

Etichetta nuovo campo: Campo test ricodificato (1=Sì 0=NO)

Esegui Incolla Reimposta Annulla Aiuto

- Per Campo campagna, selezionare *Pacchetto di controllo*.
- Per Campo risposta efficacia, selezionare *Ha risposto all'offerta di prova*.
- Selezionare *Replica*.
- Per Valore risposta positiva, selezionare *Sì* dall'elenco a discesa. Nel campo di testo viene visualizzato il valore 1 perché "Sì" è in realtà un'etichetta associata a un valore registrato di 1. Se per il valore della risposta positiva non è definita un'etichetta, è possibile immettere il valore desiderato direttamente nel campo di testo.

Viene automaticamente creato un nuovo campo, nel quale 1 rappresenta le risposte positive e 0 rappresenta le risposte negative. L'analisi viene eseguita sul nuovo campo. È possibile ignorare il nome e l'etichetta predefiniti e specificarne di diversi. In questo esempio viene utilizzato il nome di campo fornito in precedenza.

- Fare clic su *Esegui* per eseguire la procedura.

## Output

Figura 13-2  
Output di Test pacchetto di controllo

		Pacchetto di controllo			
		Controllo		Test	
		Conteggio	% casi colonna	Conteggio	% casi colonna
Efficacia (1=Si 0=No)	0	875	96.2%	945	93.8%
	1	35	3.8%	62	6.2%

È presente una differenza statisticamente significativa tra Controllo e Test.

L'output della procedura include una tabella che mostra conteggi e percentuali di risposte positive e negative per ogni gruppo definito dal Campo campagna e una tabella che indica se i tassi di risposta del gruppo differiscono in modo significativo dagli altri.

- *Efficacia* è la versione ricodificata del campo risposta, dove 1 rappresenta le risposte positive e 0 le risposte negative.
- Il tasso di risposta positiva per il pacchetto di controllo è del 3,8%, mentre il tasso di risposta positiva per il pacchetto di prova è del 6,2%.

La descrizione in testo semplice sotto la tabella indica che la differenza tra i gruppi è significativa, cioè che il tasso di risposta più elevato per il pacchetto di prova probabilmente non dipende dal caso. Questa tabella di testo conterrà un confronto per ogni possibile coppia di gruppi inclusa nell'analisi. Dal momento che in questi esempi vi sono solo due gruppi, il confronto è solo uno. Se i gruppi sono più di cinque, la tabella descrittiva in formato testo è sostituita dalla tabella Confronti delle proporzioni di colonna.

## Riepilogo

Il Test pacchetto di controllo confronta le campagne di marketing per rilevare eventuali differenze significative nell'efficacia di diversi pacchetti o offerte. In questo esempio, la risposta positiva del 6,2% per il pacchetto di prova è stata significativamente più alta del tasso di risposta positiva del 3,8% per il pacchetto di controllo. Ciò indica che è opportuno utilizzare la confezione con il nuovo design al posto di quella vecchia, ma possono esservi altri fattori da considerare quali gli eventuali costi aggiuntivi legati all'adozione del nuovo tipo di confezione.

## ***File di esempio***

Il file di esempio installato con il prodotto si trova nella sottodirectory *Samples* della directory di installazione. La sottodirectory *Samples* contiene cartelle separate 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).
- **aflatoxin20.sav.** Questo file di dati contiene le misurazioni di Aflatossina di ciascuno dei 16 campioni di quattro raccolti e otto campioni dal file di dati *aflatoxin.sav*.
- **anorectic.sav.** Per trovare una sintomatologia standardizzata del comportamento anoressico/bulimico, i ricercatori 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.

- **autoaccidents.sav.** File di dati ipotetici che prende in esame l'impegno di un analista che opera nel campo delle assicurazioni per creare un modello del numero di incidenti automobilistici per conducente. Il modello prende in esame anche l'età e il sesso del conducente. Ciascun caso rappresenta un diverso conducente e riporta il sesso e l'età (in anni) del conducente e il numero di incidenti automobilistici negli ultimi cinque anni.
- **band.sav.** Questo file di dati ipotetici contiene le cifre sulle vendite settimanali di CD conseguite da un gruppo musicale. Il file include anche i dati di tre possibili variabili predittore.
- **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.
- **behavior.sav.** In un classico esempio, 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 elevate prestazioni. 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, 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 insieme di dati presentato e analizzato altrove 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 scalato.

