

IBM SPSS Complex Samples 20



Nota: Prima di utilizzare queste informazioni e il relativo prodotto, leggere le informazioni generali disponibili in Note legali a pag. 280.

Questa versione si applica a IBM® SPSS® Statistics 20 e a tutte le successive versioni e modifiche fino a eventuali disposizioni contrarie indicate in nuove versioni.

Le schermate dei prodotti Adobe sono state ristampate su autorizzazione di Adobe Systems Incorporated.

Le schermate dei prodotti Microsoft sono state ristampate su autorizzazione di Microsoft Corporation.

Materiali concessi in licenza - Proprietà di IBM

© **Copyright IBM Corporation 1989, 2011.**

Tutti i diritti riservati.

Prefazione

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

Informazioni su Business Analytics di IBM

Il software IBM Business Analytics fornisce informazioni complete, coerenti e accurate a cui i responsabili delle decisioni possono affidarsi per ottimizzare le prestazioni dell'azienda. Un ampio portafoglio di applicazioni di [business intelligence](#), [analisi predittiva](#), [gestione delle prestazioni e delle strategie finanziarie](#) e [analisi](#) offre una panoramica chiara, istantanea e interattiva delle prestazioni attuali e la possibilità di prevedere i risultati futuri. Utilizzato in combinazione con potenti soluzioni di settore, prassi consolidate e servizi professionali, questo software consente alle aziende di tutte le dimensioni di ottimizzare la produttività, automatizzare le decisioni senza problemi e fornire risultati migliori.

Come parte di questo portafoglio, il software IBM SPSS Predictive Analytics consente alle aziende di prevedere gli eventi futuri e di agire tempestivamente in modo da migliorare i risultati delle attività aziendali. Le aziende, gli enti governativi e le università di tutto il mondo si affidano alla tecnologia IBM SPSS perché rappresenta un vantaggio concorrenziale in termini di attrazione, retention e aumento dei clienti, riducendo al tempo stesso le frodi e limitando i rischi. Incorporando il software IBM SPSS nelle attività quotidiane, le aziende diventano imprese in grado di effettuare previsioni e di gestire e automatizzare le decisioni, per raggiungere gli obiettivi aziendali e vantaggi tangibili sulla concorrenza. Per ulteriori informazioni o per contattare un rappresentante, visitare il sito <http://www.ibm.com/spss>.

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 IBM Corp. o per l'installazione di uno degli ambienti hardware supportati. Per contattare il supporto tecnico, visitare il sito Web di IBM Corp. all'indirizzo <http://www.ibm.com/support>. 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.

Supporto tecnico per studenti

Gli studenti che utilizzano una versione accademica o grad pack di qualsiasi prodotto software IBM SPSS sono pregati di utilizzare le apposite pagine online per studenti [Solutions for Education](#) (<http://www.ibm.com/spss/rd/students/>). Gli studenti che utilizzano una copia del software IBM SPSS fornita dall'università, sono pregati di contattare il coordinatore del prodotto IBM SPSS presso l'università.

Servizio clienti

Per eventuali chiarimenti in merito alla spedizione o al proprio conto, rivolgersi alla sede locale. Tenere presente che sarà necessario fornire il numero di serie.

Corsi di formazione

IBM Corp. organizza corsi di formazione pubblici e onsite che includono esercitazioni pratiche. Tali corsi si terranno periodicamente nelle principali città. Per ulteriori informazioni su questi seminari, andare a <http://www.ibm.com/software/analytics/spss/training>.

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

1 Introduzione alle procedure di Campioni complessi 1

Proprietà dei campioni complessi	1
Uso delle procedure di Campioni complessi	2
File del piano	2
Approfondimenti	3

2 Campionamento da un disegno complesso 4

Per creare un nuovo piano di campionamento	4
Campionamento guidato: Variabili disegno	6
Controlli struttura per gli spostamenti all'interno di Campionamento guidato	7
Campionamento guidato: Metodo di campionamento	8
Campionamento guidato: Dimensione campione	10
Definisci dimensioni non uguali	11
Campionamento guidato: Variabili di output	12
Campionamento guidato: Riepilogo del piano	13
Campionamento guidato: Estrai campione - Opzioni di selezione	14
Campionamento guidato: Estrai campione - File di output	15
Campionamento guidato: Fine	16
Per modificare un piano di campionamento esistente	16
Campionamento guidato: Riepilogo del piano	17
Per eseguire un piano di campionamento esistente	18
Opzioni aggiuntive dei comandi CSPLAN e CSSELECT	18

3 Preparazione di un campione complesso per l'analisi 19

Per creare un nuovo piano analisi	19
Preparazione guidata dell'analisi: Variabili disegno	20
Controlli struttura per gli spostamenti all'interno dell'analisi guidata	21
Preparazione guidata dell'analisi: Metodo di stima	22
Preparazione guidata dell'analisi: Dimensione	23
Definisci dimensioni non uguali	24

Preparazione guidata dell'analisi: Riepilogo del piano	25
Preparazione guidata dell'analisi: Fine	26
Per modificare un piano analisi esistente	26
Preparazione guidata dell'analisi: Riepilogo del piano	27
4 Piano campioni complessi	28
5 Frequenze campioni complessi	29
Frequenze campioni complessi: Statistiche	30
Campioni complessi: Valori mancanti	31
Campioni complessi: Opzioni	32
6 Descrittive campioni complessi	33
Descrittive campioni complessi: Statistiche	34
Descrittive campioni complessi: Valori mancanti	35
Campioni complessi: Opzioni	36
7 Tavole di contingenza campioni complessi	37
Tavole di contingenza campioni complessi: Statistiche	39
Campioni complessi: Valori mancanti	40
Campioni complessi: Opzioni	41
8 Rapporti campioni complessi	42
Rapporti campioni complessi: Statistiche	43
Rapporti campioni complessi: Valori mancanti	44
Campioni complessi: Opzioni	44

9 Modello lineare generalizzato campioni complessi 46

Statistiche modello lineare generalizzato campioni complessi	49
Modello lineare generalizzato campioni complessi: test di ipotesi.	50
Medie stimate modello lineare generalizzato campioni complessi.	52
Modello lineare generalizzato campioni complessi - Salva	53
Opzioni modello lineare generalizzato campioni complessi	54
Opzioni aggiuntive del comando CSGLM	55

10 Regressione logistica campioni complessi 56

Categoria di riferimento regressione logistica campioni complessi	57
Modello regressione logistica campioni complessi	58
Statistiche regressione logistica campioni complessi	60
Modello lineare generalizzato campioni complessi: test di ipotesi.	61
Rapporto odd regressione logistica campioni complessi.	62
Regressione logistica campioni complessi - Salva	63
Opzioni regressione logistica campioni complessi	64
Opzioni aggiuntive del comando CSLOGISTIC	65

11 Regressione ordinale campioni complessi 66

Probabilità di risposta della Regressione ordinale campioni complessi	68
Modello della Regressione ordinale campioni complessi	68
Statistiche della Regressione ordinale campioni complessi.	70
Modello lineare generalizzato campioni complessi: test di ipotesi.	72
Rapporti odd regressione ordinale campioni complessi.	73
Salvataggio della regressione ordinale dei campioni complessi	74
Opzioni della regressione ordinale campioni complessi.	75
Funzioni aggiuntive del comando CSORDINAL	76

12 Regressione Cox di campioni complessi 77

Definisci evento	80
----------------------------	----

Predittori	81
Definisci predictor dipendente dal tempo	82
Sottogruppi	83
Modello	84
Statistiche	86
Grafici	88
Test di ipotesi	89
Salva	90
Esporta	92
Opzioni	94
Opzioni aggiuntive del comando CSCOXREG	95

Parte II: Esempi

13 Campionamento guidato di Campioni complessi **98**

Ottenere un campione da una struttura di campionamento completa	98
Utilizzo della procedura guidata	98
Riepilogo del piano	108
Riepilogo del campionamento	108
Risultati del campione	109
Ottenere un campione da una struttura di campionamento parziale	110
Utilizzo della procedura guidata per campionare dalla prima struttura parziale	110
Risultati del campione	123
Utilizzo della procedura guidata per campionare dalla seconda struttura parziale	123
Risultati del campione	128
Campionamento con PPS (Probability Proportional to Size)	128
Utilizzo della procedura guidata	129
Riepilogo del piano	140
Riepilogo del campionamento	140
Risultati del campione	142
Procedure correlate	144

14 Preparazione guidata dell'analisi Campioni complessi 145

Utilizzo della Preparazione guidata dell'analisi Campioni complessi per leggere i dati pubblici NHIS	145
Utilizzo della procedura guidata	145
Riepilogo	148
Preparazione per l'analisi quando i pesi di campionamento non sono inclusi nel file di dati	148
Calcolo delle probabilità di inclusione e dei pesi di campionamento	148
Utilizzo della procedura guidata	151
Riepilogo	158
Procedure correlate	159

15 Frequenze campioni complessi 160

Utilizzo di Frequenze campioni complessi per analizzare il consumo dei supplementi nutrizionali	160
Esecuzione dell'analisi.	160
Tabella di frequenza (Analisi delle corrispondenze)	163
Frequenza per sottopopolazione	163
Riepilogo	164
Procedure correlate	164

16 Descrittive campioni complessi 165

Utilizzo di Descrittive campioni complessi per l'analisi dei livelli di attività.	165
Esecuzione dell'analisi.	165
Statistiche univariate.	168
Statistiche univariate per sottopopolazione	168
Riepilogo	169
Procedure correlate	169

17 Tavole di contingenza campioni complessi 170

Utilizzo delle tavole di contingenza campioni complessi per misurare il coefficiente di rischio di un evento	170
Esecuzione dell'analisi.	170
Tavola di contingenza	173
Stima del rischio	174

Stima del rischio in base alla sottopopolazione	175
Riepilogo	175
Procedure correlate	175

18 Rapporti campioni complessi 177

Utilizzo dei rapporti campioni complessi per agevolare la valutazione del valore della proprietà	177
Esecuzione dell'analisi.	177
Rapporti.	180
Tabella pivot Rapporti	180
Riepilogo	181
Procedure correlate	181

19 Modello lineare generalizzato campioni complessi 183

Utilizzo del Modello lineare generalizzato campioni complessi per l'adattamento a un'ANOVA a due fattori.	183
Esecuzione dell'analisi.	183
Riepilogo del modello (Regressione output)	188
Test degli effetti del modello	188
Stime di parametri	189
Medie marginali stimate	190
Riepilogo	192
Procedure correlate	192

20 Regressione logistica campioni complessi 194

Utilizzo della Regressione logistica campioni complessi per la valutazione del rischio di credito.	194
Esecuzione dell'analisi.	194
Pseudo R-quadrati.	198
Classificazione	199
Test degli effetti del modello	200
Stime di parametri	200
Rapporti Odd	201
Riepilogo	202
Procedure correlate	203

21 *Regressione ordinale campioni complessi*

204

Utilizzo di Regressione ordinale campioni complessi per l'analisi dei risultati di indagini	204
Esecuzione dell'analisi.	204
Pseudo R-quadrati.	209
Test degli effetti del modello	210
Stime di parametri	210
Classificazione	212
Rapporti Odd	213
Modello cumulativo generalizzato	215
Rimozione dei predittori non significativi	216
Avvisi	218
Confronto di modelli.	219
Riepilogo	220
Procedure correlate	220

22 *Regressione Cox di campioni complessi*

221

Uso di Regressione Cox di campioni complessi: Definisci predictor dipendente dal tempo	221
Preparazione dei dati.	221
Esecuzione dell'analisi.	227
Informazioni disegno di campionamento	232
Test degli effetti del modello	233
Test dei rischi proporzionali	233
Aggiunta di un predittore dipendente dal tempo.	234
Più casi per soggetto in Regressione Cox di campioni complessi	237
Preparazione dei dati per l'analisi	238
Creazione di un piano di analisi di campionamento casuale semplice.	253
Esecuzione dell'analisi.	257
Informazioni disegno di campionamento	265
Test degli effetti del modello	266
Stime di parametri	266
Valori modello	267
Grafico log meno log	268
Riepilogo	268

Appendici

A File di esempio **270**

B Note legali **280**

Bibliografia **283**

Indice **285**

Parte I:
Manuale dell'utente

Introduzione alle procedure di Campioni complessi

Le procedure di analisi usate dai software tradizionali assumono la rappresentazione di un campione casuale semplice della popolazione desiderata nelle osservazioni di un file di dati. Un numero sempre crescente di società e di ricercatori considera ormai superata questa ipotesi e ritiene più vantaggioso dal punto di vista economico creare i campioni in un modo maggiormente strutturato.

Il modulo Campioni complessi consente di selezionare un campione in base a un disegno complesso e di inserire le specifiche del disegno nell'analisi dei dati, garantendo la validità dei risultati.

Proprietà dei campioni complessi

Le differenze tra un campione complesso e un campione casuale semplice possono essere molteplici. In un campione casuale semplice, le singole unità campionarie sono selezionate in modo casuale con probabilità uguali e senza sostituzione dall'intera popolazione. Per un campione complesso è invece possibile specificare alcune o tutte le seguenti caratteristiche:

Stratificazione. Il campionamento stratificato comporta la selezione di campioni indipendenti all'interno di sottogruppi non sovrapposti della popolazione, o strati. Gli strati possono essere rappresentati, ad esempio, da gruppi socioeconomici, categorie lavorative, gruppi di età o gruppi etnici. La stratificazione consente di garantire dimensioni del campione adeguate per i sottogruppi desiderati, di aumentare la precisione delle stime complessive e di utilizzare metodi di campionamento diversi in base allo strato.

Raggruppamento. Il campionamento di raggruppamenti comporta la selezione di gruppi di unità campionarie o raggruppamenti. I raggruppamenti possono ad esempio essere rappresentati da scuole, ospedali o aree geografiche e le unità campionarie possono essere gli studenti, i pazienti o i cittadini. I raggruppamenti vengono in genere usati nei campioni con disegni a più stadi e ad area (geografici).

Stadi multipli. Nel campionamento a più stadi, è possibile selezionare un campione di primo stadio basato sui raggruppamenti. È quindi possibile creare un campione di secondo stadio mediante l'estrazione di sottocampioni dai raggruppamenti selezionati. Se il campione di secondo stadio è basato su sottoraggruppamenti, è quindi possibile aggiungere un terzo stadio al campione. Ad esempio, il primo stadio di un'indagine può prevedere l'estrazione di un campione di città. Dalle città selezionate è quindi possibile campionare le abitazioni e infine dalle abitazioni è possibile estrarre gli individui. Per specificare i tre stadi di un disegno è possibile usare Campionamento guidato e Preparazione guidata dell'analisi.

Campionamento non casuale. Se è difficile ottenere la selezione casuale, è possibile campionare le unità in modo sistematico (a un intervallo fisso) o sequenziale.

Probabilità di selezione non uguali. Se si campionano raggruppamenti che contengono numeri di unità non uguali, è possibile usare il campionamento PPS (Probability Proportional to Size) per fare in modo che la probabilità di selezione di un raggruppamento sia uguale alla proporzione di unità in esso contenute. Il campionamento PPS può inoltre usare schemi di ponderazione più generali per la selezione delle unità.

Campionamento senza limiti. Il campionamento senza limiti seleziona le unità con sostituzione (WR). Quindi, è possibile selezionare una singola unità più volte per il campione.

Pesi di campionamento. I pesi di campionamento vengono calcolati automaticamente durante l'estrazione di un campione complesso e corrispondono idealmente alla frequenza rappresentata da ciascuna unità campionaria all'interno della popolazione desiderata. La somma dei pesi relativi al campione rappresenta pertanto una stima dell'ampiezza popolazione. Per poter analizzare in modo corretto un campione complesso, è necessario che le procedure di analisi di Campioni complessi prendano in considerazione i pesi di campionamento. Tenere presente che questi pesi devono essere utilizzati unicamente con il modulo Complex Samples e non con altre procedure analitiche mediante la procedura Pesa casi, che considera i pesi come repliche di casi.

Uso delle procedure di Campioni complessi

L'uso delle procedure di Campioni complessi varia in base alle specifiche esigenze. I principali tipi di utenti hanno le seguenti esigenze:

- Pianificare ed eseguire indagini basate su disegni complessi e possibilmente analizzare il campione in un secondo momento. Lo strumento principale degli studiosi è [Campionamento guidato](#).
- Analizzare i file di dati dei campioni ottenuti in precedenza sulla base di disegni complessi. Prima di usare le procedure di analisi di Campioni complessi, potrebbe essere necessario utilizzare la [Preparazione guidata dell'analisi](#).

Indipendentemente dalle proprie esigenze, è necessario specificare le informazioni relative al disegno nelle procedure di Campioni complessi. Queste informazioni vengono memorizzate in un **file del piano** che può essere riutilizzato.

File del piano

Un file del piano contiene le specifiche del campione complesso. Esistono due tipi di file del piano:

Piano di campionamento. Le specifiche inserite in Campionamento guidato definiscono un disegno di campionamento mediante il quale è possibile estrarre un campione complesso. Tali specifiche sono contenute nel file del piano di campionamento, il quale contiene anche un piano analisi predefinito che usa i metodi di stima appropriati per il disegno di campionamento specificato.

Piano analisi. Questo file del piano contiene le informazioni necessarie alle procedure di analisi di Campioni complessi per calcolare in modo corretto le stime della varianza relative a un campione complesso. Il piano include la struttura del campione, i metodi di stima per ciascuno stadio e i riferimenti alle variabili richieste, ad esempio i pesi del campione. Preparazione guidata dell'analisi consente di creare e di modificare i piani analisi.

Il salvataggio delle specifiche in un file del piano offre numerosi vantaggi, ad esempio:

- Uno studioso può specificare il primo stadio di un piano di campionamento a più stadi ed estrarre immediatamente le unità del primo stadio, raccogliere le informazioni sulle unità campionarie del secondo stadio e quindi modificare il piano di campionamento in modo da includervi il secondo stadio.
- Un analista che non è in grado di accedere al file del piano di campionamento, può definire un piano analisi e farvi riferimento da ciascuna procedura di analisi di Campioni complessi.
- Un progettista di campioni a uso pubblico su larga scala può pubblicare il file del piano di campionamento. In questo modo potrà semplificare le istruzioni per gli analisti, evitando loro di dover specificare i relativi piani analisi.

Approfondimenti

Per ulteriori informazioni sulle tecniche di campionamento, vedere i seguenti testi:

Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.

Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.

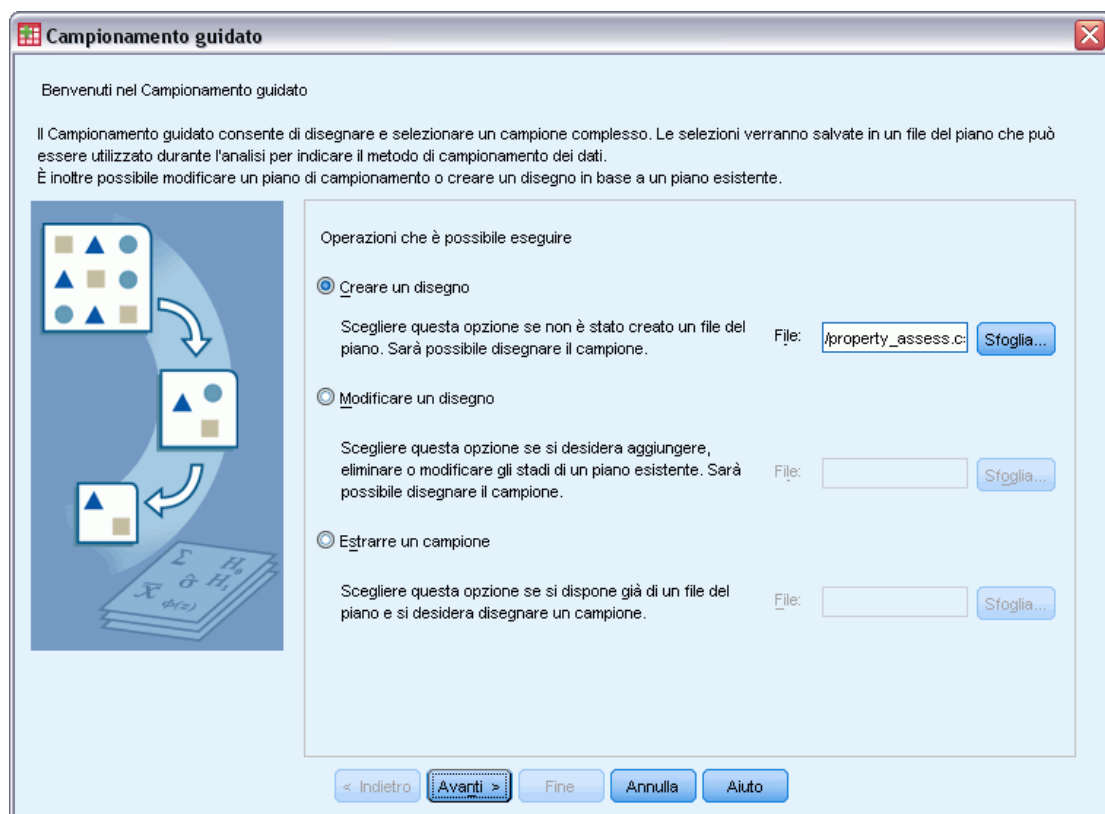
Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.

Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.

Särndal, C., B. Swensson, e J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Campionamento da un disegno complesso

Figura 2-1
Campionamento guidato, Finestra di benvenuto



Campionamento guidato consente di eseguire le varie fasi di creazione, modifica o esecuzione di un file del piano di campionamento. Per usare la procedura guidata, è necessario definire correttamente una popolazione di destinazione, disporre di un elenco di unità di campionamento e avere in mente un disegno di campionamento.

Per creare un nuovo piano di campionamento

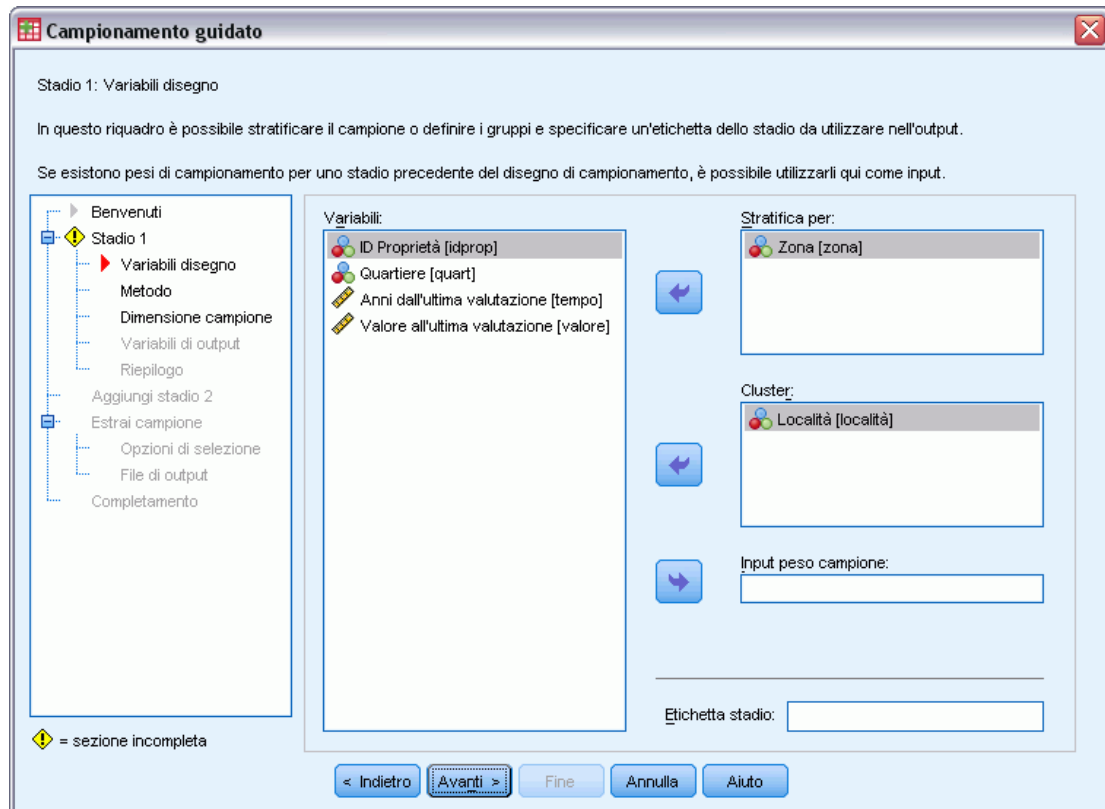
- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Seleziona un campione...
- ▶ Selezionare Creare un disegno e scegliere un nome per il file del piano in cui si desidera salvare il piano di campionamento.

- ▶ Per proseguire la procedura guidata, fare clic su Avanti.
- ▶ Se necessario, nel passo Variabili disegno è possibile definire gli strati, i cluster e l'input dei pesi del campione. Dopo avere specificato le opzioni desiderate, fare clic su Avanti.
- ▶ Se necessario, nel passo Metodo di campionamento è possibile scegliere un metodo di selezione degli elementi.

Se si seleziona PPS di Brewer o PPS di Murthy, è possibile fare clic su Fine per estrarre il campione. In caso contrario, fare clic su Avanti e quindi procedere come segue:
- ▶ Nel passo Dimensione campione, specificare il numero o la proporzione di unità da campionare.
- ▶ Fare quindi clic su Fine per estrarre il campione.
Se necessario, nei passi successivi è possibile:
 - Scegliere le variabili di output da salvare.
 - Aggiungere un secondo o un terzo stadio al disegno.
 - Impostare varie opzioni di selezione, ad esempio gli stadi da cui estrarre i campioni, il seme dei numeri casuali e se i valori mancanti definiti dall'utente dovranno essere considerati come valori validi delle variabili disegno.
 - Specificare la posizione in cui verranno salvati i dati di output.
 - Incollare le selezioni come sintassi dei comandi.

Campionamento guidato: Variabili disegno

Figura 2-2
Campionamento guidato - Passo Variabili disegno



In questo passo è possibile selezionare le variabili di stratificazione e di raggruppamento e definire l'input dei pesi del campione. È inoltre possibile definire un'etichetta per lo stadio.

Stratifica per. La classificazione incrociata delle variabili di stratificazione consente di definire sottopopolazioni distinte, o strati. Da ciascuno strato si ottengono campioni distinti. Per aumentare la precisione delle stime, è opportuno che le unità all'interno degli strati siano il più possibile omogenee in relazione alle caratteristiche desiderate.

Raggruppamenti. Le variabili di raggruppamento definiscono gruppi di unità di osservazione, o raggruppamenti. I raggruppamenti sono utili quando campionare direttamente unità di osservazione dalla popolazione è costoso o impossibile. È invece possibile campionare gruppi dalla popolazione e poi unità di osservazione dai gruppi selezionati. L'uso dei raggruppamenti può tuttavia comportare correlazioni fra le unità di campionamento e una conseguente perdita di precisione. Per ridurre al minimo questo effetto, è opportuno che le unità all'interno dei gruppi siano il più possibile eterogenee in relazione alle caratteristiche desiderate. Per creare il piano di un disegno a più stadi è necessario definire almeno una variabile di raggruppamento. I raggruppamenti sono inoltre necessari per l'uso di numerosi metodi di campionamento. [Per ulteriori informazioni, vedere l'argomento Campionamento guidato: Metodo di campionamento a pag. 8.](#)

Input peso campione. Se il disegno di campionamento corrente rientra in un disegno più ampio, potrebbero essere disponibili pesi del campione di uno stadio precedente di tale disegno. Nel primo stadio del disegno corrente è possibile specificare una variabile numerica che contiene tali pesi. I pesi vengono calcolati automaticamente per gli stadi successivi del disegno corrente.

Etichetta stadio. Per ciascuno stadio è possibile specificare un'etichetta facoltativa sotto forma di stringa, che nell'output semplificherà l'identificazione delle informazioni per stadi.

Nota: L'elenco di variabili sorgente è sempre uguale in tutti i passi della procedura guidata. Ciò significa che se in un passo viene rimossa una variabile dall'elenco, tale variabile non sarà disponibile negli stessi elenchi di tutti gli altri passi. Le variabili reinserite nell'elenco sorgente sono disponibili in tutti i passi.

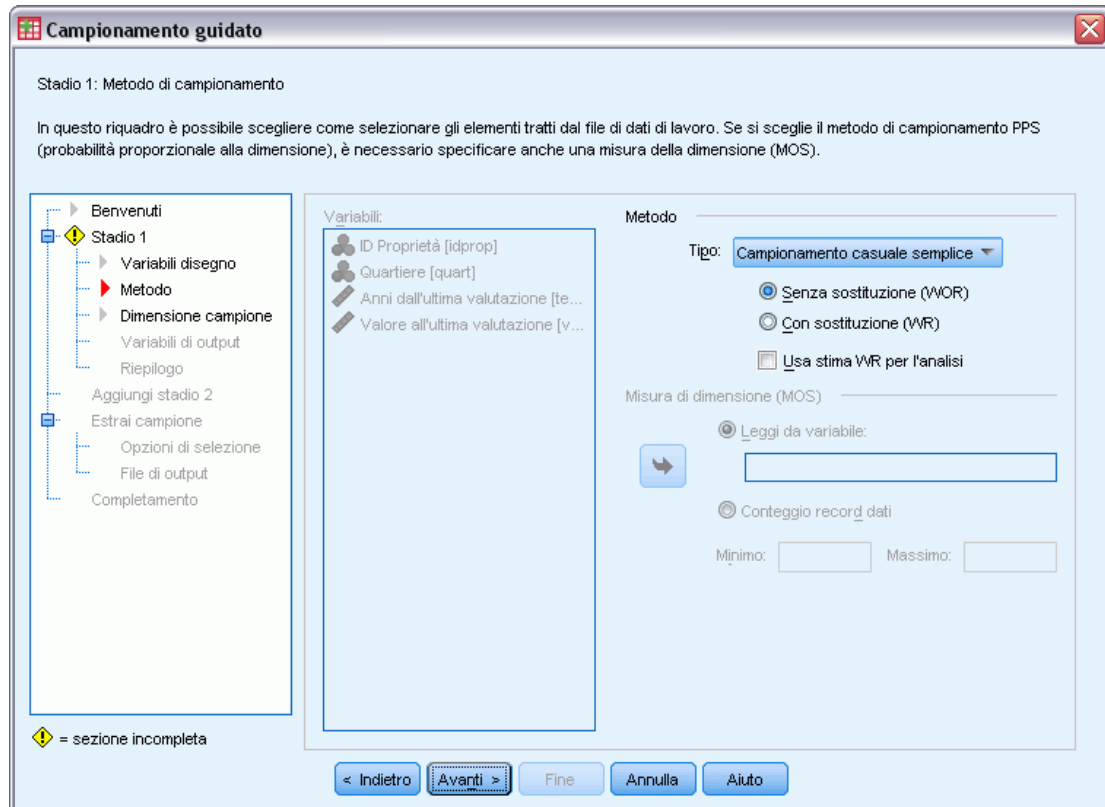
Controlli struttura per gli spostamenti all'interno di Campionamento guidato

Nella parte sinistra di ciascun passo di Campionamento guidato è disponibile una vista riassuntiva di tutti i passi, all'interno della quale è possibile fare clic sul nome di un passo attivo per spostarsi nella procedura guidata. I passi vengono attivati se tutti i passi precedenti sono validi, ovvero se per ciascun passo precedente sono state fornite le specifiche minime richieste. Per ulteriori informazioni sui motivi per cui un passo può non essere valido, vedere gli argomenti dell'Aiuto relativi ai singoli passi.

Campionamento guidato: Metodo di campionamento

Figura 2-3

Campionamento guidato - Passo Metodo di campionamento



Questo passo consente di specificare la modalità di selezione dei casi dal file di dati attivo.

Metodo. I controlli di questo gruppo consentono di scegliere un metodo di selezione. Alcuni tipi di campionamento consentono di scegliere se eseguire il campionamento con sostituzione (WR) o senza sostituzione (WOR). Per ulteriori informazioni, vedere le descrizioni dei diversi tipi. Si noti che alcuni tipi PPS (Probability Proportional to Size) sono disponibili solo dopo che sono stati definiti i raggruppamenti e che tutti i tipi PPS sono disponibili solo nel primo stadio del disegno. In generale, i metodi WR sono disponibili solo nel primo stadio di un disegno.

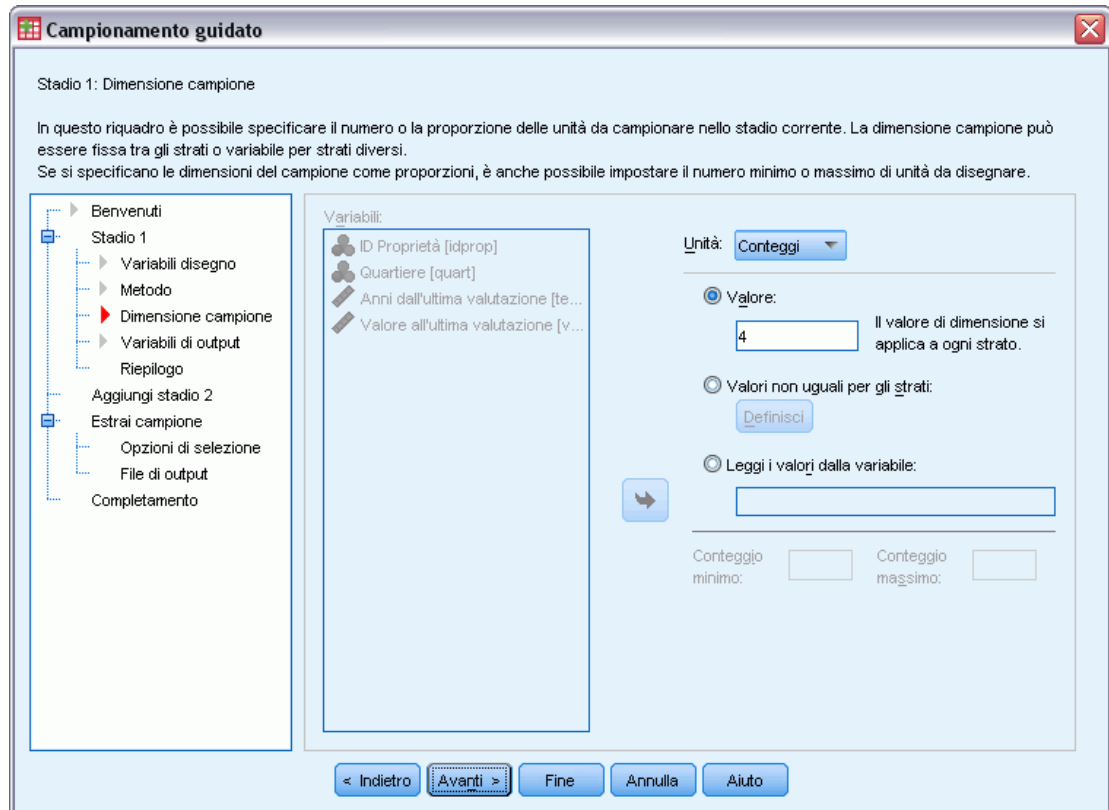
- **Campionamento casuale semplice.** Le unità vengono selezionate con probabilità uguale e possono essere selezionate con o senza sostituzione.
- **Sistematico semplice.** Le unità vengono selezionate a un intervallo fisso nella struttura di campionamento (o negli strati, se sono stati specificati) e vengono estratte senza sostituzione. Come punto di partenza viene scelta un'unità a selezione casuale all'interno del primo intervallo
- **Sequenziale semplice.** Le unità vengono selezionate in modo sequenziale con probabilità uguale e senza sostituzione.

- **PPS.** Metodo del primo stadio che consente di selezionare le unità in modo casuale con probabilità proporzionale alla dimensione. Tutte le unità possono essere selezionate con sostituzione, ma solo i raggruppamenti possono essere campionati senza sostituzione.
- **PPS sistematico.** Metodo del primo stadio che consente di selezionare le unità in modo sistematico con probabilità proporzionale alla dimensione. Le unità vengono selezionate senza sostituzione.
- **PPS sequenziale.** Metodo del primo stadio che consente di selezionare le unità in modo sequenziale con probabilità proporzionale alla dimensione del raggruppamento e senza sostituzione.
- **PPS di Brewer.** Metodo del primo stadio che consente di selezionare due raggruppamenti da ciascun strato con probabilità proporzionale alla dimensione del raggruppamento e senza sostituzione. Per usare questo metodo è necessario specificare una variabile di raggruppamento.
- **PPS di Murthy.** Metodo del primo stadio che consente di selezionare due raggruppamenti da ciascun strato con probabilità proporzionale alla dimensione del raggruppamento e senza sostituzione. Per usare questo metodo è necessario specificare una variabile di raggruppamento.
- **PPS di Sampford.** Metodo del primo stadio che consente di selezionare più di due raggruppamenti da ciascun strato con probabilità proporzionale alla dimensione del raggruppamento e senza sostituzione. È un'estensione del metodo di Brewer. Per usare questo metodo è necessario specificare una variabile di raggruppamento.
- **Usa stima WR per l'analisi.** Nel file del piano viene specificato per impostazione predefinita un metodo di stima coerente con il metodo di campionamento selezionato. In questo modo è possibile usare la stima con sostituzione anche se il metodo di campionamento implica una stima WOR. Questa opzione è disponibile solo nello stadio 1.

Misura di dimensione (MOS). Se si seleziona un metodo PPS, è necessario specificare una misura di dimensione che definisce la dimensione di ciascuna unità. Le dimensioni possono essere definite in modo esplicito in una variabile oppure possono essere calcolate dai dati. Se necessario, è possibile impostare un limite inferiore o superiore alla misura della dimensione, a prescindere dai valori presenti nella variabile o calcolati dai dati. Queste opzioni sono disponibili solo nello stadio 1.

Campionamento guidato: Dimensione campione

Figura 2-4
Campionamento guidato - Passo Dimensione campione



Questo passo consente di specificare il numero o la proporzione di unità da campionare nello stadio corrente. La dimensione del campione può essere fissa o variare in base agli strati. Ai fini della specifica della dimensione del campione, è possibile definire gli strati mediante i raggruppamenti selezionati negli stadi precedenti.

Unità. È possibile specificare una dimensione del campione esatta o una proporzione di unità da campionare.

- **Valore.** A tutti gli strati viene applicato un singolo valore. Se come metrica di unità è selezionato Conteggi, è necessario inserire un intero positivo. Se è selezionato Proporzioni, è necessario inserire un valore non negativo. A meno che non venga eseguito il campionamento con sostituzione, i valori della proporzione non devono essere maggiori di 1.
- **Valori non uguali per gli strati.** Consente di inserire dimensioni che variano in base allo strato mediante la finestra di dialogo Definisci dimensioni non uguali.
- **Leggi i valori dalla variabile.** Consente di selezionare una variabile numerica che contiene le dimensioni per gli strati.

Se è selezionato Proporzioni, è possibile impostare per il limite inferiore e il limite superiore il numero di unità campionate.

Definisci dimensioni non uguali

Figura 2-5
Finestra di dialogo Definisci dimensioni non uguali



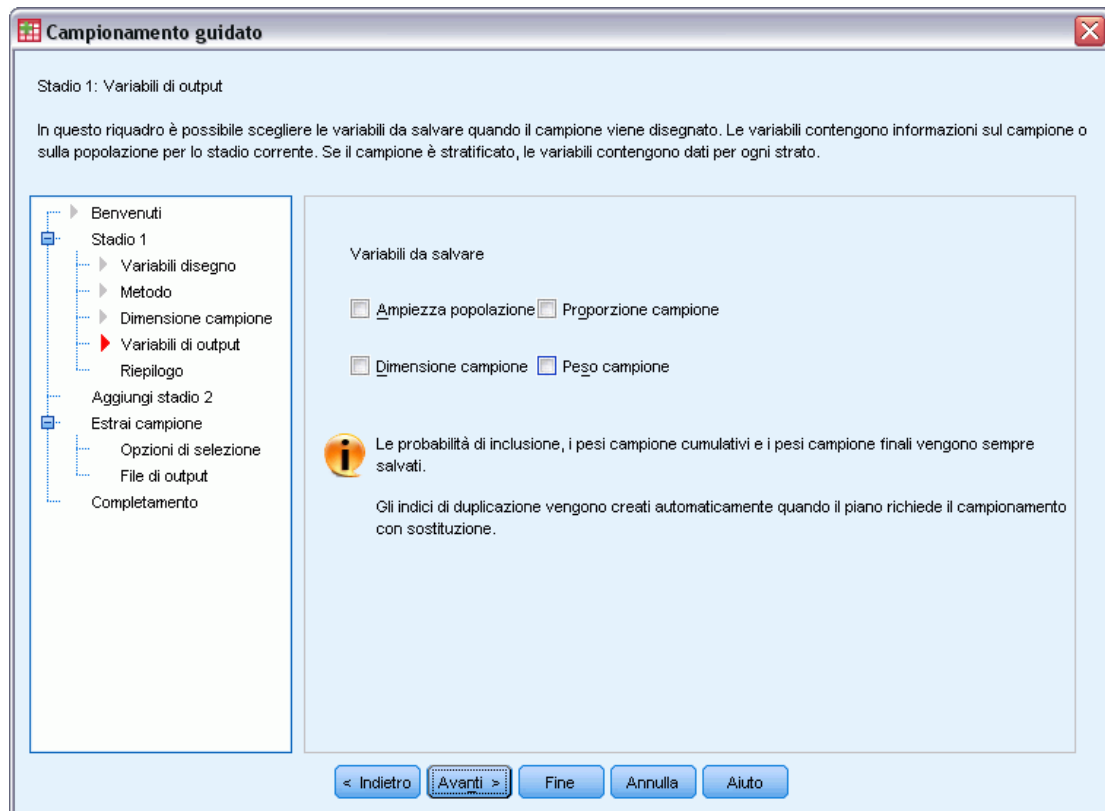
Nella finestra di dialogo Definisci dimensioni non uguali è possibile inserire le dimensioni di ciascuno strato.

Griglia Specifiche di dimensione. Nella griglia è visualizzata la classificazione incrociata di un massimo di cinque variabili di stratificazione o di raggruppamento,—con una combinazione strato/raggruppamento per ciascuna riga. Le variabili della griglia valide sono tutte le variabili di stratificazione dello stadio corrente e degli stadi precedenti e tutte le variabili di raggruppamento degli stadi precedenti. È possibile riordinare le variabili all'interno della griglia oppure spostarle nell'elenco Escludi. Inserire le dimensioni nella colonna all'estrema destra. Per visualizzare o nascondere le etichette dei valori e i valori delle variabili di stratificazione e di raggruppamento all'interno delle celle della griglia, fare clic su Etichette o Valori. Nelle celle che contengono valori senza etichette vengono sempre visualizzati i valori. Per reinserire nella griglia ciascuna combinazione di valori con etichetta per le variabili rappresentate, fare clic su Ridisegna gli strati.

Escludi. Per specificare le dimensioni di un sottoinsieme di combinazioni di strato/raggruppamento, spostare una o più variabili nell'elenco Escludi. Tali variabili non vengono usate per definire le dimensioni dei campioni.

Campionamento guidato: Variabili di output

Figura 2-6
Campionamento guidato - Passo Variabili di output



Questo passo consente di scegliere le variabili da salvare durante l'estrazione del campione.

Ampiezza popolazione. Il numero stimato di unità nella popolazione a uno stadio specifico. Il nome radice della variabile salvata è *PopulationSize_*.

Proporzione campione. La frequenza di campionamento a uno stadio specifico. Il nome radice della variabile salvata è *SamplingRate_*.

Dimensione campione. Il numero di unità estratte a uno stadio specifico. Il nome radice della variabile salvata è *SampleSize_*.

Peso campione. L'inverso delle probabilità di inclusione. Il nome radice della variabile salvata è *SampleWeight_*.

Alcune variabili per stadi vengono generate automaticamente, ad esempio:

Probabilità di inclusione. La proporzione di unità estratte a uno stadio specifico. Il nome radice della variabile salvata è *InclusionProbability_*.

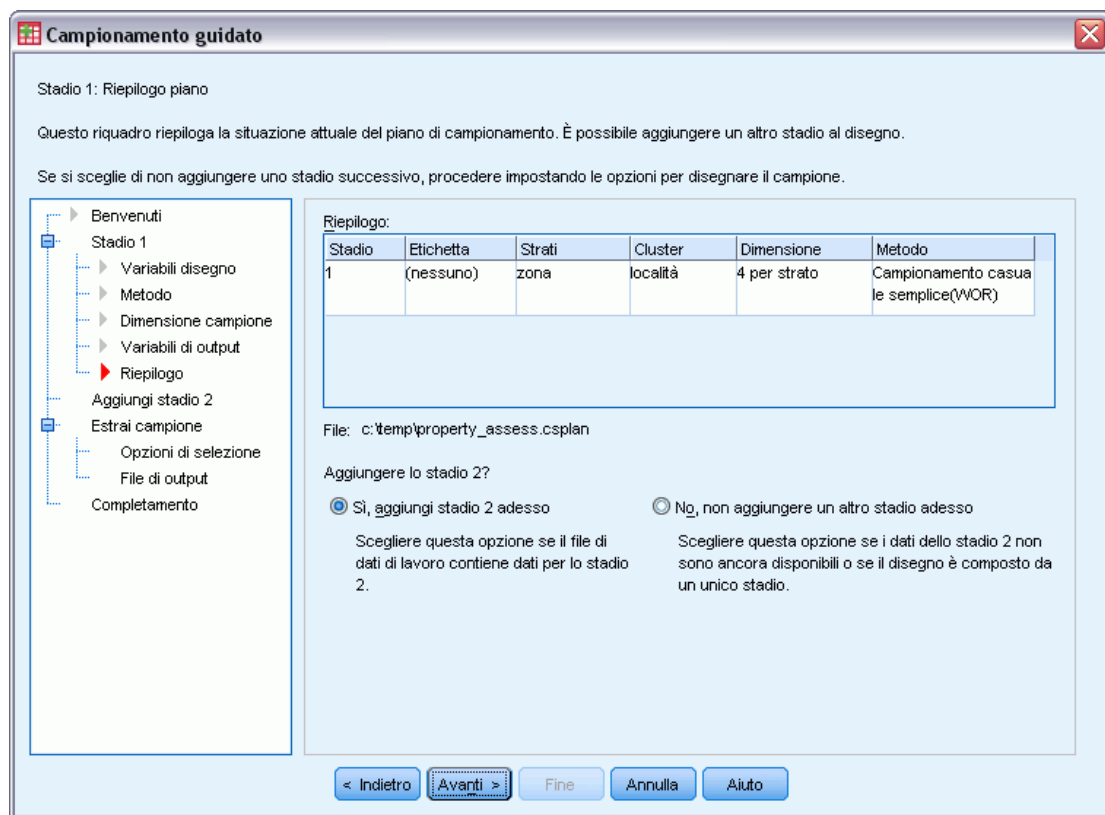
Peso cumulato. Peso cumulato del campione negli stadi precedenti e incluso lo stadio corrente. Il nome radice della variabile salvata è *SampleWeightCumulative_*.

Indice. Identifica le unità selezionate più volte all'interno di uno stadio specifico. Il nome radice della variabile salvata è *Index_*.

Nota: I nomi radice delle variabili salvate includono un suffisso rappresentato da un intero che specifica il numero dello stadio, ad esempio *PopulationSize_1_* rappresenta l'ampiezza popolazione salvata per lo stadio 1.

Campionamento guidato: Riepilogo del piano

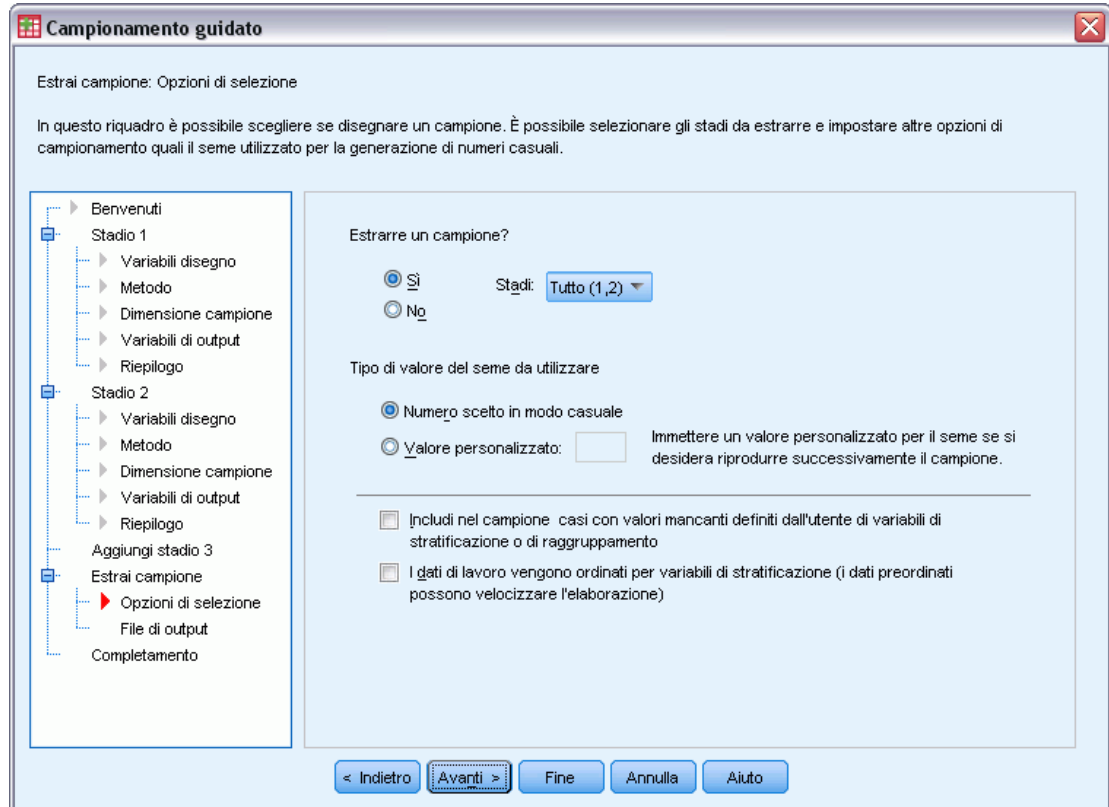
Figura 2-7
Campionamento guidato - Passo Riepilogo del piano



Questo è l'ultimo passo di ciascunostadio e fornisce un riepilogo delle specifiche di disegno dell'analisi fino allo stadio corrente. È quindi possibile procedere allo stadio successivo (creandolo se necessario) oppure impostare le opzioni per l'estrazione del campione.

Campionamento guidato: Estrai campione - Opzioni di selezione

Figura 2-8
Campionamento guidato, Passo Estrai campione - Opzioni di selezione



Questo passo consente di scegliere se estrarre un campione. È inoltre possibile impostare altre opzioni di campionamento, ad esempio il seme casuale e la gestione dei valori mancanti.

Estrai campione. Oltre a scegliere se estrarre o meno un campione, è possibile scegliere di eseguire una parte del disegno di campionamento. L'estrazione degli stadi deve essere eseguita in successione, ovvero non è possibile estrarre lo stadio 2 se non viene estratto anche lo stadio 1. Durante la modifica o l'esecuzione di un piano non è possibile ricampionare gli stadi bloccati.

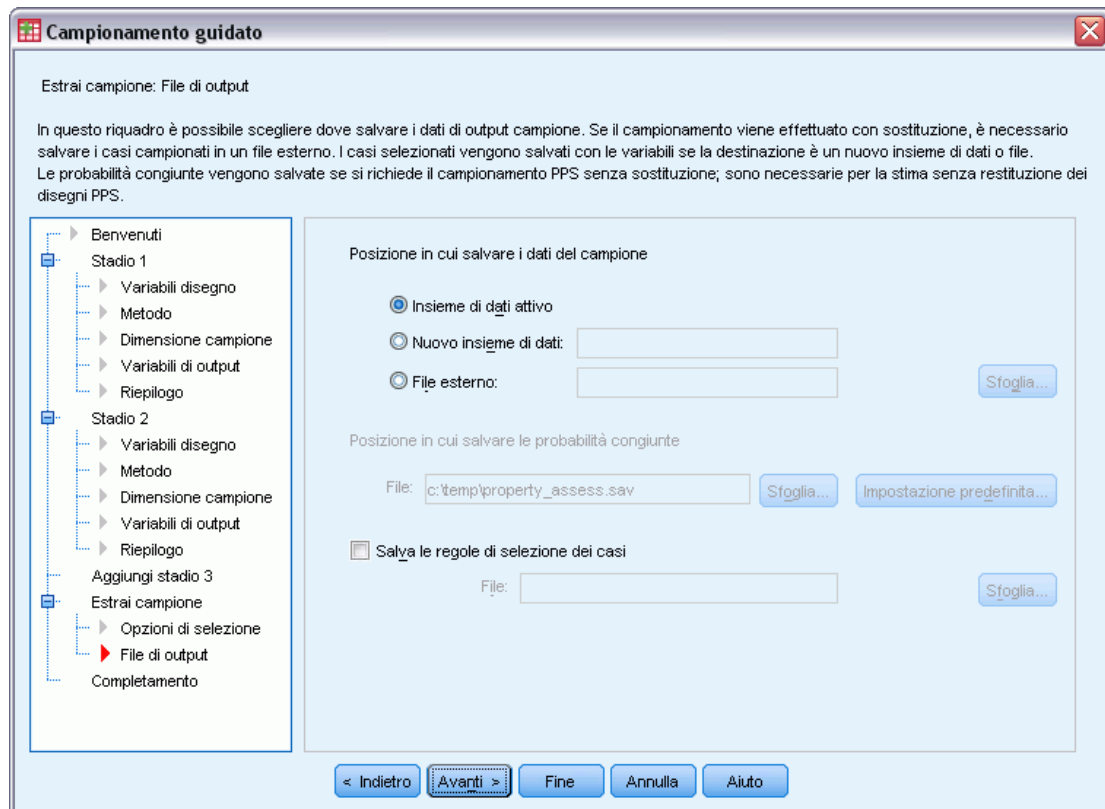
Seme. Consente di scegliere un seme per la generazione di numeri casuali.