- **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 comprende versioni trasformate dei campi.
- **carpet.sav.** Come esempio tipico, un'azienda interessata alla commercializzazione di un nuovo battitappeto desidera esaminare l'influenza di cinque fattori sulle preferenze del consumatore, ovvero design della confezione, marca, prezzo, la presenza di un *marchio di qualità* e una garanzia "Soddisfatti o rimborsati". Esistono tre livelli di fattore per il design della confezione, che differiscono per la posizione 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* include il rango delle classificazioni medie 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.
- **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 churn, ovvero l'abbandono dei clienti. Agli account vengono applicati i punteggi relativi alla propensione al churn, con valori compresi tra 0 e 100. Gli account con punteggio pari a 50 o superiore è probabile che stiano cercando nuovi provider.
- **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 . 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 ipotetico 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*.
- **customers\_model.sav.** File di dati ipotetici che contiene il nominativo delle persone a cui è stata inviata una campagna di marketing. I dati includono informazioni demografiche, un riepilogo della cronologia degli acquisti e se ciascuna persona ha risposto alla campagna. Ogni caso rappresenta una persona separata.
- **customers\_new.sav.** File di dati ipotetici che contiene i nominativi delle persone che sono state evidenziate come potenziali candidati per una campagna di marketing. I dati includono informazioni demografiche e un riepilogo sulla cronologia degli acquisti di ciascuna persona. Ogni caso rappresenta una persona separata.
- **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 il primo passo che una società intraprende per compilare un database con informazioni ricavate dai sondaggi. Ogni caso rappresenta una diversa città. Sono registrate anche le informazioni sulla regione, provincia, distretto e città.

- **demo\_cs\_2.sav.** File di dati ipotetici che prende in esame il secondo passo che una società intraprende per compilare un database con informazioni ricavate dai sondaggi. Ogni caso rappresenta una diversa unità di abitazione, ricavata dalle città selezionate nel primo passo. 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.
- **dmdata.sav.** File di dati ipotetici che contiene informazioni demografiche e di acquisto di una società di direct marketing. *dmdata2.sav* contiene informazioni su un sottoinsieme di contatti che hanno ricevuto un mailing di prova e *dmdata3.sav* contiene informazioni sui contatti rimanenti che non hanno ricevuto il mailing di prova.
- **dietstudy.sav.** File di dati ipotetici che contiene il risultato di uno studio ipotetico sulla dieta chiamato “Stillman diet”. 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.
- **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 dall’insieme di dati “German Credit” del Repository of Machine Learning Databases 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 sui sondaggi 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 ha presentato una tabella per illustrare i possibili gruppi sociali. Guttman 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 vicinanza 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 vicinanza 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.
- **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. Un esperimento di laboratorio è stato condotto 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.** File di dati ipotetici che prende in esame una compagnia di assicurazioni impegnata nella creazione di un modello per contrassegnare le 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 le stesse prestazioni.
- **kinship\_dat.sav.** Rosenberg e Kim 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 similitudine. 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. Di conseguenza, sono state ottenute sei "sorgenti" in totale. Ogni sorgente corrisponde a una matrice di prossimità  $15 \times 15$ , le cui celle sono uguali al numero delle persone in una sorgente meno il numero di volte in cui gli oggetti sono stati ripartiti insieme nella sorgente.
- **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.** File di dati che prende in esame le vendite di abitazioni in un nuovo centro abitato in Algonquin, Ill., durante gli anni 1999–2000. Tali vendite sono una questione di dominio pubblico.