Includi i valori mancanti definiti dall'utente. Determina se i valori mancanti definiti dall'utente sono validi. Se sono validi, verranno considerati come una categoria distinta.

Dati già ordinati. Se il campione è preordinato in base ai valori delle variabili di stratificazione, consente di velocizzare il processo di selezione.

Campionamento guidato: Estrai campione - File di output

Figura 2-9
Campionamento guidato - Passo Estrai campione: File di output



Questo passo consente di scegliere la posizione in cui inserire i casi campionati, le variabili di peso, le probabilità congiunte e le regole di selezione dei casi.

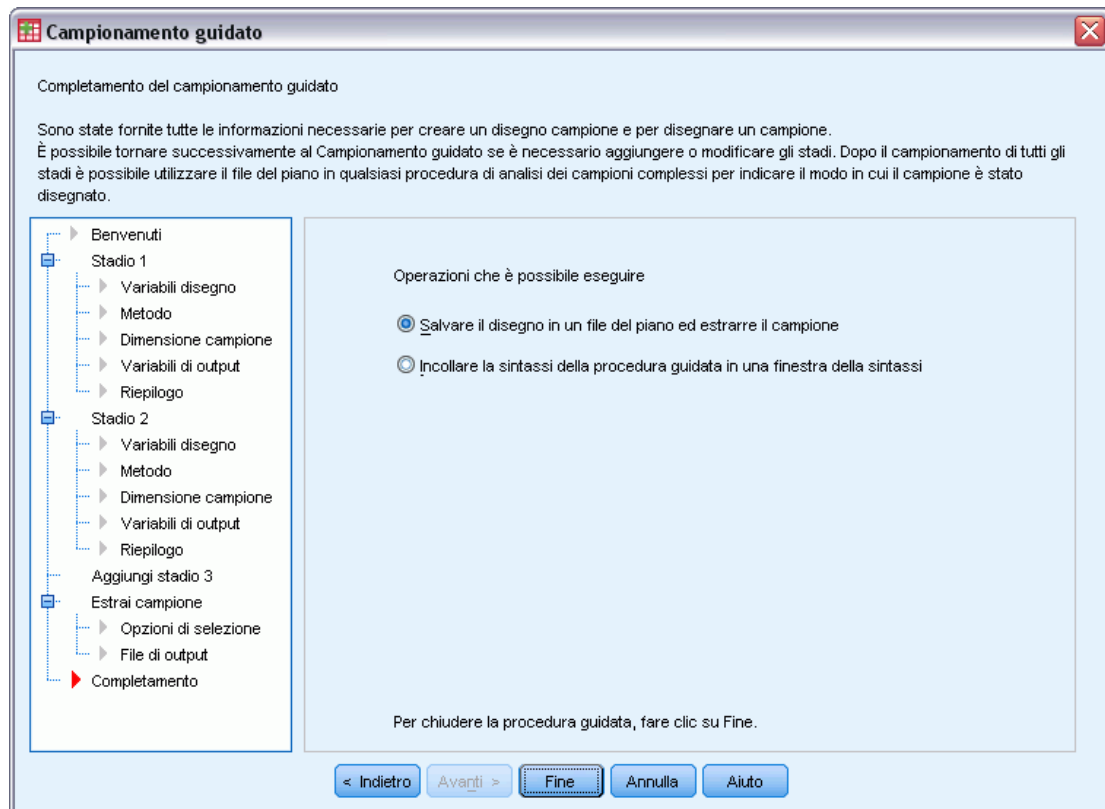
Dati del campione. Queste opzioni consentono di specificare la posizione in cui verrà scritto l'output del campione, che è possibile aggiungere al file di dati attivo, scrivere in un file di dati nuovo o salvare in un file di dati esterno in formato IBM® SPSS® Statistics. I file di dati sono disponibili durante la sessione corrente, ma non lo sono in quelle successive a meno che non li si salvi esplicitamente come file di dati. I nomi degli insiemi di dati devono rispettare le regole dei nomi delle variabili. Se si specifica un file esterno o un file di dati nuovo, le variabili di output del campionamento e le variabili del file di dati attivo per i casi selezionati verranno scritte in tale file.

Probabilità congiunte. Queste opzioni consentono di specificare la posizione in cui verranno scritte le probabilità congiunte, che vengono salvate in un file di dati esterno in formato SPSS Statistics. Le probabilità congiunte vengono prodotte se viene selezionato il metodo PPS WOR, PPS di Brewer, PPS di Sampford o PPS di Murthy e se non è specificata la stima WR.

Regole di selezione dei casi. Se si crea il campione utilizzando uno stadio alla volta, è possibile salvare le regole di selezione dei casi in un file di testo. Tali regole risultano utili per creare la sottostruttura degli stadi successivi.

Campionamento guidato: Fine

Figura 2-10
Campionamento guidato, Passo Fine



Passo finale. È possibile salvare ora il file del piano ed estrarre il campione oppure incollare le selezioni in una finestra di sintassi.

Quando si apportano modifiche agli stadi nel file del piano esistente, è possibile salvare il piano modificato in un nuovo piano oppure sovrascrivere il file esistente. Quando si aggiungono stadi senza apportare modifiche a quelle esistenti, la procedura guidata sovrascrive automaticamente il file del piano esistente. Per salvare il piano in un nuovo file, selezionare Incolla la sintassi generata dalla procedura guidata nella finestra della sintassi e modificare il nome del file nei comandi della sintassi.

Per modificare un piano di campionamento esistente

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Seleziona un campione...
- ▶ Selezionare Modificare un disegno e scegliere un file del piano da modificare.
- ▶ Per proseguire la procedura guidata, fare clic su Avanti.

- Rivedere il piano di campionamento nel passo Riepilogo del piano, quindi fare clic su Avanti.

I passi successivi sono molto simili a quelli previsti per un nuovo disegno. Per ulteriori informazioni, vedere gli argomenti dell’Aiuto relativi ai singoli passi.

- Andare al passo Fine e specificare un nuovo nome per il file del piano modificato oppure scegliere di sovrascrivere il file del piano esistente.

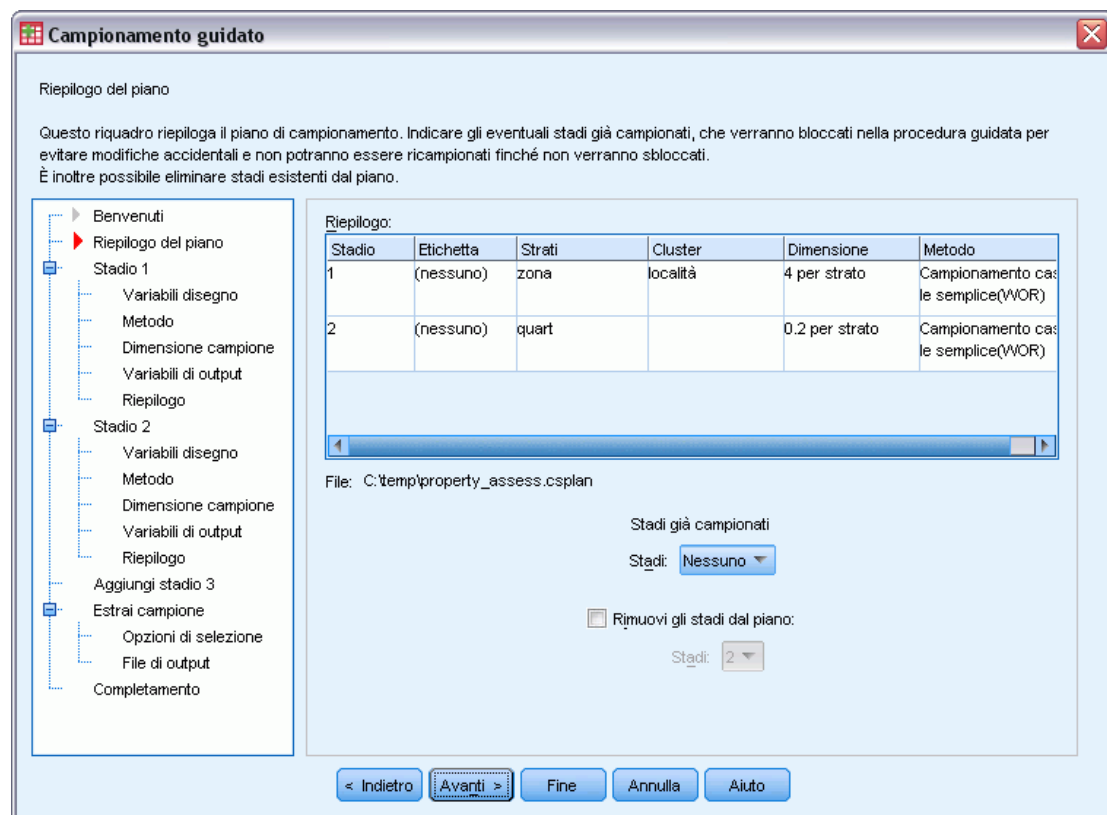
Se lo si desidera, è possibile:

- Specificare gli stadi che sono già stati campionati.
- Rimuovere stadi dal piano.

Campionamento guidato: Riepilogo del piano

Figura 2-11

Campionamento guidato - Passo Riepilogo del piano



Questo passo consente di rivedere il piano di campionamento e di specificare gli stadi che sono già stati campionati. Se si modifica un piano, è inoltre possibile rimuovere stadi dal piano.

Stadi già campionati. Se non è disponibile una struttura di campionamento estesa, sarà necessario eseguire un disegno di campionamento a più stadi utilizzando uno stadio alla volta. Selezionare gli stadi già campionati dall’elenco a discesa. Gli stadi eseguiti sono bloccati, non sono disponibili nel passo Estrai campione: Opzioni di selezione e non possono essere alterati durante la modifica di un piano.

Rimuovi gli stadi. È possibile rimuovere gli stadi 2 e 3 da un disegno a più stadi.

Per eseguire un piano di campionamento esistente

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Seleziona un campione...
- ▶ Selezionare Estrarre un campione e scegliere un file del piano da eseguire.
- ▶ Per proseguire la procedura guidata, fare clic su Avanti.
- ▶ Rivedere il piano di campionamento nel passo Riepilogo del piano, quindi fare clic su Avanti.
- ▶ Quando si esegue un piano di campionamento, i singoli passi contenenti informazioni sugli stadi vengono ignorati. È ora possibile andare in qualsiasi momento al passo Fine.
Facoltativamente, è possibile specificare gli stadi che sono già stati campionati.

Opzioni aggiuntive dei comandi CSPLAN e CSSELECT

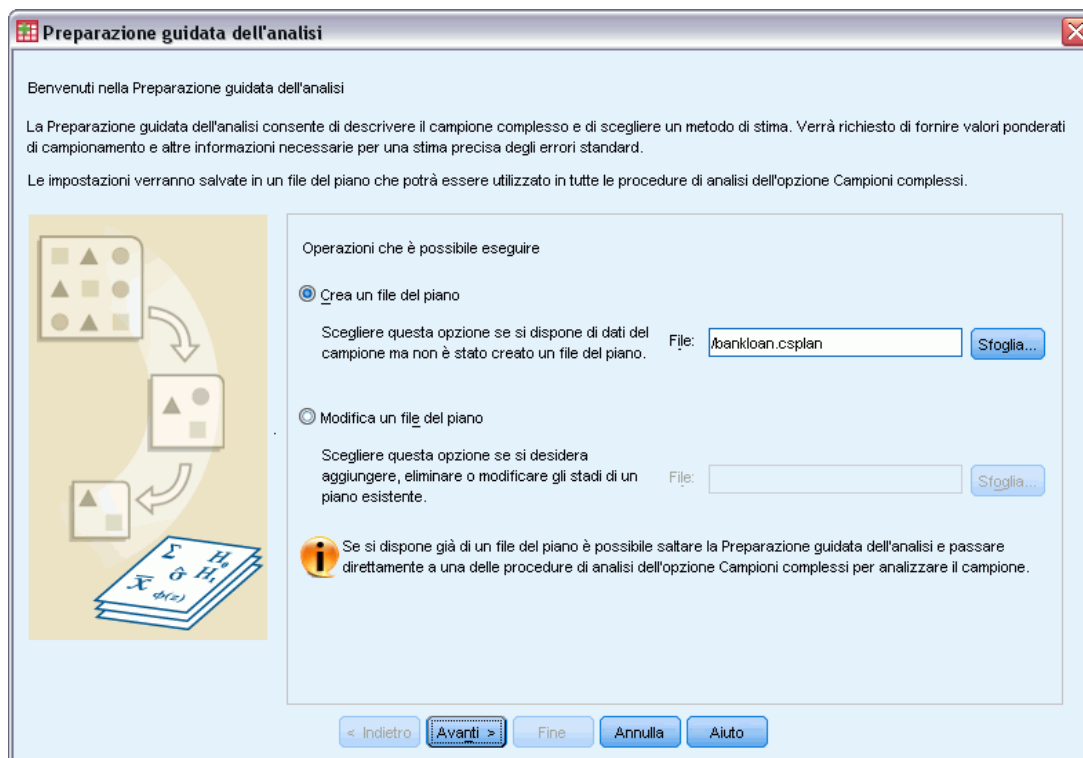
Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare nomi personalizzati per le variabili di output.
- Controllare l'output nel Viewer. Ad esempio, è possibile sopprimere il riepilogo del piano per stadi che viene visualizzato se si crea o modifica un campione, sopprimere il riepilogo della distribuzione dei casi campionati per stadi che viene visualizzato se si esegue il disegno di campionamento e richiedere un riepilogo dei casi.
- Scegliere un sottoinsieme di variabili del file di dati attivo da scrivere in un file del campione esterno o in un file di dati diverso.

Per informazioni dettagliate sulla sintassi, vedere *Command Syntax Reference*.

Preparazione di un campione complesso per l'analisi

Figura 3-1
Preparazione guidata dell'analisi, Finestra di benvenuto



Preparazione guidata dell'analisi consente di eseguire le varie fasi di creazione o modifica di un piano analisi da usare con le diverse procedure di analisi di Campioni complessi. Prima di usare la procedura guidata, è consigliabile estrarre un campione in base a un disegno complesso.

Se non si ha accesso al file del piano di campionamento mediante il quale è stato estratto il campione, è utile creare un nuovo piano. Tenere presente che il piano di campionamento contiene un piano analisi predefinito. Se invece si ha accesso al file del piano di campionamento, è possibile usare il piano analisi predefinito in esso contenuto oppure ignorare le specifiche dell'analisi predefinite e salvare le proprie modifiche in un nuovo file.

Per creare un nuovo piano analisi

- Dai menu, scegliere:
Analizza > Campioni complessi > Prepara per l'analisi...

- ▶ Selezionare Crea un file del piano e scegliere un nome per il file del piano in cui si desidera salvare il piano analisi.
- ▶ Per proseguire la procedura guidata, fare clic su Avanti.
- ▶ Specificare la variabile che include i pesi del campione nel passo Variabili disegno e, se necessario, definire gli strati e i raggruppamenti.
- ▶ Fare quindi clic su Fine per salvare il piano.

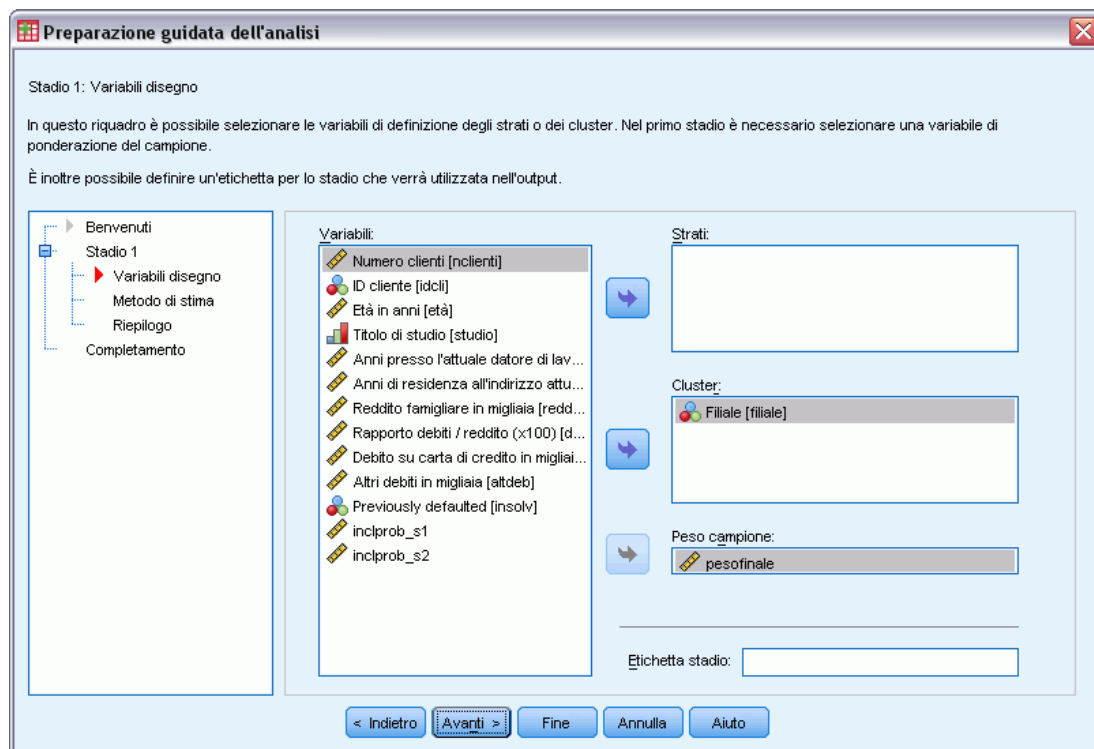
Se necessario, nei passi successivi è possibile:

- Selezionare il metodo di stima degli errori standard nel passo Metodo di stima.
- Specificare il numero di unità campionate oppure la probabilità di inclusione per unità nel passo Dimensione.
- Aggiungere un secondo o un terzo stadio al disegno.
- Incollare le selezioni come sintassi dei comandi.

Preparazione guidata dell'analisi: Variabili disegno

Figura 3-2

Preparazione guidata dell'analisi - Passo Variabili disegno



In questo passo è possibile specificare le variabili di stratificazione e di raggruppamento e i pesi del campione. È inoltre possibile definire un'etichetta per lo stadio.

Strati. La classificazione incrociata delle variabili di stratificazione consente di definire sottopopolazioni distinte, o strati. Il campione totale rappresenta la combinazione di campioni indipendenti da ciascuno strato.

Raggruppamenti. Le variabili di raggruppamento definiscono gruppi di unità di osservazione, o raggruppamenti. I campioni estratti in più stadi selezionano i raggruppamenti negli stadi precedenti e quindi creano sottocampioni delle unità dai raggruppamenti selezionati. Se si analizza un file di dati ottenuto mediante il campionamento di raggruppamenti con sostituzione, è consigliabile includere l'indice di duplicazione come una variabile di raggruppamento.

Peso campione. È necessario specificare i pesi del campione nel primo stadio. I pesi vengono calcolati automaticamente per gli stadi successivi del disegno corrente.

Etichetta stadio. Per ciascuno stadio è possibile specificare un'etichetta facoltativa sotto forma di stringa, che nell'output semplificherà l'identificazione delle informazioni per stadi.

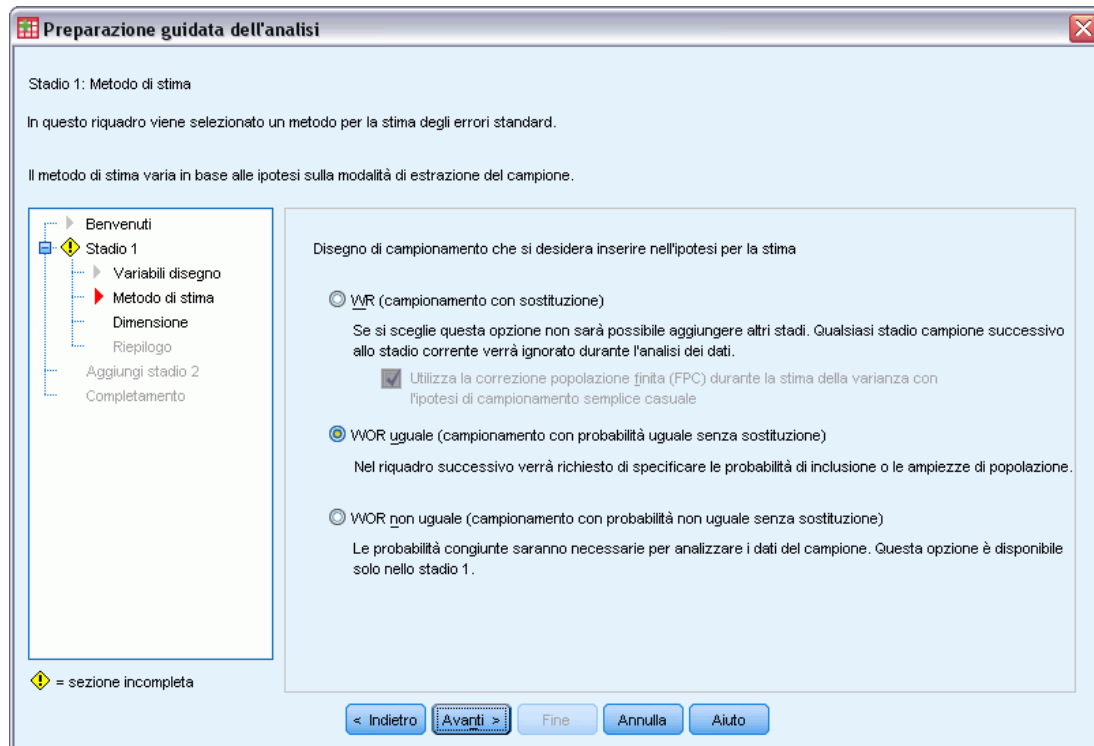
Nota: L'elenco di variabili sorgente è sempre uguale in tutti i passi della procedura guidata. Ciò significa che se in un passo viene rimossa una variabile dall'elenco, tale variabile non sarà disponibile negli stessi elenchi di tutti gli altri passi. Le variabili reinserite nell'elenco sorgente sono disponibili in tutti i passi.

Controlli struttura per gli spostamenti all'interno dell'analisi guidata

Nella parte sinistra di ciascun passo dell'analisi guidata è disponibile una vista riassuntiva di tutti i passi, all'interno della quale è possibile fare clic sul nome di un passo attivo per spostarsi nella procedura guidata. I passi vengono attivati se tutti i passi precedenti sono validi, — ovvero se per ciascun passo precedente sono state fornite le specifiche minime richieste. Per ulteriori informazioni sui motivi per cui un passo può non essere valido, vedere gli argomenti dell'Aiuto relativi ai singoli passi.

Preparazione guidata dell'analisi: Metodo di stima

Figura 3-3
Preparazione guidata dell'analisi - Passo Metodo di stima



In questo passo è possibile specificare un metodo di stima per lo stadio.

WR (campionamento con sostituzione). La stima WR non include la correzione del campionamento da una popolazione finita (FPC) se la stima della varianza viene effettuata utilizzando un disegno di campionamento complesso. Tuttavia, è possibile scegliere di includere o escludere l'FPC quando si stima la varianza utilizzando un campionamento casuale semplice (SRS).

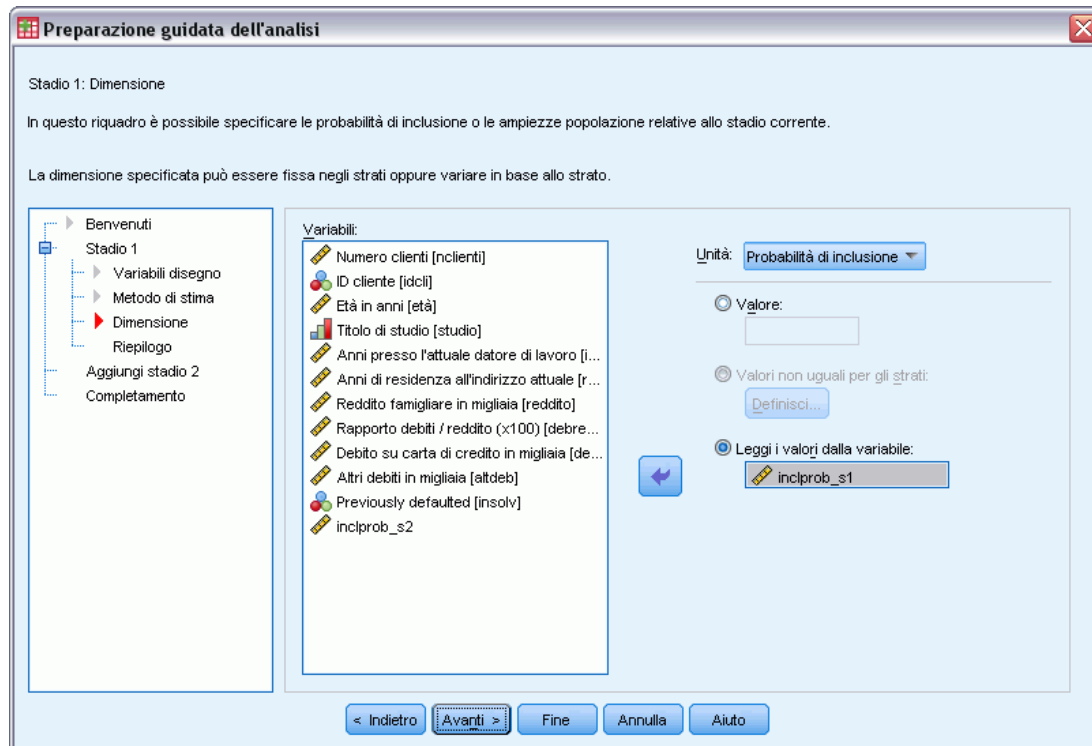
È generalmente preferibile non includere l'FPC per la stima della varianza SRS se i pesi dell'analisi sono stati scalati in modo che non possano essere aggiunti all'ampiezza della popolazione. La stima della varianza SRS viene usata per calcolare statistiche come l'effetto disegno. È possibile specificare la stima WR solo nello stadio finale di un disegno. Se si seleziona la stima WR, la procedura guidata non consentirà di aggiungere un altro stadio.

WOR uguale (campionamento con probabilità uguale senza sostituzione). La stima WOR uguale include la correzione della popolazione finita e assume che le unità siano campionate con probabilità uguale. È possibile specificare la stima WOR uguale in qualsiasi stadio di un disegno.

WOR non uguale (campionamento con probabilità non uguale senza sostituzione). Oltre a usare la correzione della popolazione finita, la stima WOR non uguale determina le unità campionarie (in genere raggruppamenti) selezionate con probabilità non uguale. Questo metodo di stima è disponibile solo nel primo stadio.

Preparazione guidata dell'analisi: Dimensione

Figura 3-4
Preparazione guidata dell'analisi - Passo Dimensione



In questo passo è possibile specificare le probabilità di inclusione o le ampiezze popolazione per lo stadio corrente. Le ampiezze possono essere fisse o variare negli strati. Ai fini della specifica delle ampiezze, è possibile definire gli strati mediante i raggruppamenti specificati nei passi precedenti. È utile notare che questo passaggio è necessario solo se si seleziona WOR uguale come metodo di stima.

Unità. È possibile specificare le ampiezze popolazione esatte oppure le probabilità con cui sono state campionate le unità.

- **Valore.** A tutti gli strati viene applicato un singolo valore. Se come metrica di unità è selezionato Ampiezza popolazione, è necessario inserire un intero non negativo. Se è selezionato Probabilità di inclusione, è necessario inserire un valore compreso fra 0 e 1 inclusi.
- **Valori non uguali per gli strati.** Consente di inserire dimensioni che variano in base allo strato mediante la finestra di dialogo Definisci dimensioni non uguali.
- **Leggi i valori dalla variabile.** Consente di selezionare una variabile numerica che contiene le dimensioni per gli strati.

Definisci dimensioni non uguali

Figura 3-5
Finestra di dialogo Definisci dimensioni non uguali



Nella finestra di dialogo Definisci dimensioni non uguali è possibile inserire le dimensioni di ciascuno strato.

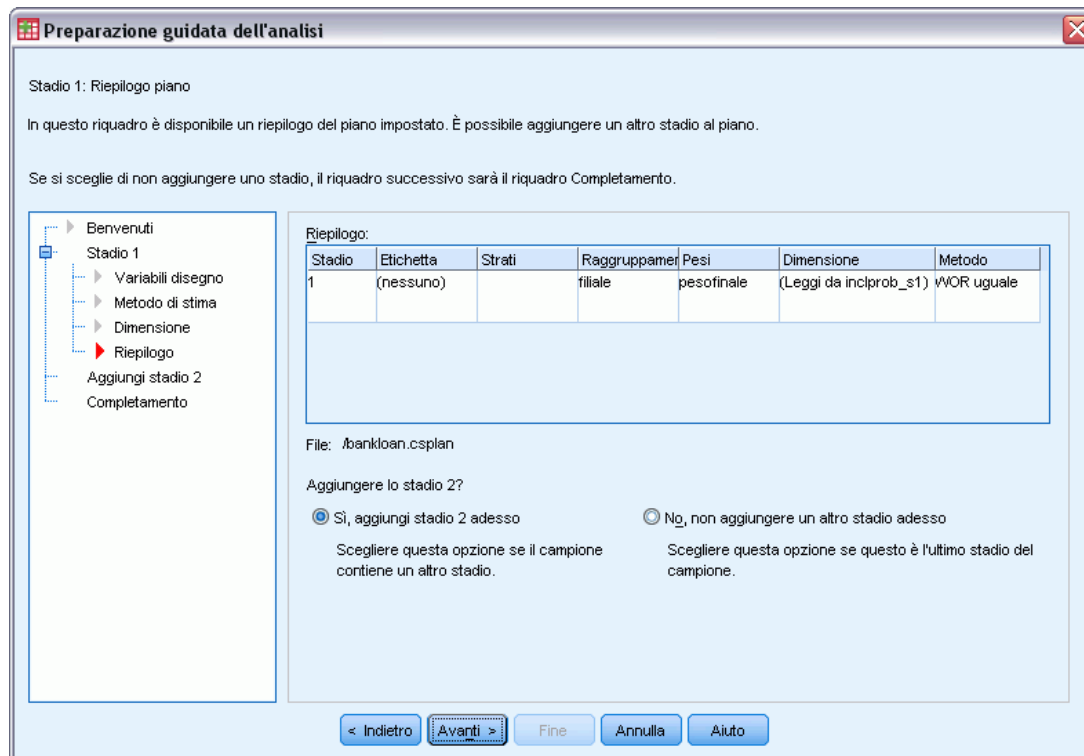
Griglia Specifiche di dimensione. Nella griglia è visualizzata la classificazione incrociata di un massimo di cinque variabili di stratificazione o di raggruppamento,—con una combinazione strato/raggruppamento per ciascuna riga. Le variabili della griglia valide sono tutte le variabili di stratificazione dello stadio corrente e degli stadi precedenti e tutte le variabili di raggruppamento degli stadi precedenti. È possibile riordinare le variabili all'interno della griglia oppure spostarle nell'elenco Escludi. Inserire le dimensioni nella colonna all'estrema destra. Per visualizzare o nascondere le etichette dei valori e i valori delle variabili di stratificazione e di raggruppamento all'interno delle celle della griglia, fare clic su Etichette o Valori. Nelle celle che contengono valori senza etichette vengono sempre visualizzati i valori. Per reinserire nella griglia ciascuna combinazione di valori con etichetta per le variabili rappresentate, fare clic su Ridisegna gli strati.

Escludi. Per specificare le dimensioni di un sottoinsieme di combinazioni di strato/raggruppamento, spostare una o più variabili nell'elenco Escludi. Tali variabili non vengono usate per definire le dimensioni dei campioni.

Preparazione guidata dell'analisi: Riepilogo del piano

Figura 3-6

Preparazione guidata dell'analisi - Passo Riepilogo del piano



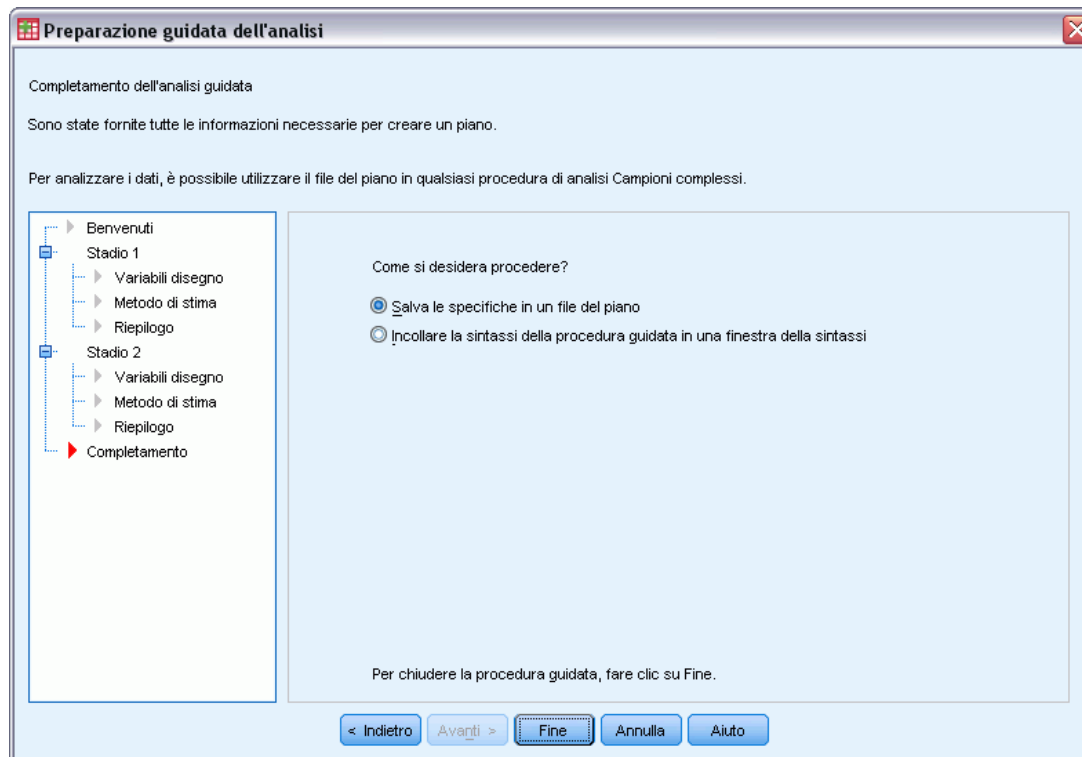
Questo è l'ultimo passo di ciascun stadio e fornisce un riepilogo delle specifiche di disegno dell'analisi fino allo stadio corrente. È quindi possibile procedere allo stadio successivo (creandolo se necessario) oppure salvare le specifiche dell'analisi.

Se non è possibile aggiungere un altro stadio, è probabile che ciò sia dovuto ai motivi seguenti:

- Nel passo Variabili disegno non sono state specificate variabili di raggruppamento.
- Nel passo Metodo di stima è stata selezionata la stima WR.
- Questo è il terzo stadio dell'analisi e la procedura guidata supporta un massimo di tre stadi.

Preparazione guidata dell'analisi: Fine

Figura 3-7
Preparazione guidata dell'analisi - Passo Fine



Passo finale. È possibile salvare ora il file del piano oppure incollare le selezioni in una finestra di sintassi.

Quando si apportano modifiche agli stadi nel file del piano esistente, è possibile salvare il piano modificato in un nuovo piano oppure sovrascrivere il file esistente. Quando si aggiungono stadi senza apportare modifiche a quelle esistenti, la procedura guidata sovrascrive automaticamente il file del piano esistente. Per salvare il piano in un nuovo file, selezionare Incolla la sintassi generata dalla procedura guidata nella finestra della sintassi e modificare il nome del file nei comandi della sintassi.

Per modificare un piano analisi esistente

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Prepara per l'analisi...
- ▶ Selezionare Modifica un file del piano e scegliere un nome per il file del piano in cui si desidera salvare il piano analisi.
- ▶ Per proseguire la procedura guidata, fare clic su Avanti.

- Rivedere il piano analisi nel passo Riepilogo del piano, quindi fare clic su Avanti.

I passi successivi sono molto simili a quelli previsti per un nuovo disegno. Per ulteriori informazioni, vedere i singoli argomenti della Guida.

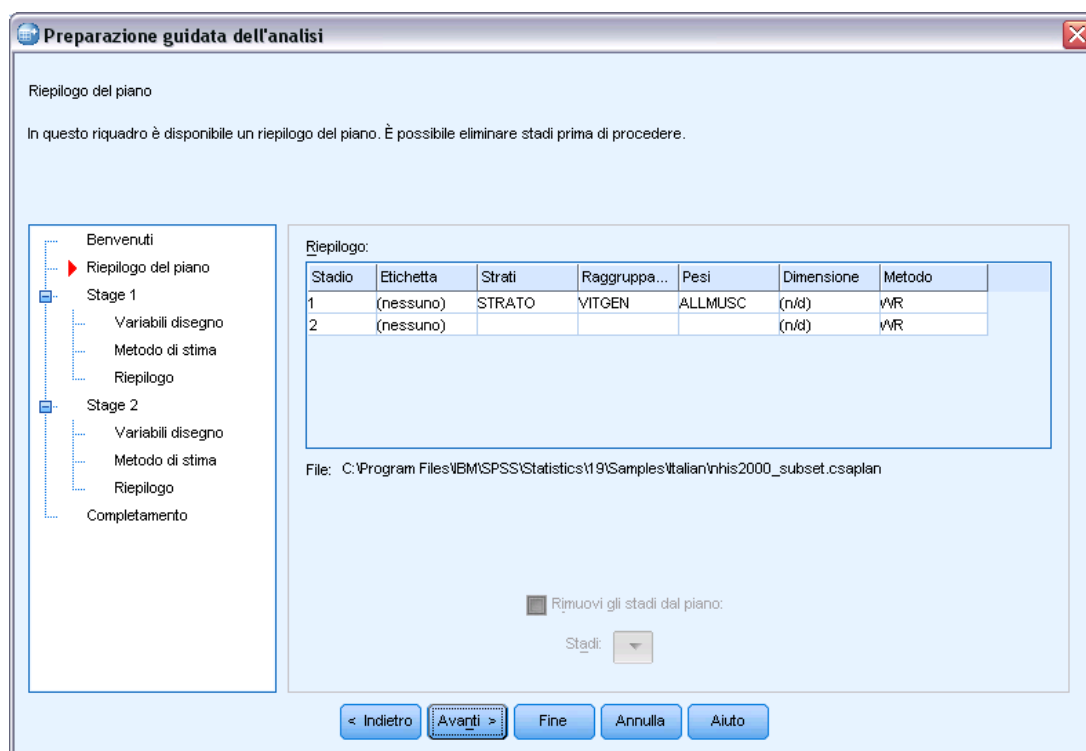
- Andare al passo Fine e specificare un nuovo nome per il file del piano modificato oppure scegliere di sovrascrivere il file del piano esistente.

Facoltativamente, è possibile eliminare gli stadi dal piano.

Preparazione guidata dell'analisi: Riepilogo del piano

Figura 3-8

Preparazione guidata dell'analisi - Passo Riepilogo del piano



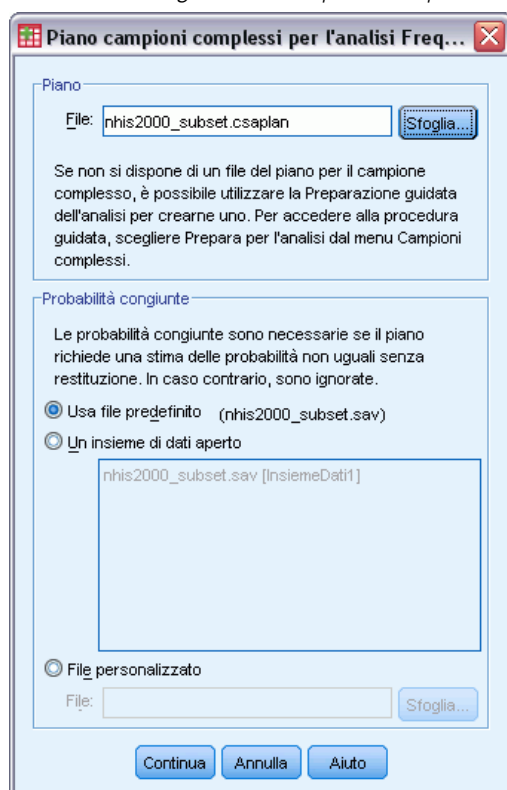
Questo passo consente di rivedere il piano analisi e di rimuovere stadi dal piano.

Rimuovi gli stadi. È possibile rimuovere gli stadi 2 e 3 da un disegno a più stadi. Poiché un piano deve avere almeno uno stadio, lo stadio 1 può essere modificato ma non rimosso dal disegno.

Piano campioni complessi

Per ottenere risultati validi dalle procedure di analisi di Campioni complessi, è necessario che siano disponibili le specifiche dell'analisi di un file del piano analisi o del piano di campionamento.

Figura 4-1
Finestra di dialogo Piano campioni complessi



Piano. Specificare il percorso di un file del piano analisi o del piano di campionamento.

Probabilità congiunte. Per usare la stima senza restituzione (WOR) non uguale per i gruppi estratti con un metodo PPS senza restituzione, è necessario specificare un file distinto o un file di dati attivo contenente le probabilità congiunte. Tale file o file di dati viene creato da Campionamento guidato durante il campionamento.

Frequenze campioni complessi

La procedura Frequenze campioni complessi consente di ottenere tabelle di frequenza per le variabili selezionate e di visualizzare statistiche univariate. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Esempio. Mediante la procedura Frequenze campioni complessi, è possibile ottenere statistiche tabellari univariate sull'uso di vitamine nei cittadini americani, basate sulle informazioni del National Health Interview Survey (NHIS) e con un piano analisi appropriato per questi dati di uso pubblico.

Statistiche. La procedura consente di ottenere le stime delle ampiezze popolazione di cella e le percentuali di tabella, nonché gli errori standard, gli intervalli di confidenza, i coefficienti di variazione, gli effetti disegno, le radici quadrate degli effetti disegno, i valori cumulati e i conteggi non pesati per ciascuna stima. Per il test di uguaglianza proporzioni di cella vengono inoltre calcolate le statistiche chi-quadrato e del rapporto di verosimiglianza.

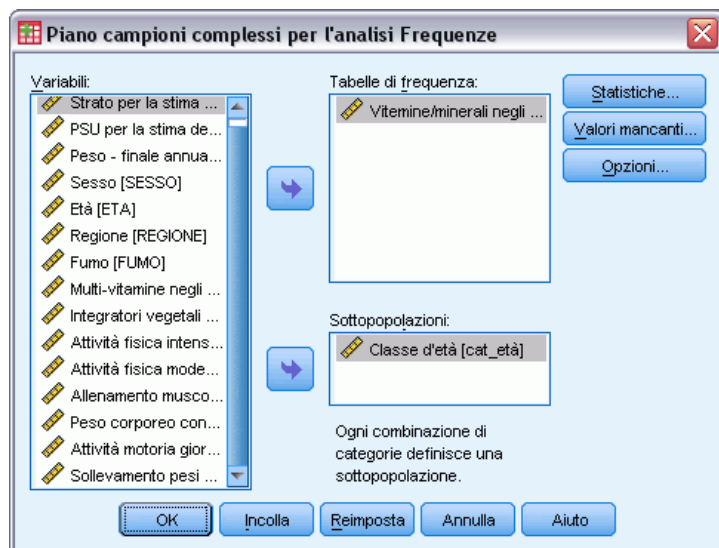
Dati. Le variabili per cui vengono create le tabelle di frequenza devono essere categoriali. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

Per ottenere frequenze campioni complessi

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Frequenze...
- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 5-1
Finestra di dialogo Frequenze

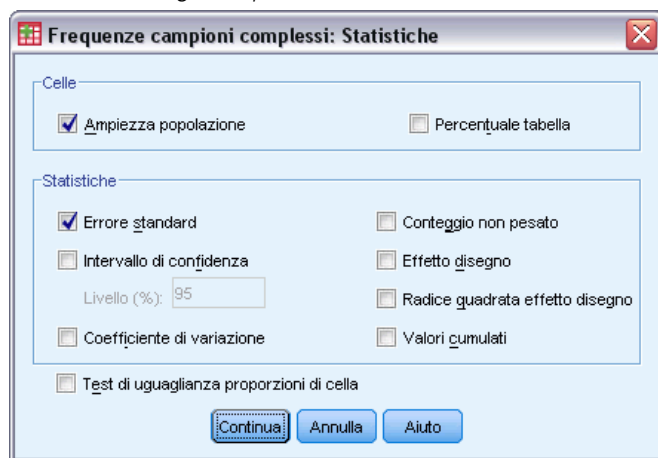


- Selezionare almeno una variabile di frequenza.

Se lo si desidera, è possibile specificare variabili che definiscono sottopopolazioni. Le statistiche vengono calcolate separatamente per ciascuna sottopopolazione.

Frequenze campioni complessi: Statistiche

Figura 5-2
Finestra di dialogo Frequenze: Statistiche



Celle. In questo gruppo è possibile richiedere le stime delle ampiezze popolazione di cella e le percentuali di tabella.

Statistiche. Questo gruppo consente di ottenere statistiche associate all'ampiezza popolazione o alla percentuale di tabella.

- **Errore standard.** Errore standard della stima.

- **Intervallo di confidenza.** L'intervallo di confidenza per la stima, basato sul livello specificato.
- **Coefficiente di variazione.** Il rapporto fra l'errore standard della stima e la stima.
- **Conteggio non pesato.** Il numero di unità utilizzate per il calcolo della stima.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.
- **Valori cumulati.** La stima cumulata per ogni valore della variabile.

Test di uguaglianza proporzioni di cella. Consente di ottenere test chi-quadrato e del rapporto di verosimiglianza relativi all'ipotesi che le categorie di una variabile abbiano frequenze uguali. Per ciascuna variabile vengono eseguiti test distinti.

Campioni complessi: Valori mancanti

Figura 5-3
Finestra di dialogo Valori mancanti

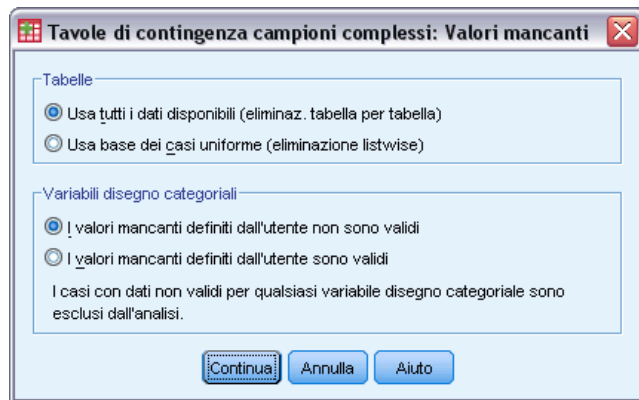


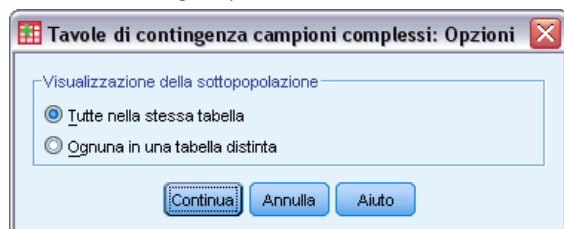
Tabelle. Questo gruppo consente di definire i casi che verranno utilizzati nell'analisi.

- **Usa tutti i dati disponibili.** I valori mancanti vengono determinati in base alle singole tabelle. Quindi, i casi usati per calcolare le statistiche possono variare nella tavole della frequenza o di contingenza.
- **Usa base dei casi uniforme.** I valori mancanti vengono rilevati in tutte le variabili. Quindi, i casi usati per calcolare le statistiche sono uniformi in tutte le tabelle.

Variabili disegno categoriali. Questo gruppo determina se i valori mancanti definiti dall'utente sono validi o non validi.

Campioni complessi: Opzioni

Figura 5-4
Finestra di dialogo Opzioni



Visualizza sottopopolazioni. È possibile scegliere di visualizzare le sottopopolazioni nella stessa tabella o in tabelle distinte.

Descrittive campioni complessi

La procedura Descrittive campioni complessi consente di visualizzare statistiche riassuntive univariate per numerose variabili. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Esempio. Mediante la procedura Descrittive campioni complessi, è possibile ottenere statistiche descrittive univariate relative ai livelli di attività dei cittadini americani, basate sulle informazioni del National Health Interview Survey (NHIS) e con un piano analisi appropriato per questi dati di uso pubblico.

Statistiche. La procedura consente di ottenere le medie e le somme, nonché i test t , gli errori standard, gli intervalli di confidenza, i coefficienti di variazione, i conteggi non pesati, le ampiezze popolazione, gli effetti disegno e le radici quadrate degli effetti disegno per ciascuna stima.

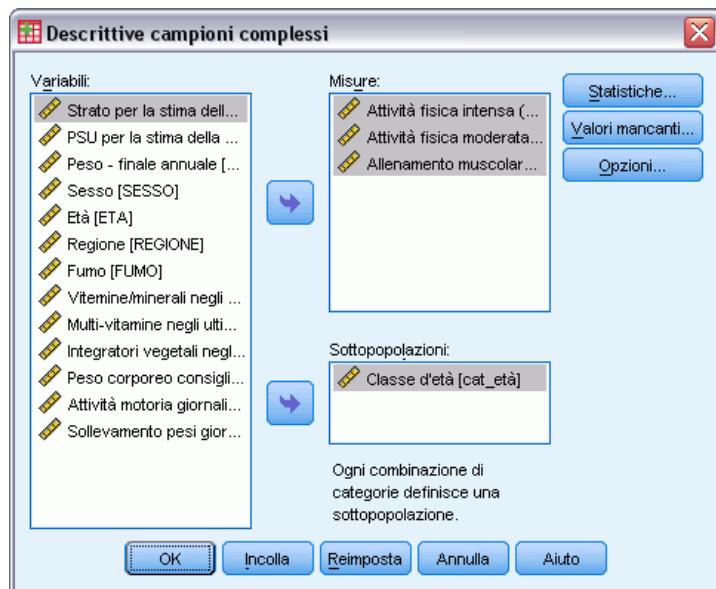
Dati. Le misure devono essere variabili scala. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

Per ottenere analisi descrittive campioni complessi

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Descrittive...
- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 6-1
Descrittive

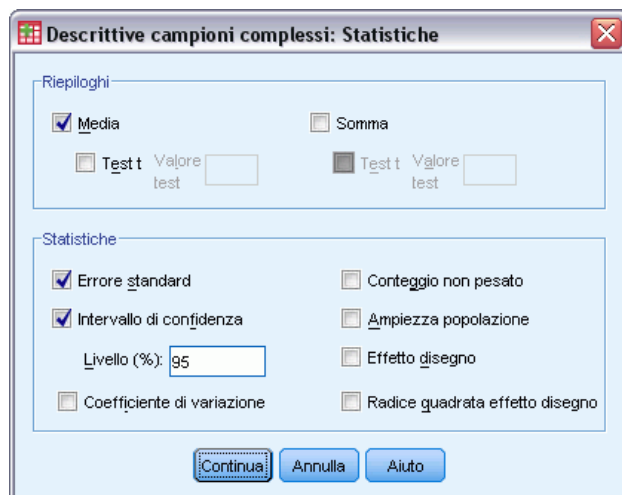


- Selezionare almeno una variabile di misura.

Se lo si desidera, è possibile specificare variabili che definiscono sottopopolazioni. Le statistiche vengono calcolate separatamente per ciascuna sottopopolazione.

Descrittive campioni complessi: Statistiche

Figura 6-2
Finestra di dialogo Descrittive: Statistiche



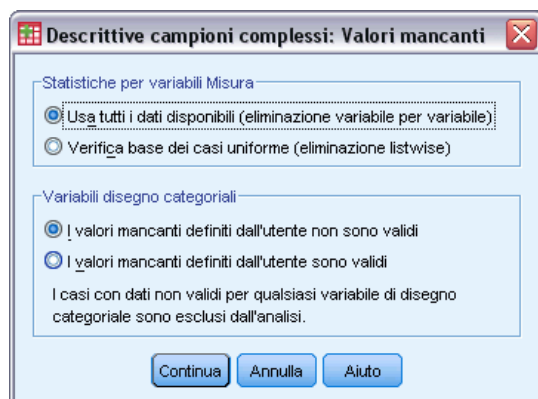
Riepiloghi. Questo gruppo consente di richiedere stime delle medie e somme delle variabili di misura. È inoltre possibile richiedere test t delle stime mediante il confronto con un valore specificato.

Statistiche. Questo gruppo consente di ottenere le statistiche associate alla media o alla somma.

- **Errore standard.** Errore standard della stima.
- **Intervallo di confidenza.** L'intervallo di confidenza per la stima, basato sul livello specificato.
- **Coefficiente di variazione.** Il rapporto fra l'errore standard della stima e la stima.
- **Conteggio non pesato.** Il numero di unità utilizzate per il calcolo della stima.
- **Ampiezza popolazione.** Il numero stimato di unità nella popolazione.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.