- **nhis2000\_subset.sav.** Il National Health Interview Survey (NHIS) è un sondaggio di grandi dimensioni condotto 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 dominio 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, e , hanno rilevato non linearità tra queste variabili, che impediscono un approccio di regressione standard.
- **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.
- **polishing.sav.** File di dati "Nambeware Polishing Times" di Data and Story Library. Prende in esame l'impegno di un'industria di stoviglie in metallo (Nambe Mills, Santa Fe, N. M.) nel tentativo di pianificare il proprio piano di produzione. Ogni caso rappresenta un diverso articolo nella linea dei prodotti. Per ciascun articolo sono indicati il diametro, il tempo di lucidatura, il prezzo e il tipo di prodotto.
- **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 selezionato 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 selezionato 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 trasgressore precedente, rilasciato dopo il primo arresto durante il mese di giugno del 2003 e registra le relative informazioni demografiche, alcuni dettagli sul primo crimine commesso e i dati del secondo arresto, se si è verificato prima della fine di giugno del 2006. I trasgressori sono stati selezionati dai dipartimenti sottoposti a campione in base al piano di campionamento specificato nel file *recidivism\_cs\_csplan*. Poiché viene utilizzato un 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 delle transazioni 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. Alla fine 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 ipotetico 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 .
- **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. I valori accettati sono compresi tra 4,5 e 5,5.
- **ships.sav.** Ad esempio, un insieme di dati presentato e analizzato altrove 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>) Il primo passo nell'analisi di questo file di dati consiste quindi nel pesare i dati per rispecchiare le tendenze della popolazione.
- **stroke\_clean.sav.** File di dati ipotetici che riporta lo stato di un database medico dopo averne eseguito la pulizia utilizzando le procedure del modulo Data Preparation.
- **stroke\_invalid.sav.** File di dati ipotetici che 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 alla fine 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 i valori di alcuni dati siano stati modificati e siano state aggiunte variabili fittizie a scopo dimostrativo.
- **telco.sav.** File di dati ipotetici che prende in esame l'impegno di un'azienda di telecomunicazioni nel tentativo di ridurre il churn, 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 di 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 dei valori.
- **tv-survey.sav.** File di dati ipotetici che prende in esame un sondaggio condotto da una emittente televisiva che deve stabilire se estendere la durata di un programma di successo. A un campione di 906 intervistati è stato chiesto se preferisce guardare il programma con diverse condizioni. Ciascuna riga rappresenta un diverso intervistato 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 .
- **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 alla fine. È stato presentato e analizzato in altri luoghi .
- **verd1985.sav.** Questo file di dati prende in esame un'indagine . 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.

- **wheeze\_steubenville.sav.** Questo file è un sottoinsieme di uno studio longitudinale degli effetti che l'inquinamento provoca sulla salute dei bambini . 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.



# Notices

Licensed Materials – Property of SPSS Inc., an IBM Company. © Copyright SPSS Inc. 1989, 2010.

Patent No. 7,023,453

**The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:** SPSS INC., AN IBM COMPANY, PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. SPSS Inc. may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-SPSS and non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this SPSS Inc. product and use of those Web sites is at your own risk.

When you send information to IBM or SPSS, you grant IBM and SPSS a nonexclusive right to use or distribute the information in any way it believes appropriate without incurring any obligation to you.

Information concerning non-SPSS products was obtained from the suppliers of those products, their published announcements or other publicly available sources. SPSS has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-SPSS products. Questions on the capabilities of non-SPSS products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

## COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to SPSS Inc., for the purposes of developing,

using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. SPSS Inc., therefore, cannot guarantee or imply reliability, serviceability, or function of these programs. The sample programs are provided “AS IS”, without warranty of any kind. SPSS Inc. shall not be liable for any damages arising out of your use of the sample programs.

### **Trademarks**

IBM, the IBM logo, and [ibm.com](http://www.ibm.com) are trademarks of IBM Corporation, registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>.

SPSS is a trademark of SPSS Inc., an IBM Company, registered in many jurisdictions worldwide.

Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java-based trademarks and logos are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.

This product uses WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Other product and service names might be trademarks of IBM, SPSS, or other companies.

Adobe product screenshot(s) reprinted with permission from Adobe Systems Incorporated.

Microsoft product screenshot(s) reprinted with permission from Microsoft Corporation.



analisi cluster, 15, 53

cluster, 15

file di esempio  
posizione, 100

legal notices, 111

profili dei potenziali clienti, 20, 70  
propensione all'acquisto, 34, 85

regressione logistica , 34, 85  
RFM, 2, 9, 11, 13, 46  
categorizzazione, 6  
dati clienti, 5  
dati transazione, 3, 46

tassi di risposta cap, 26  
tassi di risposta codice postale, 26, 77  
test pacchetto di controllo, 42, 97  
trademarks, 112