Descrittive campioni complessi: Valori mancanti

Figura 6-3
Finestra di dialogo Descrittive: Valori mancanti



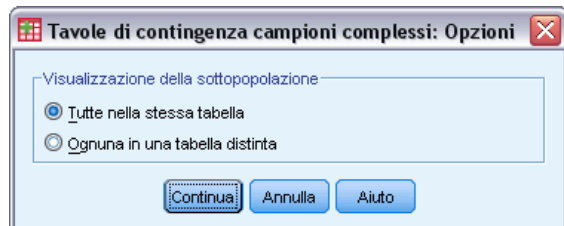
Statistiche per variabili di misura. Questo gruppo consente di definire i casi che verranno utilizzati nell'analisi.

- **Usa tutti i dati disponibili.** I valori mancanti sono determinati in base alla variabile, pertanto i casi usati per il calcolo delle statistiche possono variare nelle variabili di misura.
- **Usa base dei casi uniforme.** I valori mancanti sono determinati in tutte le variabili, pertanto i casi usati per il calcolo delle statistiche sono uniformi.

Variabili disegno categoriali. Questo gruppo determina se i valori mancanti definiti dall'utente sono validi o non validi.

Campioni complessi: Opzioni

Figura 6-4
Finestra di dialogo Opzioni



Visualizza sottopopolazioni. È possibile scegliere di visualizzare le sottopopolazioni nella stessa tabella o in tabelle distinte.

Tavole di contingenza campioni complessi

La procedura Tavole di contingenza campioni complessi consente di creare tavole di contingenza per le coppie di variabili selezionate e di visualizzare statistiche a due vie. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Esempio. Mediante la procedura Tavole di contingenza campioni complessi, è possibile ottenere statistiche di classificazione incrociata relative alla frequenza del fumo di sigaretta e all'uso di vitamine nei cittadini americani, basate sulle informazioni del National Health Interview Survey (NHIS) e con un piano analisi appropriato per questi dati di uso pubblico.

Statistiche. La procedura consente di ottenere le stime delle ampiezze popolazione di cella e le percentuali di riga, colonna e tabella, nonché gli errori standard, gli intervalli di confidenza, i coefficienti di variazione, i valori attesi, gli effetti disegno, le radici quadrate degli effetti disegno, i residui, i residui corretti e i conteggi non pesati per ciascuna stima. Per le tabelle 2x2 vengono calcolati il rapporto Odd, il rischio relativo e la differenza di rischio. Per il test di indipendenza per le variabili di riga e di colonna vengono inoltre calcolate le statistiche di Pearson e del rapporto di verosimiglianza.

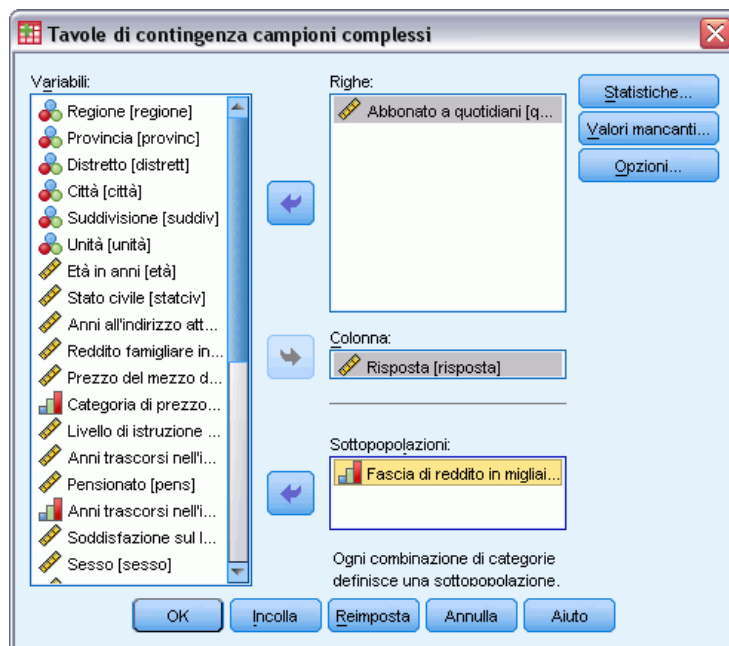
Dati. Le variabili di riga e di colonna devono essere categoriali. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

Per ottenere tavole di contingenza di campioni complessi

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Tavole di contingenza...
- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 7-1
Finestra di dialogo Tavole di contingenza

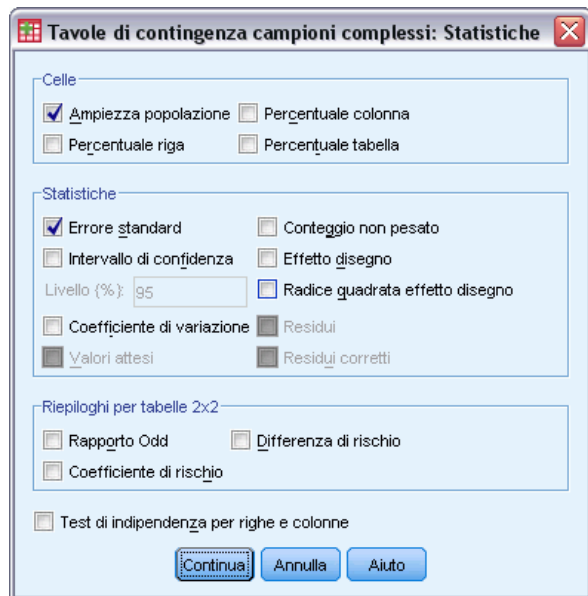


- Selezionare almeno una variabile di riga e una variabile di colonna.

Se lo si desidera, è possibile specificare variabili che definiscono sottopopolazioni. Le statistiche vengono calcolate separatamente per ciascuna sottopopolazione.

Tavole di contingenza campioni complessi: Statistiche

Figura 7-2
Finestra di dialogo Tavole di contingenza: Statistiche



Celle. In questo gruppo è possibile richiedere le stime delle ampiezze popolazione di cella e le percentuali di riga, colonna e tabella.

Statistiche. Questo gruppo consente di creare le statistiche associate all'ampiezza popolazione e alle percentuali di riga, colonna e tabella.

- **Errore standard.** Errore standard della stima.
- **Intervallo di confidenza.** L'intervallo di confidenza per la stima, basato sul livello specificato.
- **Coefficiente di variazione.** Il rapporto fra l'errore standard della stima e la stima.
- **Valori attesi.** Il valore atteso della stima, in base all'ipotesi di indipendenza della variabile di riga e di colonna.
- **Conteggio non pesato.** Il numero di unità utilizzate per il calcolo della stima.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.
- **Residui.** Il valore atteso è il numero di casi attesi nella cella nel caso in cui non esista alcuna relazione fra le due variabili. Un residuo positivo indica che ci sono più casi nella cella di quanti ce ne sarebbero se le variabili di riga e di colonna fossero indipendenti.
- **Residui corretti.** Il residuo di una cella (valore osservato meno valore atteso) diviso per una stima del relativo errore standard. Viene espresso in unità di deviazione standard sopra o sotto la media.

Riepiloghi per tabelle 2x2. Questo gruppo consente di creare statistiche per tabelle in cui sono presenti due categorie per la variabile di riga e di colonna. Ognuna misura l'intensità dell'associazione fra la presenza di un fattore e la manifestazione di un evento.

- **Rapporto Odd.** Il rapporto Odd può essere usato come stima del rischio relativo quando la presenza del fattore è rara.
- **Rischio relativo.** Il rapporto fra il rischio di un evento in presenza del fattore e il rischio in assenza del fattore.
- **Differenza di rischio.** La differenza fra il rischio di un evento in presenza del fattore e il rischio in assenza del fattore.

Test di indipendenza per righe e colonne. Consente di ottenere test chi-quadrato e del rapporto di verosimiglianza relativi all'ipotesi di indipendenza per una variabile di riga e una variabile di colonna. Per ciascuna coppia di variabili vengono eseguiti test distinti.

Campioni complessi: Valori mancanti

Figura 7-3
Finestra di dialogo Valori mancanti

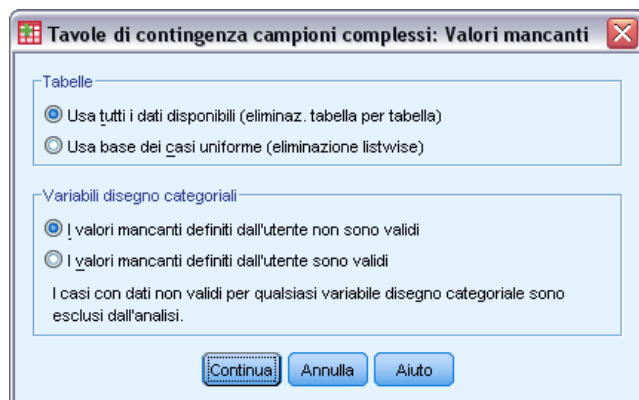


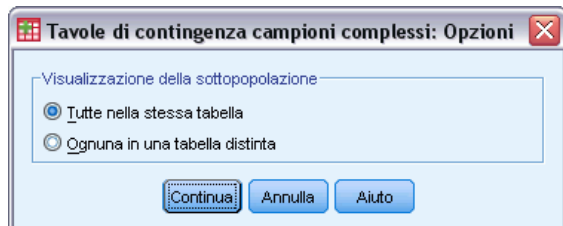
Tabelle. Questo gruppo consente di definire i casi che verranno utilizzati nell'analisi.

- **Usa tutti i dati disponibili.** I valori mancanti vengono determinati in base alle singole tabelle. Quindi, i casi usati per calcolare le statistiche possono variare nella tavole della frequenza o di contingenza.
- **Usa base dei casi uniforme.** I valori mancanti vengono rilevati in tutte le variabili. Quindi, i casi usati per calcolare le statistiche sono uniformi in tutte le tabelle.

Variabili disegno categoriali. Questo gruppo determina se i valori mancanti definiti dall'utente sono validi o non validi.

Campioni complessi: Opzioni

Figura 7-4
Finestra di dialogo Opzioni



Visualizza sottopopolazioni. È possibile scegliere di visualizzare le sottopopolazioni nella stessa tabella o in tabelle distinte.

Rapporti campioni complessi

La procedura Rapporti campioni complessi consente di visualizzare statistiche riassuntive univariate per rapporti di variabili. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Esempio. Mediante la procedura Rapporti campioni complessi è possibile ottenere statistiche univariate per il rapporto fra il valore corrente della proprietà e l'ultima valutazione del valore, in base ai risultati di uno studio condotto a livello regionale e in base a un disegno complesso che prevede un piano analisi appropriato per i dati.

Statistiche. La procedura consente di ottenere le stime del rapporto, nonché i test t , gli errori standard, gli intervalli di confidenza, i coefficienti di variazione, i conteggi non pesati, le ampiezze popolazione, gli effetti disegno e le radici quadrate degli effetti disegno.

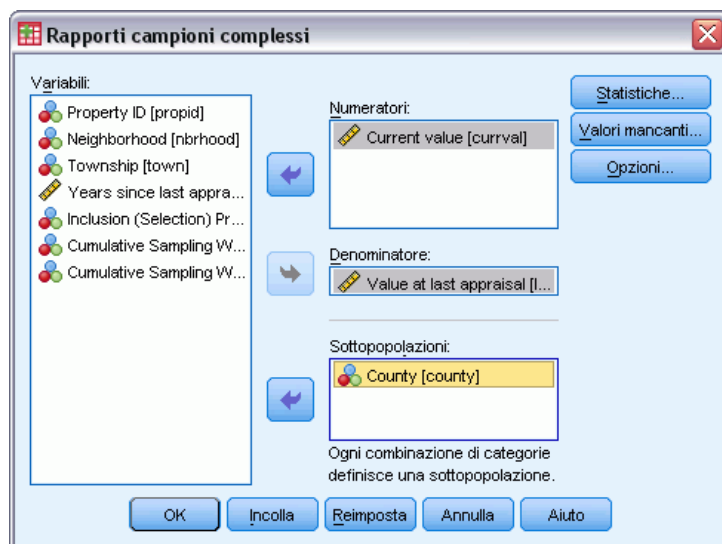
Dati. I numeratori e i denominatori devono essere variabili di scala con valori positivi. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

Per ottenere rapporti campioni complessi

- ▶ Dai menu, scegliere:
Analizza > Campioni complessi > Rapporti...
- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 8-1
Finestra di dialogo Rapporti

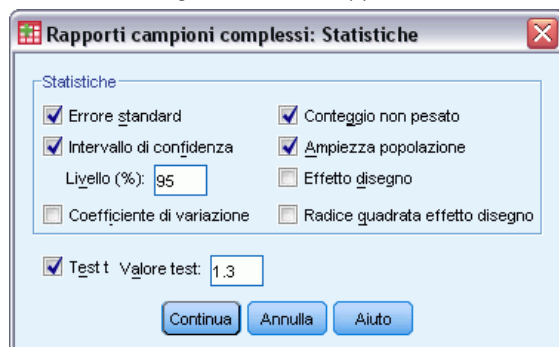


- Selezionare almeno una variabile numeratore e una variabile denominatore.

Opzionalmente, è possibile specificare variabili per definire i sottogruppi per cui vengono prodotte le statistiche.

Rapporti campioni complessi: Statistiche

Figura 8-2
Finestra di dialogo Statistiche rapporti



Statistiche. Questo gruppo consente di ottenere statistiche associate alla stima del rapporto.

- **Errore standard.** Errore standard della stima.
- **Intervallo di confidenza.** L'intervallo di confidenza per la stima, basato sul livello specificato.
- **Coefficiente di variazione.** Il rapporto fra l'errore standard della stima e la stima.
- **Conteggio non pesato.** Il numero di unità utilizzate per il calcolo della stima.
- **Ampiezza popolazione.** Il numero stimato di unità nella popolazione.

- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.

Test T. È possibile richiedere test t delle stime mediante il confronto con un valore specificato.

Rapporti campioni complessi: Valori mancanti

Figura 8-3
Finestra di dialogo Rapporti valori mancanti



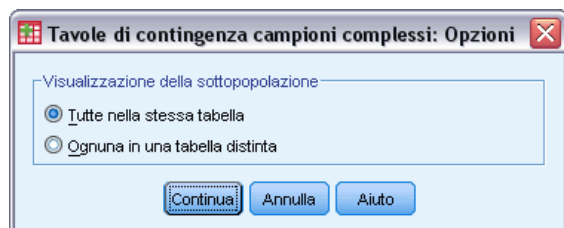
Rapporti. Questo gruppo consente di definire i casi che verranno utilizzati nell'analisi.

- **Usa tutti i dati disponibili.** I valori mancanti vengono determinati in base ai singoli rapporti. Quindi, i casi usati per calcolare le statistiche possono variare nelle copie numeratore-denominatore.
- **Usa base dei casi uniforme.** I valori mancanti vengono rilevati in tutte le variabili. Quindi, i casi usati per calcolare le statistiche sono uniformi.

Variabili disegno categoriali. Questo gruppo determina se i valori mancanti definiti dall'utente sono validi o non validi.

Campioni complessi: Opzioni

Figura 8-4
Finestra di dialogo Opzioni



Visualizza sottopopolazioni. È possibile scegliere di visualizzare le sottopopolazioni nella stessa tabella o in tabelle distinte.

Modello lineare generalizzato campioni complessi

La procedura Modello lineare generalizzato campioni complessi (CSGLM) esegue l'analisi di regressione lineare, nonché l'analisi della varianza e della covarianza su campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Esempio. Una catena di drogherie ha effettuato un'indagine su un gruppo di clienti in relazione alle loro abitudini di acquisto, in base a un disegno complesso. Dati i risultati dell'indagine e la spesa di ogni cliente nel mese precedente, si vuole verificare se la frequenza di acquisto della clientela è correlata all'importo di spesa mensile, controllando il sesso del cliente e inserendo il disegno di campionamento.

Statistiche. La procedura consente di ottenere stime, errori standard, intervalli di confidenza, test t , effetti disegno e radici quadrate degli effetti disegno per i parametri del modello, nonché correlazioni e covarianze tra le stime dei parametri. Sono inoltre disponibili le misure dell'adattamento del modello e le statistiche descrittive per le variabili dipendenti e indipendenti. In aggiunta è possibile richiedere le medie marginali stimate per livelli di fattori di modelli e interazioni tra fattori.

Dati. La variabile dipendente è quantitativa. I fattori sono categoriali. Le covariate sono variabili quantitative correlate alla variabile dipendente. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

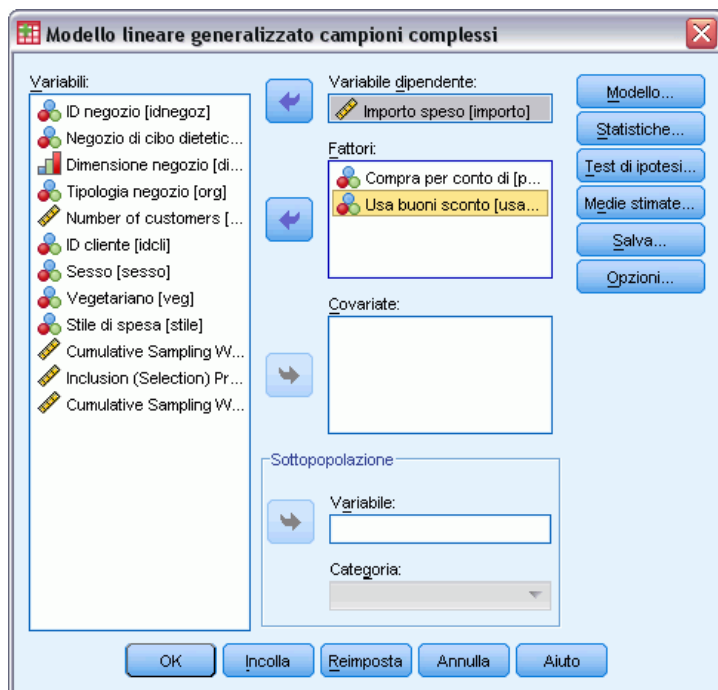
Per ottenere un modello lineare generalizzato campioni complessi

Dai menu, scegliere:

Analizza > Complex Samples > Modello lineare generalizzato...

- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 9-1
Finestra di dialogo Modello lineare generalizzato

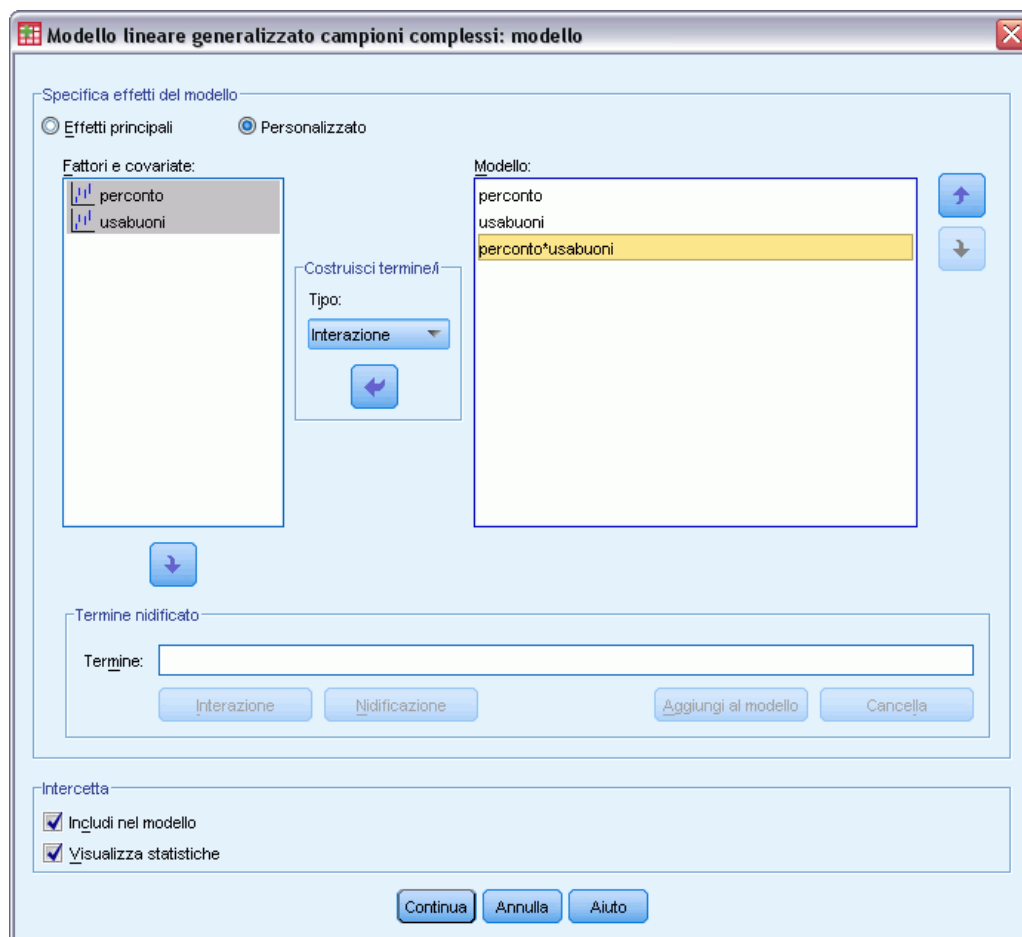


- Selezionare una variabile dipendente.

Se lo si desidera, è possibile:

- Selezionare le variabili per l'opzione Fattori e Covariate, a seconda dei dati in uso.
- Specificare una variabile che definisce una sottopopolazione. L'analisi viene eseguita solo per la categoria selezionata della variabile della sottopopolazione.

Figura 9-2
Finestra di dialogo Modello



Specifica modello effetti. Per impostazione predefinita, la procedura crea un modello di effetti principali usando i fattori e le covariate specificati nella finestra di dialogo principale. È possibile anche creare un modello personalizzato che comprenda effetti di interazione e termini nidificati.

Termini non nidificati

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Termini nidificati

Questa procedura consente di costruire termini nidificati per il modello. I termini nidificati sono utili per modellare l'effetto di un fattore o di una covariata i cui valori non interagiscono con i livelli di un altro fattore. Ad esempio, una catena di drogherie può seguire le abitudini di spesa dei propri clienti in più sedi di negozi. Poiché ogni cliente frequenta solo una delle sedi, si può affermare che l'effetto *Cliente* è **nidificato all'interno** dell'effetto *Ubicazione negozio*.

È possibile anche includere gli effetti di interazione, come i termini polinomiali che interessano la stessa covariata, oppure aggiungere più livelli di nidificazione al termine nidificato.

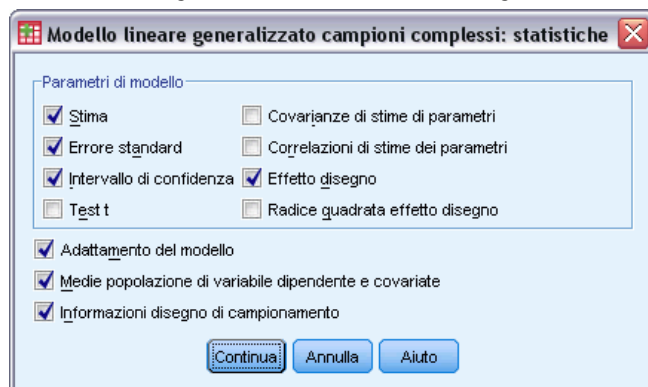
Limitazioni. I termini nidificati sono sottoposti alle seguenti restrizioni:

- Tutti i fattori compresi in un'interazione devono essere univoci. Quindi, se A è un fattore, non è consentito specificare $A*A$.
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Pertanto, se A è un fattore, non è consentito specificare $A(A)$.
- Nessun effetto può essere nidificato all'interno di una covariata. Quindi, se A è un fattore e X è una covariata, non è possibile specificare $A(X)$.

Intercetta. L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercettazione può essere esclusa. Anche se si include l'intercettazione nel modello, è possibile scegliere di disattivare la visualizzazione delle statistiche associate al modello.

Statistiche modello lineare generalizzato campioni complessi

Figura 9-3
Finestra di dialogo Statistiche modello lineare generalizzato



Parametri modello. Il gruppo consente di controllare la visualizzazione delle statistiche relative ai parametri del modello.

- **Stima.** Visualizza le stime dei coefficienti.
- **Errore standard.** Visualizza l'errore standard per ogni stima dei coefficienti.

- **Intervallo di confidenza.** Visualizza un intervallo di confidenza per ogni stima dei coefficienti. Il livello di confidenza per l'intervallo è impostato nella finestra di dialogo Opzioni.
- **Test T.** Visualizza un test t per ogni stima dei coefficienti. L'ipotesi nulla per ogni test corrisponde al caso in cui il valore del coefficiente è 0.
- **Covarianze di stime di parametri** Visualizza una stima della matrice di covarianza per i coefficienti del modello.
- **Correlazioni di stime di parametri** Visualizza una stima della matrice di correlazione per i coefficienti del modello.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.

Adattamento del modello. Visualizza le statistiche relative all' R quadrato e alla radice quadrata dell'errore medio.

Medie della popolazione delle variabili dipendenti e delle covariate. Visualizza informazioni di riepilogo sulla variabile dipendente, sulle covariate e sui fattori.

Informazioni sul disegno di campionamento. Visualizza informazioni di riepilogo sul campione, compreso il conteggio non pesato e le dimensioni della popolazione.

Modello lineare generalizzato campioni complessi: test di ipotesi

Figura 9-4
Finestra di dialogo Test di ipotesi

Regressione logistica campioni complessi: test di ipotesi

Statistica test

- F
- F corretto
- Chi-quadrato
- Chi-quadrato corretto

Gradi di libertà campionamento

- In base a disegno di campionamento
- Fissa Valore:

Correzione per confronti multipli

- Differenza meno significativa
- Sidak sequenziale
- Bonferroni sequenziale
- Sidak
- Bonferroni

Continua Annulla Aiuto

Statistica test. Questo gruppo permette di selezionare il tipo di statistica da usare per i test delle ipotesi. È possibile scegliere tra F , F corretta, Chi-quadrato e Chi-quadrato corretto.

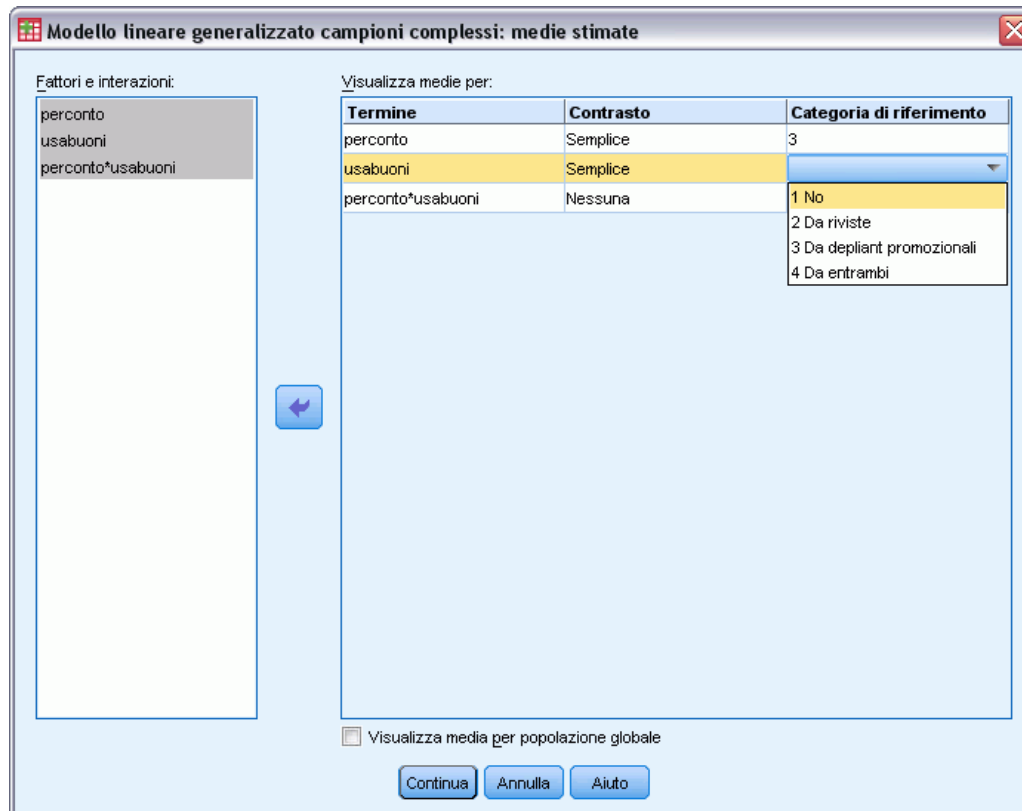
Campionamento dei gradi di libertà. Questo gruppo permette di controllare i gradi di libertà del disegno di campionamento utilizzati per calcolare i valori p di tutte le statistiche di test. Se si utilizza il disegno di campionamento, il valore è pari alla differenza tra il numero delle unità di campionamento primario e il numero di strati della prima fase di campionamento. In alternativa, è possibile impostare un valore personalizzato per i gradi di libertà specificando un intero positivo.

Correzione per confronti multipli. Quando si eseguono test di ipotesi con contrasti multipli, è possibile adattare il livello di significatività globale ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di adattamento.

- **Differenza meno significativa.** Questo metodo non permette di controllare la probabilità complessiva di rifiutare le ipotesi che alcuni contrasti lineari siano diversi dai valori delle ipotesi nulle.
- **Sidak sequenziale.** Una procedura di Sidak con scarti sequenzialmente discendenti, molto meno conservativa per quanto riguarda il rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Bonferroni sequenziale.** Una procedura di Bonferroni con scarti sequenzialmente decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Sidak.** Questo metodo fornisce limiti più stretti rispetto all'approccio di Bonferroni.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.

Medie stimate modello lineare generalizzato campioni complessi

Figura 9-5
Finestra di dialogo Medie stimate modello lineare generalizzato



La finestra di dialogo Medie stimate consente di visualizzare le medie marginali stimate del modello per livelli di fattori di modelli e interazioni tra fattori specificate nella finestra di dialogo secondaria Modello. È inoltre possibile impostare la visualizzazione della media della popolazione globale.

Termine. Le medie stimate vengono calcolate per i fattori e le interazioni tra fattori selezionate.

Contrasto. Il contrasto determina in che modo i test di ipotesi sono impostati per il confronto delle medie stimate.

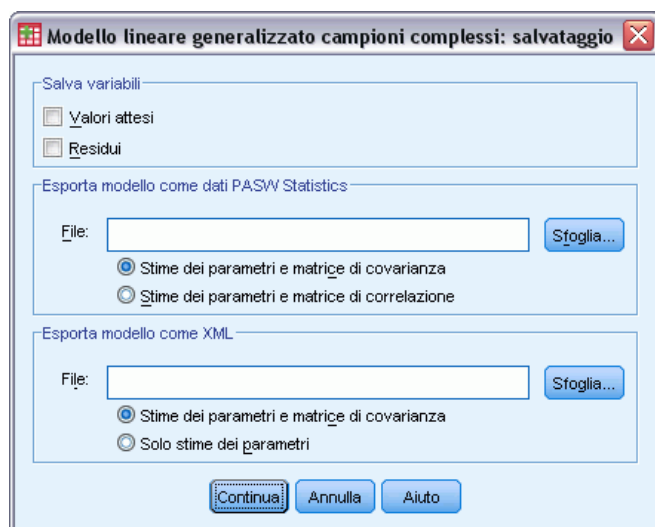
- **Semplice.** Consente di confrontare la media di ciascun livello con la media di un livello specifico. Questo tipo di contrasto risulta utile quando è disponibile un gruppo di controllo.
- **Deviazione.** Consente di confrontare la media di ciascun livello, ad eccezione di una categoria di riferimento, con la media di tutti i livelli (media principale). L'ordine dei livelli del fattore può essere un ordine qualsiasi.
- **Differenza.** Consente di confrontare la media di ciascun livello (ad eccezione del primo) con la media dei livelli precedenti. A volte definiti contrasti di Helmert.
- **Helmert.** Confronta la media di ciascun livello del fattore (eccetto l'ultimo) con la media dei livelli successivi.

- **Ripetuta.** Consente di confrontare la media di ciascun livello (a eccezione dell'ultimo) con la media del livello successivo..
- **Polinomiale.** Consente di confrontare l'effetto lineare, quadratico, cubico e così via. Tutte le categorie del primo grado di libertà includono l'effetto lineare, quelle del secondo includono l'effetto quadratico e così via. Questi contrasti sono spesso usati per la stima di trend polinomiali.

Categoria di riferimento. Per i contrasti deviazione e i contrasti semplici è necessaria una categoria di riferimento o un livello di fattore rispetto al quale confrontare gli altri.

Modello lineare generalizzato campioni complessi - Salva

Figura 9-6
Finestra di dialogo Salva modello lineare generalizzato



Salva variabili. Il gruppo consente di salvare i valori attesi e i residui per il modello come nuove variabili nel file di lavoro.

Esporta modello come dati SPSS Statistics. Genera un insieme di dati in formato IBM® SPSS® Statistics che contiene la correlazione dei parametri o la matrice di covarianza utilizzando le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file della matrice è come segue.

- **rowtype_.** Assume valori (ed etichette dei valori) COV (Covarianze), CORR (Correlazioni), EST (Stime dei parametri), SE (Errori standard), SIG (Livelli di significatività) e DF (Gradi di libertà del disegno di campionamento). Per ogni modello di parametro vi è un caso separato con tipo di riga COV (o CORR) e anche un caso separato per ciascuno degli altri tipi di riga.

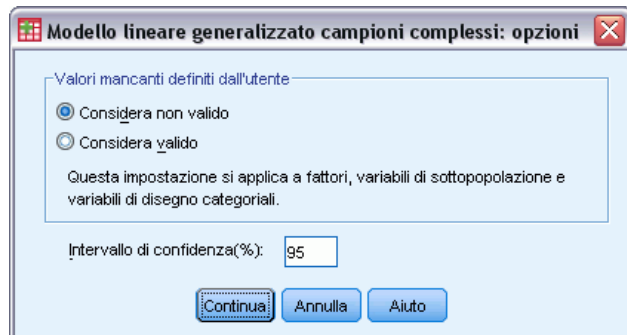
- **varname_.** Assume i valori P1, P2, ..., che corrispondono a un elenco ordinato di tutti i parametri dei modelli per i tipi di riga COV o CORR, con etichette dei valori corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri. Le celle sono vuote per gli altri tipi di riga.
- **P1, P2, ...** Queste variabili corrispondono a un elenco ordinato di tutti i parametri dei modelli con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri e assumono i valori in base al tipo di riga. Per i parametri ridondanti, tutte le covarianze sono impostate su zero e le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate su zero e tutti gli errori standard, i livelli di significatività e i gradi di libertà dei residui sono impostati sul valore mancante di sistema.

Nota: questo file non può essere utilizzato immediatamente per ulteriori analisi in altre procedure che eseguono la lettura di un file di matrice a meno che queste procedure non accettino tutti i tipi di riga esportati qui.

Esporta modello come XML. Consente di salvare le stime dei parametri e la matrice di covarianza dei parametri, se selezionata, in formato XML (PMML). È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Opzioni modello lineare generalizzato campioni complessi

Figura 9-7
Finestra di dialogo Opzioni modello lineare generalizzato



Valori mancanti definiti dall'utente. Tutte le variabili di disegno, nonché la variabile dipendente ed eventuali covariate, devono avere dati validi. I casi con dati non validi per qualsiasi tra queste variabili vengono esclusi dall'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente vengono considerati come validi nelle variabili di stratificazione, di raggruppamento, di sottopopolazione e di fattore.

Intervallo di confidenza. Il livello dell'intervallo di confidenza per le stime dei coefficienti e le medie marginali stimate. Specificare un valore maggiore o uguale a 50 e minore di 100.

Opzioni aggiuntive del comando CSGLM

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test personalizzati di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `CUSTOM`).
- Fissare le covariate sui valori diversi dalle relative medie durante il calcolo delle medie marginali stimate (utilizzando il sottocomando `EMMEANS`).
- Specificare una metrica per contrasti polinomiali (tramite il sottocomando `EMMEANS`).
- Specificare un valore di tolleranza per la verifica della singolarità (tramite il sottocomando `CRITERIA`).
- Creare nomi specificati dall'utente per le variabili salvate (tramite il sottocomando `SAVE`).
- Generare una tabella di forma funzionale generalizzata (tramite il sottocomando `PRINT`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Regressione logistica campioni complessi

La procedura Regressione logistica campioni complessi esegue l'analisi di regressione logistica su una variabile dipendente binaria o multinomiale per campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Esempio. Un funzionario mutui ha raccolto i dati precedenti relativi a clienti cui sono stati accordati prestiti da parte di varie filiali, in base a un disegno complesso. Durante l'inserimento del disegno di campionamento, il funzionario desidera verificare se la probabilità di mancato rimborso da parte di un cliente è connessa all'età, alla storia lavorativa e all'importo dell'esposizione debitoria.

Statistiche. La procedura consente di ottenere stime, stime elevate a potenza, errori standard, intervalli di confidenza, test t , effetti disegno e radici quadrate degli effetti disegno per i parametri del modello, nonché correlazioni e covarianze tra le stime dei parametri. Sono inoltre disponibili statistiche relative all' R^{quadrato} , tabelle di classificazione e statistiche descrittive per le variabili dipendenti e indipendenti.

Dati. La variabile dipendente è categoriale. I fattori sono categoriali. Le covariate sono variabili quantitative correlate alla variabile dipendente. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

Per ottenere una regressione logistica con campioni complessi.

Dai menu, scegliere:

Analizza > Complex Samples > Regressione logistica...

- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 10-1
Finestra di dialogo Regressione logistica



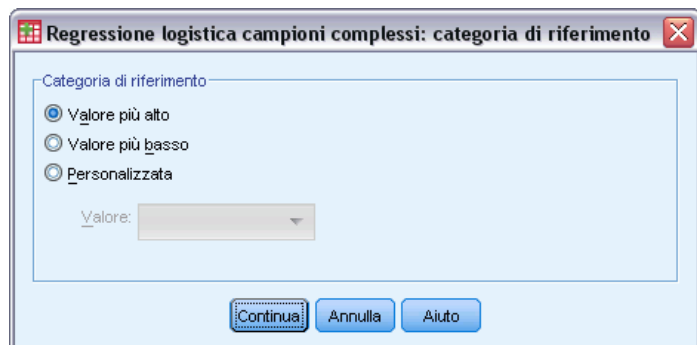
- Selezionare una variabile dipendente.

Se lo si desidera, è possibile:

- Selezionare le variabili per l'opzione Fattori e Covariate, a seconda dei dati in uso.
- Specificare una variabile che definisce una sottopopolazione. L'analisi viene eseguita solo per la categoria selezionata della variabile della sottopopolazione.

Categoria di riferimento regressione logistica campioni complessi

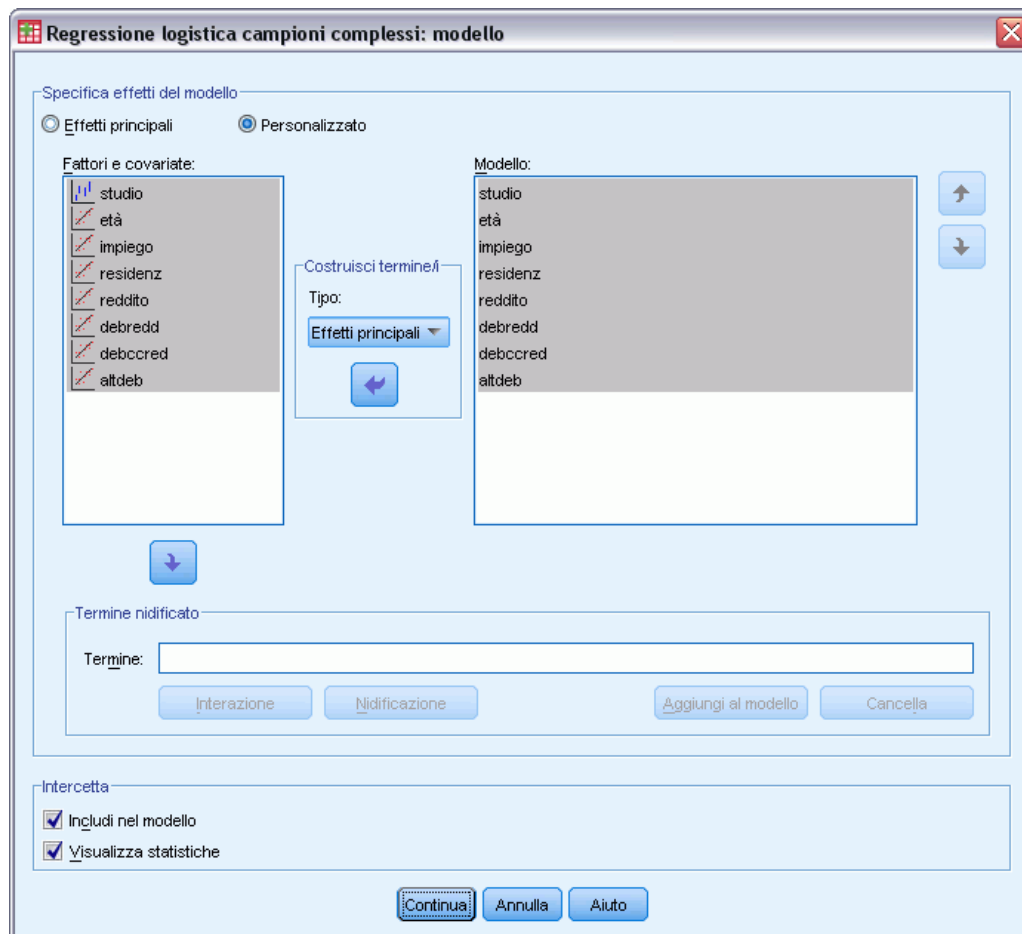
Figura 10-2
Finestra di dialogo Regressione logistica: Categoria riferimento



Per impostazione predefinita, la procedura Regressione logistica campioni complessi definisce la categoria più alta come categoria di riferimento. Questa finestra di dialogo permette di specificare il valore più alto, il valore più basso o una categoria personalizzata come categoria di riferimento.

Modello regressione logistica campioni complessi

Figura 10-3
Finestra di dialogo Modello regressione logistica



Specifica modello effetti. Per impostazione predefinita, la procedura crea un modello di effetti principali usando i fattori e le covariate specificati nella finestra di dialogo principale. È possibile anche creare un modello personalizzato che comprenda effetti di interazione e termini nidificati.

Termini non nidificati

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Termini nidificati

Questa procedura consente di costruire termini nidificati per il modello. I termini nidificati sono utili per modellare l'effetto di un fattore o di una covariata i cui valori non interagiscono con i livelli di un altro fattore. Ad esempio, una catena di drogherie può seguire le abitudini di spesa dei propri clienti in più sedi di negozi. Poiché ogni cliente frequenta solo una delle sedi, si può affermare che l'effetto *Cliente* è **nidificato all'interno** dell'effetto *Ubicazione negozio*.

È possibile anche includere gli effetti di interazione, come i termini polinomiali che interessano la stessa covariata, oppure aggiungere più livelli di nidificazione al termine nidificato.

Limitazioni. I termini nidificati sono sottoposti alle seguenti restrizioni:

- Tutti i fattori compresi in un'interazione devono essere univoci. Quindi, se A è un fattore, non è consentito specificare $A*A$.
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Pertanto, se A è un fattore, non è consentito specificare $A(A)$.
- Nessun effetto può essere nidificato all'interno di una covariata. Quindi, se A è un fattore e X è una covariata, non è possibile specificare $A(X)$.

Intercetta. L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercettazione può essere esclusa. Anche se si include l'intercettazione nel modello, è possibile scegliere di disattivare la visualizzazione delle statistiche associate al modello.

Statistiche regressione logistica campioni complessi

Figura 10-4
Finestra di dialogo Regressione logistica: Statistiche



Adattamento del modello. Controlla la visualizzazione delle statistiche che misurano l'attendibilità globale del modello.

- **Pseudo R-quadrato.** La statistica R^{quadrato} dalla regressione lineare non ha un corrispondente esatto nei modelli di regressione logistica. Sono invece presenti più misure che tentano di simulare le proprietà della statistica R^2 .
- **Tabella classificazioni.** Visualizza le classificazioni incrociate a tabelle della categoria osservata dichiarata per modello sulla variabile dipendente.

Parametri. Il gruppo consente di controllare la visualizzazione delle statistiche relative ai parametri del modello.

- **Stima.** Visualizza le stime dei coefficienti.
- **Stima elevata a potenza.** Visualizza la base del logaritmo naturale elevata alla potenza delle stime dei coefficienti. Mentre la stima ha proprietà adatte alla verifica statistica, la stima elevata a potenza, o $\exp(B)$, è più facile da interpretare.
- **Errore standard.** Visualizza l'errore standard per ogni stima dei coefficienti.
- **Intervallo di confidenza.** Visualizza un intervallo di confidenza per ogni stima dei coefficienti. Il livello di confidenza per l'intervallo è impostato nella finestra di dialogo Opzioni.
- **Test T.** Visualizza un test t per ogni stima dei coefficienti. L'ipotesi nulla per ogni test corrisponde al caso in cui il valore del coefficiente è 0.
- **Covarianze di stime di parametri** Visualizza una stima della matrice di covarianza per i coefficienti del modello.
- **Correlazioni di stime di parametri** Visualizza una stima della matrice di correlazione per i coefficienti del modello.

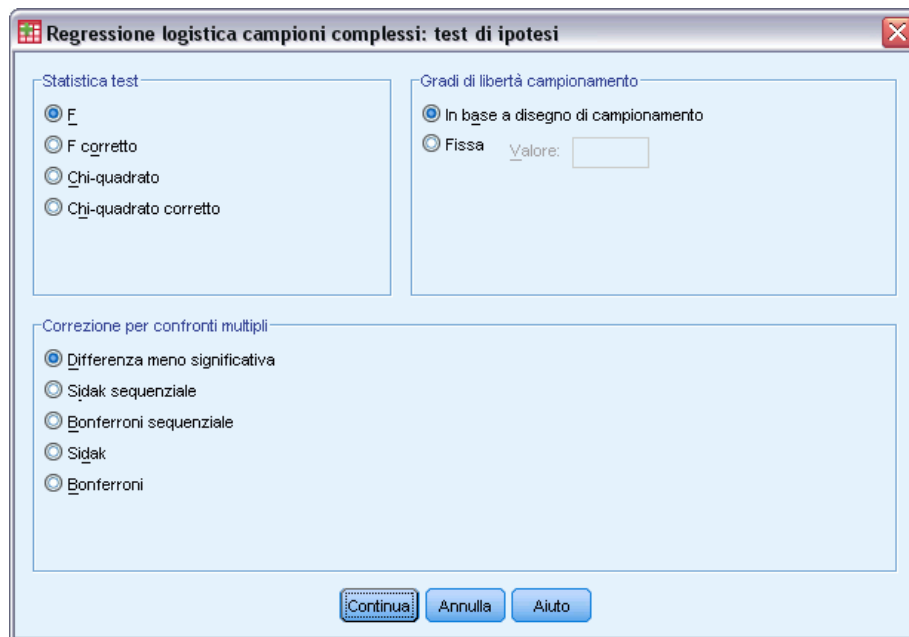
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.

Statistiche riassuntive per le variabili di modello. Visualizza informazioni di riepilogo sulla variabile dipendente, sulle covariate e sui fattori.

Informazioni sul disegno di campionamento. Visualizza informazioni di riepilogo sul campione, compreso il conteggio non pesato e le dimensioni della popolazione.

Modello lineare generalizzato campioni complessi: test di ipotesi

Figura 10-5
Finestra di dialogo Test di ipotesi



Statistica test. Questo gruppo permette di selezionare il tipo di statistica da usare per i test delle ipotesi. È possibile scegliere tra F , F corretta, Chi-quadrato e Chi-quadrato corretto.

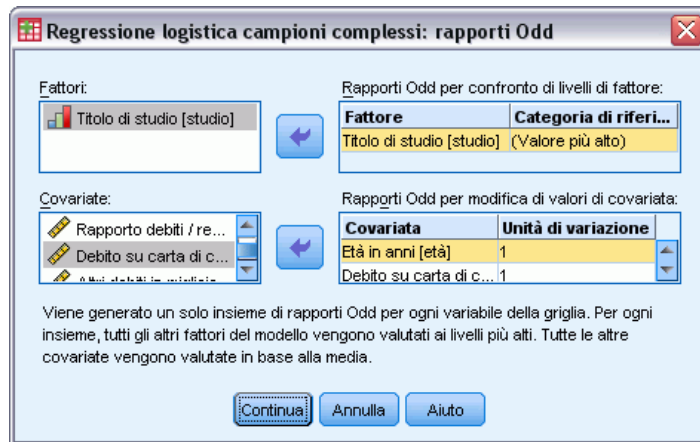
Campionamento dei gradi di libertà. Questo gruppo permette di controllare i gradi di libertà del disegno di campionamento utilizzati per calcolare i valori p di tutte le statistiche di test. Se si utilizza il disegno di campionamento, il valore è pari alla differenza tra il numero delle unità di campionamento primario e il numero di strati della prima fase di campionamento. In alternativa, è possibile impostare un valore personalizzato per i gradi di libertà specificando un intero positivo.

Correzione per confronti multipli. Quando si eseguono test di ipotesi con contrasti multipli, è possibile adattare il livello di significatività globale ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di adattamento.

- **Differenza meno significativa.** Questo metodo non permette di controllare la probabilità complessiva di rifiutare le ipotesi che alcuni contrasti lineari siano diversi dai valori delle ipotesi nulle.
- **Sidak sequenziale.** Una procedura di Sidak con scarti sequenzialmente discendenti, molto meno conservativa per quanto riguarda il rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Bonferroni sequenziale.** Una procedura di Bonferroni con scarti sequenzialmente decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Sidak.** Questo metodo fornisce limiti più stretti rispetto all'approccio di Bonferroni.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.

Rapporto odd regressione logistica campioni complessi

Figura 10-6
Finestra di dialogo Regressione logistica: Rapporti odd



Nella finestra di dialogo Rapporti odd è possibile visualizzare i rapporti odd stimati in base al modello per fattori e covariate specificate. Un gruppo separato di rapporti odd viene calcolato per ogni categoria della variabile dipendente, fatta eccezione per la categoria di riferimento.

Fattori. Per ogni fattore selezionato, visualizza il rapporto degli odd in corrispondenza di ogni categoria del fattore rispetto agli odd in corrispondenza della categoria di riferimento specificata.

Covariate. Per ogni covariata selezionata, visualizza il rapporto degli odd in corrispondenza del valore della media della covariata più le unità di variazione specificate rispetto agli odd in corrispondenza della media.

Quando si calcolano i rapporti odd per un fattore o una covariata, la procedura fissa tutti gli altri fattori sui livelli massimi e tutte le altre covariate sulle relative medie. Se un fattore o una covariata interagisce con altri stimatori nel modello, i rapporti odd dipenderanno non solo dalla variazione nella variabile specificata, ma anche dai valori delle variabili con cui essa interagisce.

Se una covariata specificata interagisce con se stessa nel modello (ad esempio, $età * età$), i rapporti odd dipenderanno sia dalla variazione nella covariata sia dal valore di quest'ultima.

Regressione logistica campioni complessi - Salva

Figura 10-7
Finestra di dialogo Salva regressione logistica



Salva variabili. Il gruppo consente di salvare la categoria e le probabilità previste per il modello come nuove variabili nell'insieme di dati attivo.

Esporta modello come dati SPSS Statistics. Genera un insieme di dati in formato IBM® SPSS® Statistics che contiene la correlazione dei parametri o la matrice di covarianza utilizzando le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file della matrice è come segue.

- **rowtype_.** Assume valori (ed etichette dei valori) COV (Covarianze), CORR (Correlazioni), EST (Stime dei parametri), SE (Errori standard), SIG (Livelli di significatività) e DF (Gradi di libertà del disegno di campionamento). Per ogni modello di parametro vi è un caso separato con tipo di riga COV (o CORR) e anche un caso separato per ciascuno degli altri tipi di riga.
- **varname_.** Assume i valori P1, P2, ..., che corrispondono a un elenco ordinato di tutti i parametri dei modelli per i tipi di riga COV o CORR, con etichette dei valori corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri. Le celle sono vuote per gli altri tipi di riga.
- **P1, P2, ...** Queste variabili corrispondono a un elenco ordinato di tutti i parametri dei modelli con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri e assumono i valori in base al tipo di riga. Per i parametri ridondanti, tutte le covarianze sono impostate su zero e le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate su zero e tutti gli errori standard, i livelli di significatività e i gradi di libertà dei residui sono impostati sul valore mancante di sistema.

Nota: questo file non può essere utilizzato immediatamente per ulteriori analisi in altre procedure che eseguono la lettura di un file di matrice a meno che queste procedure non accettino tutti i tipi di riga esportati qui.

Esporta modello come XML. Consente di salvare le stime dei parametri e la matrice di covarianza dei parametri, se selezionata, in formato XML (PMML). È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Opzioni regressione logistica campioni complessi

Figura 10-8
Finestra di dialogo Regressione logistica: Opzioni

Stima. Il gruppo consente di controllare i vari criteri utilizzati nella stima del modello.

- **Massimo numero di iterazioni.** Il numero massimo di iterazioni che verranno eseguite nell'algoritmo. Specifica un intero non negativo.
- **Massimo numero di dimezzamenti.** Per ciascuna iterazione, la dimensione di passo è ridotta in base al fattore 0,5 finché la verosimiglianza non aumenta o non viene raggiunto il dimezzamento massimo. Specificare un intero positivo.
- **Limitare le iterazioni in base alla variazione nelle stime dei parametri.** Quando l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nelle stime del parametro è minore del valore specificato, che non deve essere negativo.
- **Limitare le iterazioni in base alla variazione nel log verosimiglianza.** Quando l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nel log verosimiglianza è minore del valore specificato, che non deve essere negativo.

- **Verificare la completa separazione dei punti di dati.** Quando l'opzione è selezionata, l'algoritmo esegue i test per garantire che le stime dei parametri abbiano valori univoci. La separazione avviene quando la procedura può generare un modello che classifichi correttamente tutti i casi.
- **Visualizza cronologia iterazioni.** Consente di visualizzare le stime e le statistiche dei parametri per ogni n iterazioni a partire dall'iterazione 0° (stime iniziali). Quando si stamperà la cronologia dell'iterazione, verrà stampata sempre l'ultima iterazione, a prescindere dal valore n .

Valori mancanti definiti dall'utente. Tutte le variabili di disegno, nonché la variabile dipendente ed eventuali covariate, devono avere dati validi. I casi con dati non validi per qualsiasi tra queste variabili vengono esclusi dall'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente vengono considerati come validi nelle variabili di stratificazione, di raggruppamento, di sottopopolazione e di fattore.

Intervallo di confidenza. Il livello dell'intervallo di confidenza per le stime dei coefficienti, le stime dei coefficienti elevate a potenza e i rapporti odd. Specificare un valore maggiore o uguale a 50 e minore di 100.

Opzioni aggiuntive del comando CSLOGISTIC

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test personalizzati di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `CUSTOM`).
- Fissare i valori delle altre variabili di modello il calcolo dei rapporti odd per fattori e covariate (tramite il sottocomando `ODDSRATIOS`).
- Specificare un valore di tolleranza per la verifica della singolarità (tramite il sottocomando `CRITERIA`).
- Creare nomi specificati dall'utente per le variabili salvate (tramite il sottocomando `SAVE`).
- Generare una tabella di forma funzionale generalizzata (tramite il sottocomando `PRINT`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Regressione ordinale campioni complessi

La procedura Regressione ordinale campioni complessi esegue l'analisi di regressione su una variabile dipendente binaria o ordinale per campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Esempio. Prima di promuovere una nuova legge alcuni rappresentanti di governo desiderano sapere se verrà sostenuta a livello pubblico e in che misura tale supporto è correlato ai dati demografici dei votanti. Gli esperti in sondaggi di opinione sviluppano ed eseguono interviste in base a un disegno di campionamento complesso. La regressione ordinale campioni complessi consente di adattare un modello al livello di sostegno della nuova legge tenendo conto dei dati demografici dei votanti.

Dati. La variabile dipendente è ordinale. I fattori sono categoriali. Le covariate sono variabili quantitative correlate alla variabile dipendente. Le variabili di sottopopolazione possono essere numeriche o stringa, ma devono essere categoriali.

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

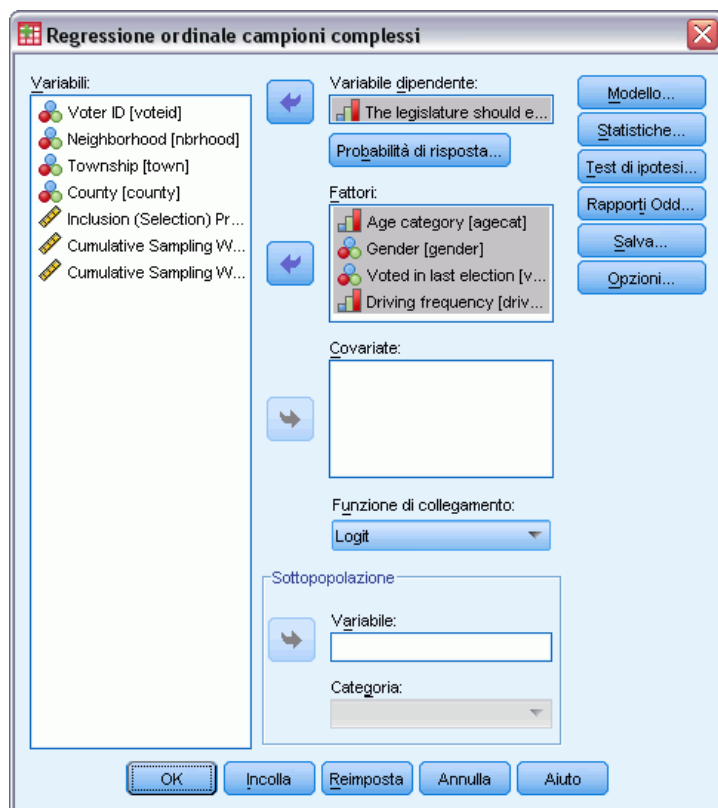
Come ottenere la regressione ordinale campioni complessi

Dai menu, scegliere:

Analizza > Complex Samples > Regressione ordinale...

- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

Figura 11-1
Finestra di dialogo Regressione ordinale



► Selezionare una variabile dipendente.

Se lo si desidera, è possibile:

- Selezionare le variabili per l'opzione Fattori e Covariate, a seconda dei dati in uso.
- Specificare una variabile che definisce una sottopopolazione. L'analisi viene eseguita solo per la categoria della variabile della sottopopolazione selezionata anche se le varianze vengono comunque stimate correttamente utilizzando l'insieme di dati completo.
- Selezionare una funzione di collegamento.

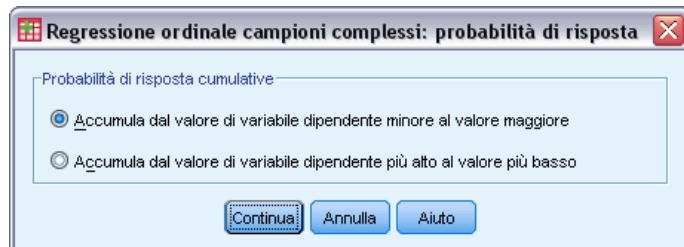
Funzione di collegamento. La funzione Collegamento è la trasformazione delle probabilità cumulate che permette di stimare il modello. Sono disponibili le cinque funzioni di collegamento descritte nella tabella che segue.

Funzione	Formato	Applicazione tipica
Logit	$\log(\xi / (1-\xi))$	Categorie distribuite uniformemente
Log-log complementare	$\log(-\log(1-\xi))$	Categorie alte più probabili
Log-log negativo	$-\log(-\log(\xi))$	Categorie basse più probabili
Probit	$\Phi^{-1}(\xi)$	La variabile latente è normalmente distribuita
Cauchit (funzione Cauchy inversa)	$\tan(\pi(\xi-0.5))$	La variabile latente ha molti valori estremi

Probabilità di risposta della Regressione ordinale campioni complessi

Figura 11-2

Finestra di dialogo Regressione ordinale: Probabilità di risposta

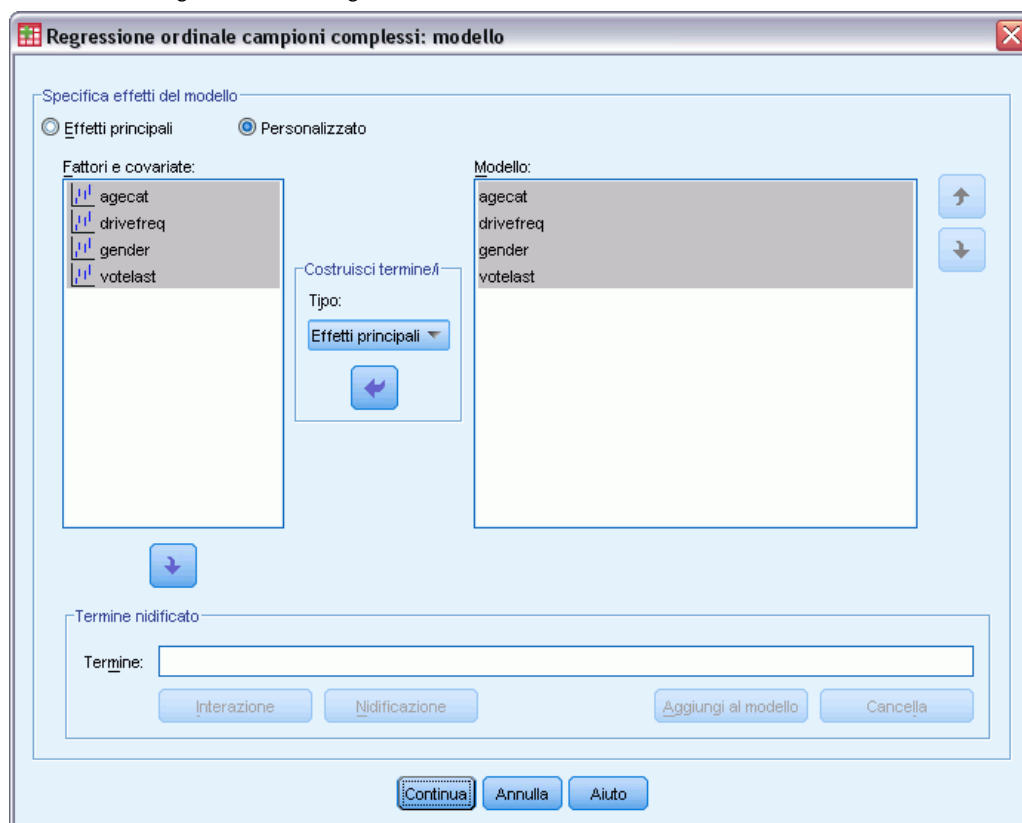


La finestra di dialogo Probabilità di risposta permette di specificare se la probabilità cumulata di una risposta (ossia la probabilità che la risposta appartenga o sia inclusa in una categoria specifica della variabile dipendente) aumenta se i valori della variabile dipendente aumentano o si riducono.

Modello della Regressione ordinale campioni complessi

Figura 11-3

Finestra di dialogo Modello - Regressione ordinale



Specifica modello effetti. Per impostazione predefinita, la procedura crea un modello di effetti principali usando i fattori e le covariate specificati nella finestra di dialogo principale. È possibile anche creare un modello personalizzato che comprenda effetti di interazione e termini nidificati.

Termini non nidificati

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Termini nidificati

Questa procedura consente di costruire termini nidificati per il modello. I termini nidificati sono utili per modellare l'effetto di un fattore o di una covariata i cui valori non interagiscono con i livelli di un altro fattore. Ad esempio, una catena di drogherie può seguire le abitudini di spesa dei propri clienti in più sedi di negozi. Poiché ogni cliente frequenta solo una delle sedi, si può affermare che l'effetto *Cliente* è **nidificato all'interno** dell'effetto *Ubicazione negozio*.

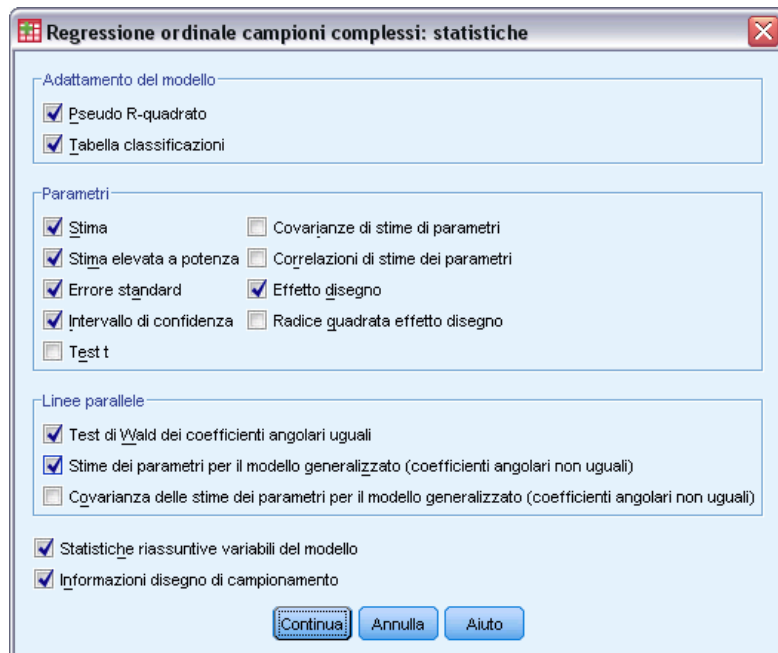
È possibile anche includere gli effetti di interazione, come i termini polinomiali che interessano la stessa covariata, oppure aggiungere più livelli di nidificazione al termine nidificato.

Limitazioni. I termini nidificati sono sottoposti alle seguenti restrizioni:

- Tutti i fattori compresi in un'interazione devono essere univoci. Quindi, se A è un fattore, non è consentito specificare $A*A$.
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Pertanto, se A è un fattore, non è consentito specificare $A(A)$.
- Nessun effetto può essere nidificato all'interno di una covariata. Quindi, se A è un fattore e X è una covariata, non è possibile specificare $A(X)$.

Statistiche della Regressione ordinale campioni complessi

Figura 11-4
Finestra di dialogo Regressione ordinale: Statistiche



Adattamento del modello. Controlla la visualizzazione delle statistiche che misurano l'attendibilità globale del modello.

- **Pseudo R-quadrato.** La statistica R^2 dalla regressione lineare non ha un corrispondente esatto nei modelli di regressione ordinale. Sono invece presenti più misure che tentano di simulare le proprietà della statistica R^2 .
- **Tabella classificazioni.** Visualizza le classificazioni incrociate a tabelle della categoria osservata dichiarata per modello sulla variabile dipendente.

Parametri. Il gruppo consente di controllare la visualizzazione delle statistiche relative ai parametri del modello.

- **Stima.** Visualizza le stime dei coefficienti.
- **Stima elevata a potenza.** Visualizza la base del logaritmo naturale elevata alla potenza delle stime dei coefficienti. Mentre la stima ha proprietà adatte alla verifica statistica, la stima elevata a potenza, o $\exp(B)$, è più facile da interpretare.
- **Errore standard.** Visualizza l'errore standard per ogni stima dei coefficienti.
- **Intervallo di confidenza.** Visualizza un intervallo di confidenza per ogni stima dei coefficienti. Il livello di confidenza per l'intervallo è impostato nella finestra di dialogo Opzioni.
- **Test T.** Visualizza un test t per ogni stima dei coefficienti. L'ipotesi nulla per ogni test corrisponde al caso in cui il valore del coefficiente è 0.
- **Covarianze di stime di parametri** Visualizza una stima della matrice di covarianza per i coefficienti del modello.

- **Correlazioni di stime di parametri** Visualizza una stima della matrice di correlazione per i coefficienti del modello.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** Questa è una misura, espressa in unità equivalenti a quelle dell'errore standard, dell'effetto risultante dall'adozione di un disegno complesso. I valori più lontani da 1 indicano effetti maggiori.

Linee parallele. Questo gruppo consente di richiedere le statistiche associate a un modello con linee non parallele in cui a ciascuna categoria di risposta (ad eccezione dell'ultima) viene adattata una linea di regressione distinta.

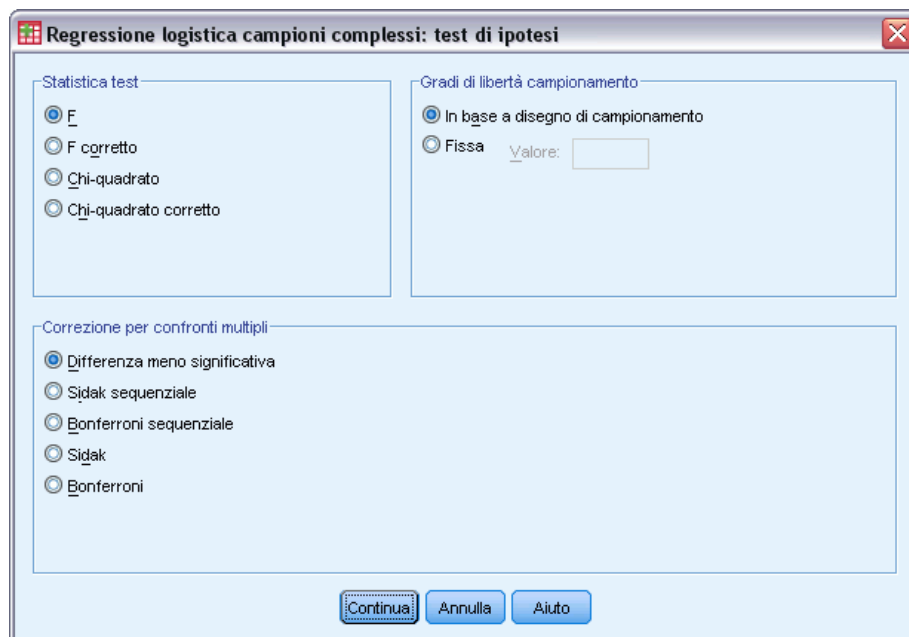
- **Test di Wald.** Genera un test dell'ipotesi nulla secondo cui i parametri di regressione sono uguali per tutte le risposte cumulate. Viene stimato il modello con linee non parallele e applicato il test di Wald dei parametri uguali.
- **Stime dei parametri.** Visualizza le stime dei coefficienti e degli errori standard per il modello con linee non parallele.
- **Covarianze di stime di parametri** Visualizza una stima della matrice di covarianza per i coefficienti del modello con linee non parallele.

Statistiche riassuntive per le variabili di modello. Visualizza informazioni di riepilogo sulla variabile dipendente, sulle covariate e sui fattori.

Informazioni sul disegno di campionamento. Visualizza informazioni di riepilogo sul campione, compreso il conteggio non pesato e le dimensioni della popolazione.

Modello lineare generalizzato campioni complessi: test di ipotesi

Figura 11-5
Finestra di dialogo Test di ipotesi



Statistica test. Questo gruppo permette di selezionare il tipo di statistica da usare per i test delle ipotesi. È possibile scegliere tra F , F corretta, Chi-quadrato e Chi-quadrato corretto.

Campionamento dei gradi di libertà. Questo gruppo permette di controllare i gradi di libertà del disegno di campionamento utilizzati per calcolare i valori p di tutte le statistiche di test. Se si utilizza il disegno di campionamento, il valore è pari alla differenza tra il numero delle unità di campionamento primario e il numero di strati della prima fase di campionamento. In alternativa, è possibile impostare un valore personalizzato per i gradi di libertà specificando un intero positivo.

Correzione per confronti multipli. Quando si eseguono test di ipotesi con contrasti multipli, è possibile adattare il livello di significatività globale ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di adattamento.

- **Differenza meno significativa.** Questo metodo non permette di controllare la probabilità complessiva di rifiutare le ipotesi che alcuni contrasti lineari siano diversi dai valori delle ipotesi nulle.
- **Sidak sequenziale.** Una procedura di Sidak con scarti sequenzialmente discendenti, molto meno conservativa per quanto riguarda il rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Bonferroni sequenziale.** Una procedura di Bonferroni con scarti sequenzialmente decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Sidak.** Questo metodo fornisce limiti più stretti rispetto all'approccio di Bonferroni.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.

Rapporti odd regressione ordinale campioni complessi

Figura 11-6
Finestra di dialogo Regressione lordinale: Rapporti odd



Nella finestra di dialogo Rapporti odd è possibile visualizzare i rapporti odd cumulati stimati in base al modello per fattori e covariate specificate. Questa funzione può essere usata solo per i modelli che utilizzano la funzione di collegamento logit. Viene calcolato un rapporto odd cumulato singolo per tutte le categorie della variabile dipendente eccetto l'ultima. Il modello odd proporzionale presume che siano tutte uguali.

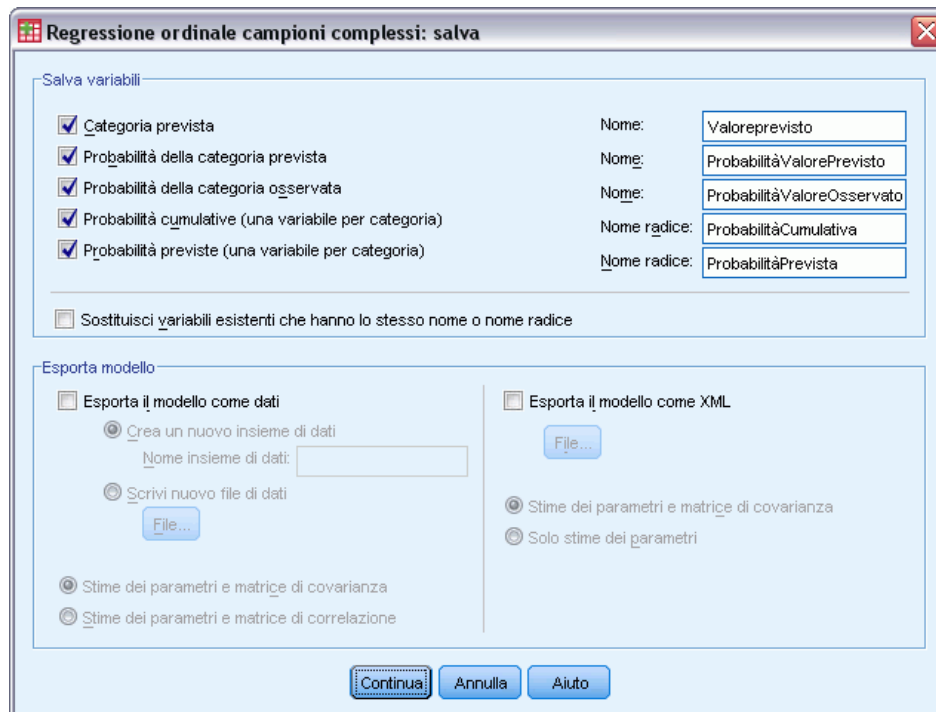
Fattori. Per ogni fattore selezionato, visualizza il rapporto degli odd cumulato in corrispondenza di ogni categoria del fattore rispetto agli odd in corrispondenza della categoria di riferimento specificata.

Covariate. Per ogni covariata selezionata, visualizza il rapporto degli odd cumulato in corrispondenza del valore della media della covariata più le unità di variazione specificate rispetto agli odd in corrispondenza della media.

Quando si calcolano i rapporti odd per un fattore o una covariata, la procedura fissa tutti gli altri fattori sui livelli massimi e tutte le altre covariate sulle relative medie. Se un fattore o una covariata interagisce con altri stimatori nel modello, i rapporti odd dipenderanno non solo dalla variazione nella variabile specificata, ma anche dai valori delle variabili con cui essa interagisce. Se una covariata specificata interagisce con se stessa nel modello (ad esempio, $età*età$), i rapporti odd dipenderanno sia dalla variazione nella covariata sia dal valore di quest'ultima.

Salvataggio della regressione ordinale dei campioni complessi

Figura 11-7
Finestra di dialogo Regressione ordinale: Salva



Salva variabili. Questo gruppo consente di salvare la categoria stimata per il modello, la probabilità della categoria stimata, la probabilità della categoria osservata, le probabilità cumulative e le probabilità previste come nuove variabili nell'insieme di dati attivo.

Esporta modello come dati SPSS Statistics. Genera un insieme di dati in formato IBM® SPSS® Statistics che contiene la correlazione dei parametri o la matrice di covarianza utilizzando le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file della matrice è come segue.

- **rowtype_.** Assume valori (ed etichette dei valori) COV (Covarianze), CORR (Correlazioni), EST (Stime dei parametri), SE (Errori standard), SIG (Livelli di significatività) e DF (Gradi di libertà del disegno di campionamento). Per ogni modello di parametro vi è un caso separato con tipo di riga COV (o CORR) e anche un caso separato per ciascuno degli altri tipi di riga.
- **varname_.** Assume i valori P1, P2, ..., che corrispondono a un elenco ordinato di tutti i parametri dei modelli per i tipi di riga COV o CORR, con etichette dei valori corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri. Le celle sono vuote per gli altri tipi di riga.
- **P1, P2, ...** Queste variabili corrispondono a un elenco ordinato di tutti i parametri dei modelli con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri e assumono i valori in base al tipo di riga. Per i parametri ridondanti, tutte le covarianze sono impostate su zero e le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate su zero e tutti gli

errori standard, i livelli di significatività e i gradi di libertà dei residui sono impostati sul valore mancante di sistema.

Nota: questo file non può essere utilizzato immediatamente per ulteriori analisi in altre procedure che eseguono la lettura di un file di matrice a meno che queste procedure non accettino tutti i tipi di riga esportati qui.

Esporta modello come XML. Consente di salvare le stime dei parametri e la matrice di covarianza dei parametri, se selezionata, in formato XML (PMML). È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Opzioni della regressione ordinale campioni complessi

Figura 11-8
Finestra di dialogo Regressione ordinale: Opzioni

Metodo di stima. È possibile selezionare il metodo di stima tra i seguenti metodi: Newton-Raphson e punteggio di Fisher. In alternativa, è possibile anche usare un metodo ibrido in cui vengono effettuate le iterazioni dei punteggi di Fisher prima di passare al metodo di Newton-Raphson. Se si ottiene la convergenza durante la fase del punteggio di Fisher del metodo ibrido, ossia prima che venga raggiunto il numero massimo di iterazioni di Fisher, l'algoritmo continua con il metodo di Newton-Raphson.

Stima. Il gruppo consente di controllare i vari criteri utilizzati nella stima del modello.

- **Massimo numero di iterazioni.** Il numero massimo di iterazioni che verranno eseguite nell'algoritmo. Specifica un intero non negativo.

- **Massimo numero di dimezzamenti.** Per ciascuna iterazione, la dimensione di passo è ridotta in base al fattore 0,5 finché la verosimiglianza non aumenta o non viene raggiunto il dimezzamento massimo. Specificare un intero positivo.
- **Limitare le iterazioni in base alla variazione nelle stime dei parametri.** Quando l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nelle stime del parametro è minore del valore specificato, che non deve essere negativo.
- **Limitare le iterazioni in base alla variazione nel log verosimiglianza.** Quando l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nel log verosimiglianza è minore del valore specificato, che non deve essere negativo.
- **Verificare la completa separazione dei punti di dati.** Quando l'opzione è selezionata, l'algoritmo esegue i test per garantire che le stime dei parametri abbiano valori univoci. La separazione avviene quando la procedura può generare un modello che classifichi correttamente tutti i casi.
- **Visualizza cronologia iterazioni.** Consente di visualizzare le stime e le statistiche dei parametri per ogni n iterazioni a partire dall'iterazione 0° (stime iniziali). Quando si stamperà la cronologia dell'iterazione, verrà stampata sempre l'ultima iterazione, a prescindere dal valore n .

Valori mancanti definiti dall'utente. Le variabili scala di disegno, nonché la variabile dipendente ed eventuali covariate, devono avere dati validi. I casi con dati non validi per qualsiasi tra queste variabili vengono esclusi dall'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente vengono considerati come validi nelle variabili di stratificazione, di raggruppamento, di sottopopolazione e di fattore.

Intervallo di confidenza. Il livello dell'intervallo di confidenza per le stime dei coefficienti, le stime dei coefficienti elevate a potenza e i rapporti odd. Specificare un valore maggiore o uguale a 50 e minore di 100.

Funzioni aggiuntive del comando CSORDINAL

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test personalizzati di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `CUSTOM`).
- Fissare i valori delle altre variabili di modello su valori diversi dalle medie durante il calcolo dei rapporti odd cumulati per i fattori e le covariate (utilizzando il sottocomando `ODDSRATIOS`).
- Utilizzare valori non etichettati come categorie di riferimento personalizzate per i fattori se sono richiesti i rapporti odd (utilizzando il sottocomando `ODDSRATIOS`).
- Specificare un valore di tolleranza per la verifica della singolarità (tramite il sottocomando `CRITERIA`).
- Generare una tabella di forma funzionale generalizzata (tramite il sottocomando `PRINT`).
- Salvare più di 25 variabili di probabilità (utilizzando il sottocomando `SAVE`).

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Regressione Cox di campioni complessi

La procedura Regressione Cox di campioni complessi esegue l'analisi di sopravvivenza su campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Esempi. Le Forze dell'Ordine sono preoccupate del tasso di recidività nella propria giurisdizione. Una delle misure di recidività è il tempo che trascorre fino al secondo arresto del trasgressore. Le Forze dell'Ordine desiderano creare un modello relativo al tempo trascorso fino a un nuovo arresto utilizzando la regressione di Cox, ma temono che l'ipotesi dei rischi proporzionali non sia valida per le diverse categorie di età.

Alcuni ricercatori medici stanno studiando i tempi di sopravvivenza dei pazienti che hanno completato un programma riabilitativo in seguito a un ictus postischemico. Poiché le anamnesi dei pazienti cambiano quando viene annotata l'occorrenza di eventi non mortali significativi e vengono registrati i tempi in cui tali eventi si sono verificati, esiste il potenziale per più casi per ogni soggetto. Il campione viene inoltre troncato a sinistra nel senso che i tempi di sopravvivenza osservati vengono "gonfiati" con la durata della riabilitazione, poiché l'insorgenza del rischio ha inizio al momento dell'attacco ischemico, ma il campione include solo i pazienti che sopravvivono al termine del programma riabilitativo.

Ora di sopravvivenza. La procedura applica la regressione di Cox all'analisi dei tempi di sopravvivenza, ovvero del periodo di tempo trascorso prima che si verifichi un evento. Esistono due modi per specificare il tempo di sopravvivenza in base all'ora di inizio dell'intervallo:

- **Ora = 0.** In genere, sono disponibili informazioni complete sull'inizio dell'intervallo per ogni soggetto ed è sufficiente utilizzare una variabile contenente le ore finali (oppure creare una singola variabile con le ore finali dalle variabili relative a data e ora; vedere quanto riportato di seguito).
- **Varia con il soggetto.** Questa funzione è appropriata con il **troncamento a sinistra**, chiamato anche **inserimento ritardato**; ad esempio se si analizzano i tempi di sopravvivenza per i pazienti che hanno terminato un programma di riabilitazione post-ictus, è possibile considerare che il rischio inizi al momento dell'ictus. Tuttavia, se vengono inclusi solo i pazienti sopravvissuti al termine del programma di riabilitazione, il campione viene troncato a sinistra nel senso che i tempi di sopravvivenza osservati vengono "gonfiati" con la durata della riabilitazione. È possibile rappresentare questo specificando il periodo in cui i pazienti hanno concluso la riabilitazione come il periodo di ingresso nel caso di studio.

Variabili di data e ora. Per definire direttamente l'inizio e la fine dell'intervallo, non è possibile utilizzare le variabili di data e ora; se sono disponibili, utilizzarle per creare variabili contenenti i tempi di sopravvivenza. Se non viene effettuato il troncamento a sinistra, è sufficiente creare una variabile contenente le ore finali in base alla differenza tra la data di inserimento nello studio e la data di osservazione. Se viene effettuato il troncamento a sinistra, creare una variabile contenente le ore iniziali in base alla differenza tra la data di inizio dello studio e la data di inserimento e

una variabile contenente le ore finali in base alla differenza tra la data di inizio dello studio e la data di osservazione.

Stato evento. È necessaria una variabile che registri se si è verificato l'evento di interesse per il soggetto nell'ambito dell'intervallo. I soggetti per i quali l'evento non si è verificato vengono troncati a destra.

Identificatore soggetto. È possibile inserire facilmente predittori dipendenti dal tempo con parti costanti suddividendo le osservazioni per un singolo soggetto tra più casi. Ad esempio, se si stanno analizzando i tempi di sopravvivenza dei pazienti colpiti da ictus, le variabili che rappresentano l'anamnesi sono utili come predittori. Nel tempo si possono verificare eventi medici importanti in grado di modificare l'anamnesi dei pazienti. La tabella riportata di seguito mostra la modalità di strutturazione di un insieme di dati: *ID paziente* è l'identificatore del soggetto, *Ora finale* definisce gli intervalli osservati, *Stato* registra i principali eventi medici e *Attacchi ischemici precedenti* e *Emorragie precedenti* sono predittori dipendenti dal tempo con parti costanti.

<i>ID paziente</i>	<i>Ora finale</i>	<i>Stato</i>	<i>Attacchi ischemici precedenti</i>	<i>Emorragie precedenti</i>
1	5	Attacco ischemico	No	No
1	7	Emorragia	Sì	No
1	8	Decessi	Sì	Sì
2	24	Decessi	No	No
3	8	Attacco ischemico	No	No
3	15	Decessi	Sì	No

Assunzioni. I casi del file di dati rappresentano un campione da un disegno complesso, che deve essere analizzato in base alle specifiche definite nel file selezionato nella [finestra di dialogo Piano campioni complessi](#).

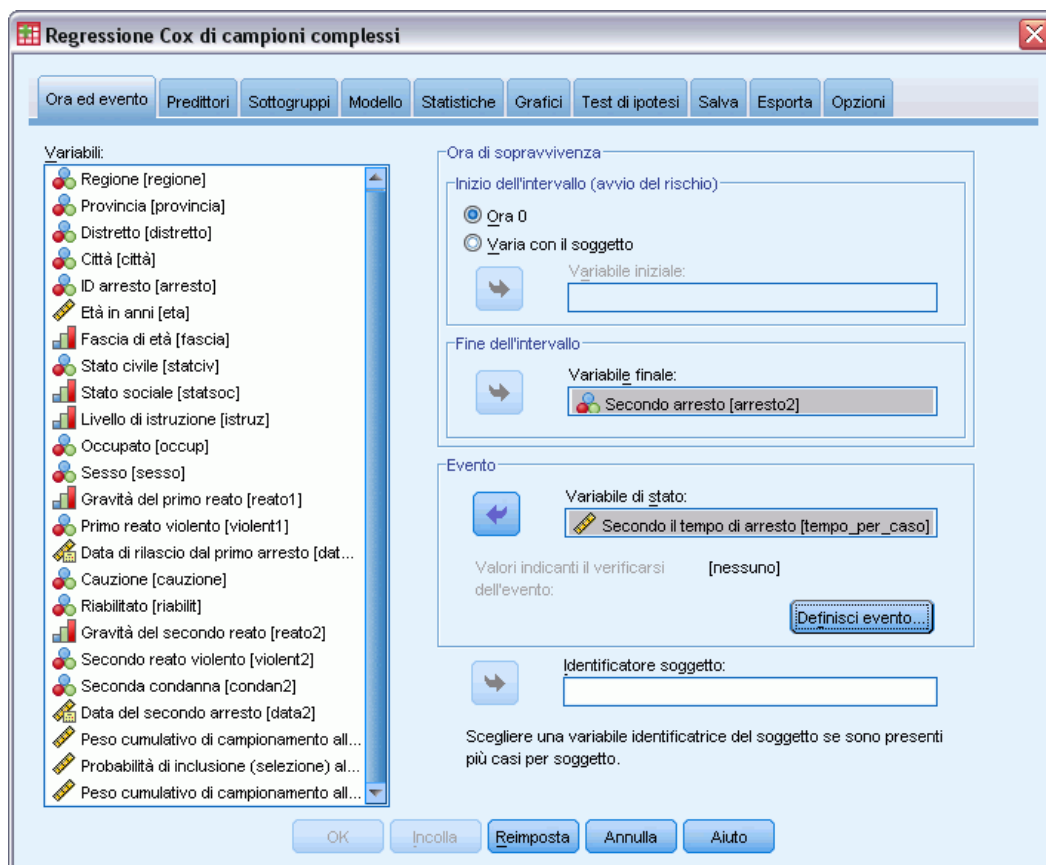
In genere i modelli di regressione di Cox presuppongono l'esistenza di rischi proporzionali, ovvero il tasso di rischi da un caso all'altro non dovrebbe variare nel tempo. Se questa ipotesi non è valida, è possibile che sia necessario aggiungere al modello predittori dipendenti dal tempo.

Analisi di Kaplan-Meier. Se non si selezionano predittori (o non si inseriscono i predittori selezionati nel modello) e si sceglie il metodo del prodotto-limite per il calcolo della curva di sopravvivenza di base nella scheda Opzioni, la procedura esegue un tipo di analisi della sopravvivenza di Kaplan-Meier.

Per ottenere una regressione di Cox di campioni complessi

- ▶ Dai menu, scegliere:
Analizza > Complex Samples > Regressione di Cox...
- ▶ Selezionare un file del piano. Se necessario, selezionare un file delle probabilità congiunte personalizzato.
- ▶ Fare clic su Continua.

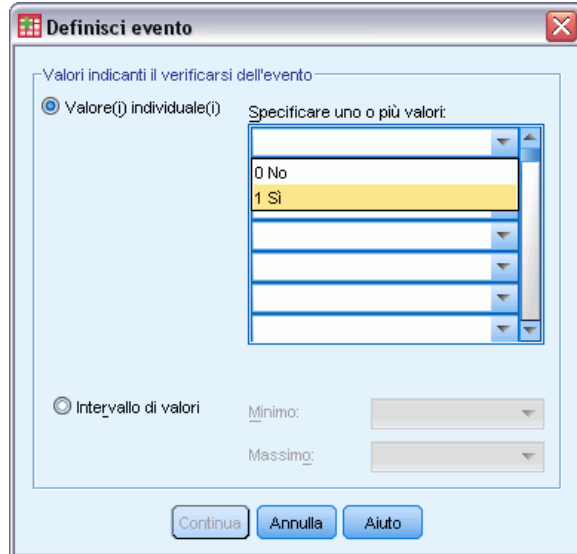
Figura 12-1
Finestra di dialogo regressione di Cox, scheda Ora ed evento



- ▶ Specificare il tempo di sopravvivenza selezionando i tempi di inserimento nello studio e di eliminazione dallo studio.
 - ▶ Selezionare una variabile relativa allo stato dell'evento.
 - ▶ Fare clic su [Definisci evento](#) e definire almeno un valore.
- In alternativa, è possibile selezionare un identificatore di soggetto.

Definisci evento

Figura 12-2
Finestra di dialogo Definisci evento

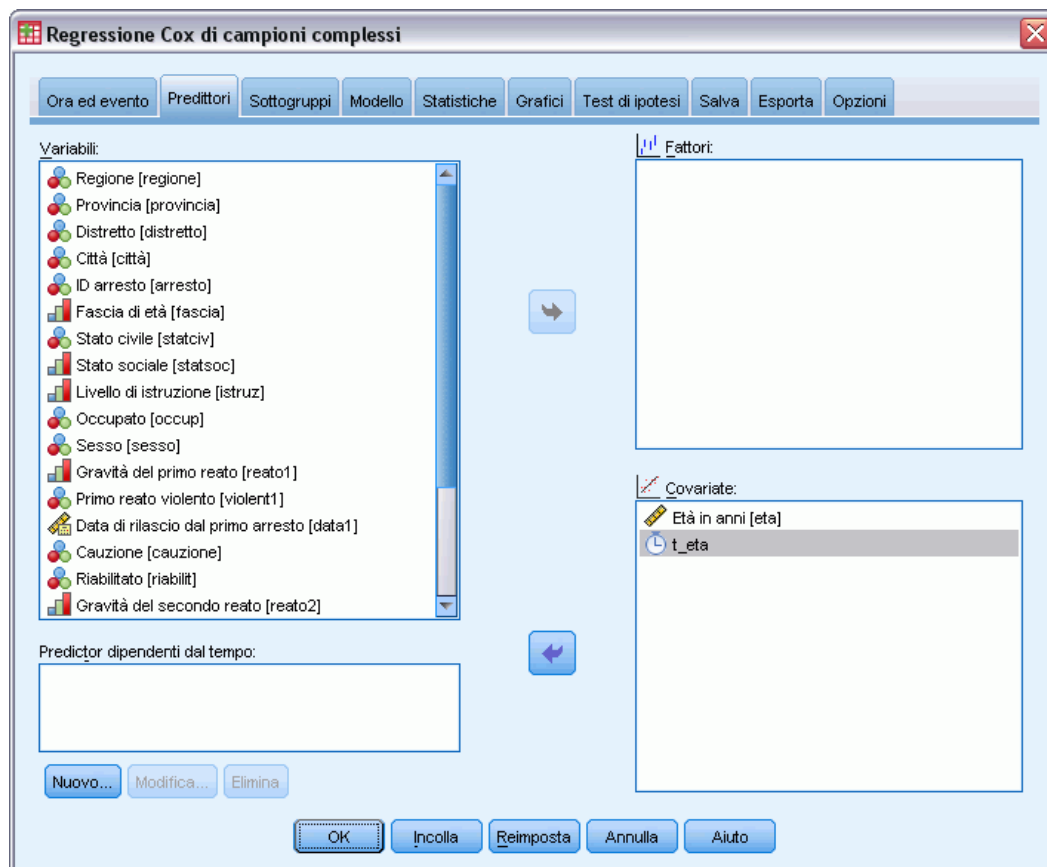


Specificare i valori che indicano che si è verificato un evento terminale.

- **Valore(i) individuale(i).** Specificare uno o più valori inserendoli nella griglia o effettuando una selezione da un elenco di valori con etichette di valori definite.
- **Intervallo di valori.** Specificare un intervallo di valori inserendo i valori minimo e massimo o effettuando una selezione da un elenco con etichette di valori definite.

Predittori

Figura 12-3
Finestra di dialogo Regressione di Cox, scheda Predittori



La scheda Predittori consente di specificare i fattori e le covariate utilizzati per definire gli effetti del modello.

Fattori. I fattori sono predittori categoriali e possono essere di tipo numerico o stringa.

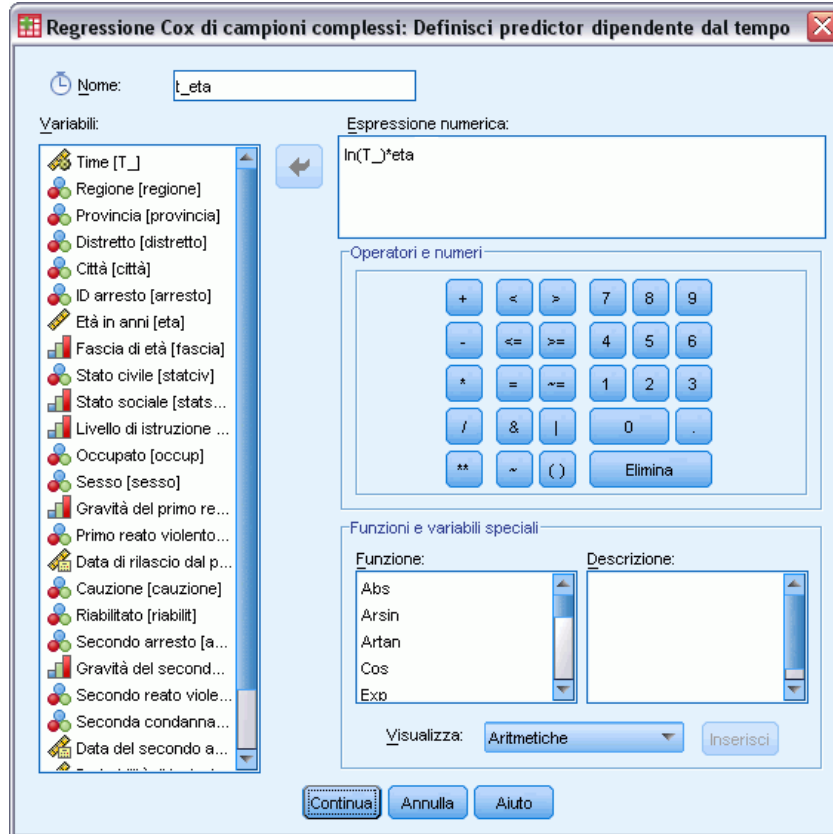
Covariate. Le covariate sono predittori scala e devono essere di tipo numerico.

Predittori dipendenti dal tempo. In determinate situazioni l'ipotesi di rischi proporzionali non è valida, ovvero i tassi di rischio variano nel tempo e i valori di uno o più predittori sono diversi in momenti differenti. In alcuni casi, è necessario specificare predittori dipendenti dal tempo. [Per ulteriori informazioni, vedere l'argomento Definisci predictor dipendente dal tempo a pag. 82.](#) È possibile selezionare i predittori dipendenti dal tempo come fattori o covariate.

Definisci predictor dipendente dal tempo

Figura 12-4

Finestra di dialogo Regressione Cox di campioni complessi: Definisci predictor dipendente dal tempo



La finestra di dialogo Definisci predictor dipendente dal tempo consente di creare un predittore dipendente da una variabile di tempo incorporata, $T_$. La variabile può essere usata in due modi diversi:

- Se si desidera stimare un modello di regressione di Cox esteso che consenta rischi non proporzionali, è possibile definire il predittore dipendente dal tempo come funzione della variabile di tempo $T_$ e della covariata specificata. Un esempio comune è il prodotto semplice della variabile di tempo e del predittore; è tuttavia possibile specificare anche funzioni più complesse.
- I valori di alcune variabili potrebbero essere diversi in periodi di tempo diversi. Tra queste variabili e il tempo non esiste sempre una relazione. In questi casi è necessario definire un **predittore dipendente dal tempo segmentato** tramite espressioni logiche. Le espressioni logiche sono uguali a 1 se sono vere e uguali a 0 se sono false. Tramite una serie di espressioni logiche è possibile creare un predittore dipendente dal tempo in base a un insieme di misure. Se, ad esempio, si esegue un controllo della pressione sanguigna una volta alla settimana per le quattro settimane dello studio (identificate da $BP1$ a $BP4$), il predittore dipendente dal tempo può essere definito nel modo seguente: $(T_ < 1) * BP1 + (T_ \geq 1 \ \& \ T_ < 2) * BP2 + (T_ \geq 2 \ \& \ T_ < 3) * BP3 + (T_ \geq 3 \ \& \ T_ < 4) * BP4$. Per qualsiasi caso considerato, uno solo dei termini tra parentesi sarà uguale a uno, mentre tutti gli altri saranno uguali a 0, ovvero questa

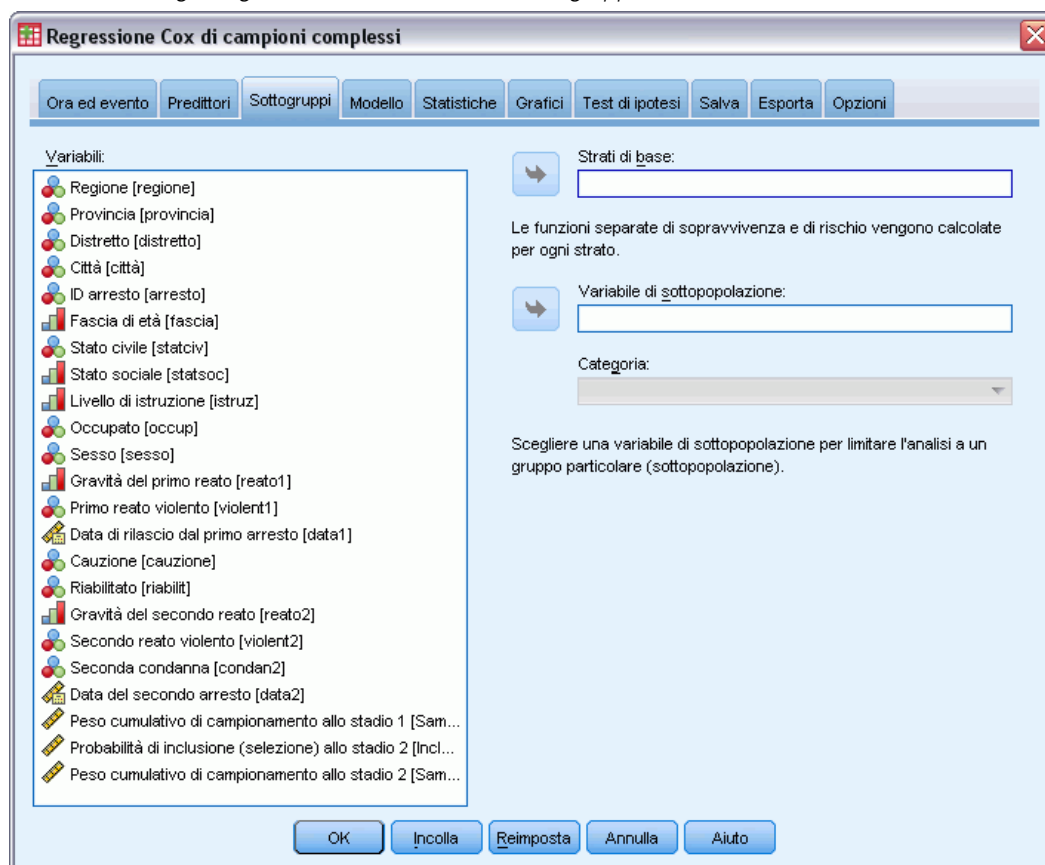
funzione può essere interpretata nel modo seguente: se il periodo di tempo è inferiore a una settimana, usare *BPI*, se è superiore a una settimana ma inferiore a due, usare *BP2* e così via.

Nota: se il predittore dipendente dal tempo segmentato è costante all'interno dei segmenti, come nell'esempio relativo alla pressione del sangue riportato in precedenza, potrebbe essere più semplice specificare il predittore dipendente dal tempo con parti costanti suddividendo i soggetti tra più casi. Per ulteriori informazioni sugli identificatori dei soggetti, vedere [Regressione Cox di campioni complessi](#) a pag. 77.

Nella finestra di dialogo Definisci predictor dipendente dal tempo è possibile utilizzare i controlli relativi alla creazione delle funzioni per creare l'espressione relativa alla covariata dipendente dal tempo oppure è possibile inserirla direttamente nell'area di testo Espressione numerica. Le costanti di stringa devono essere racchiuse tra virgolette semplici o doppie, mentre le costanti numeriche devono essere specificate nel formato americano, ovvero usando il punto come separatore decimale. La variabile risultante è data dal nome specificato e deve essere inclusa come fattore o covariata nella scheda Predittori.

Sottogruppi

Figura 12-5
Finestra di dialogo Regressione di Cox, scheda Sottogruppi

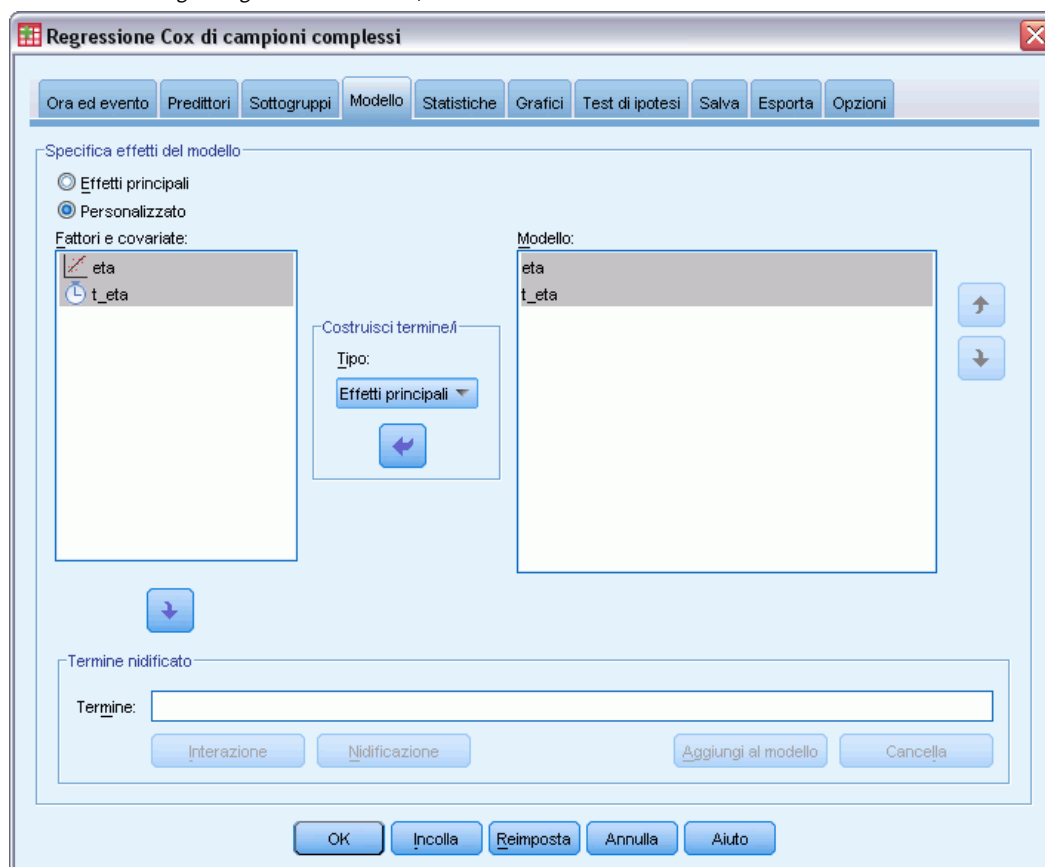


Strati di base. Per ciascun valore di questa variabile viene calcolata una funzione di sopravvivenza e di rischio di base separata, mentre per gli strati viene stimato un singolo insieme di coefficienti del modello.

Variabile di sottopopolazione. Specificare una variabile per definire una sottopopolazione. L'analisi viene eseguita solo per la categoria selezionata della variabile della sottopopolazione.

Modello

Figura 12-6
Finestra di dialogo *Regressione di Cox*, scheda *Modello*



Specifica modello effetti. Per impostazione predefinita, la procedura crea un modello di effetti principali usando i fattori e le covariate specificati nella finestra di dialogo principale. È possibile anche creare un modello personalizzato che comprenda effetti di interazione e termini nidificati.

Termini non nidificati

Per i fattori e le covariate selezionati:

Interazione. Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate.

Effetti principali. Consente di creare un termine di effetti principali per ciascuna variabile selezionata.

Tutti 2-vie. Consente di creare tutte le possibili interazioni a due vie delle variabili selezionate.

Tutti 3-vie. Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

Tutti 4-vie. Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

Tutti 5-vie. Consente di creare tutte le possibili interazioni a cinque vie delle variabili selezionate.

Termini nidificati

Questa procedura consente di costruire termini nidificati per il modello. I termini nidificati sono utili per modellare l'effetto di un fattore o di una covariata i cui valori non interagiscono con i livelli di un altro fattore. Ad esempio, una catena di drogherie può seguire le abitudini di spesa dei propri clienti in più sedi di negozi. Poiché ogni cliente frequenta solo una delle sedi, si può affermare che l'effetto *Cliente* è **nidificato all'interno** dell'effetto *Ubicazione negozio*.

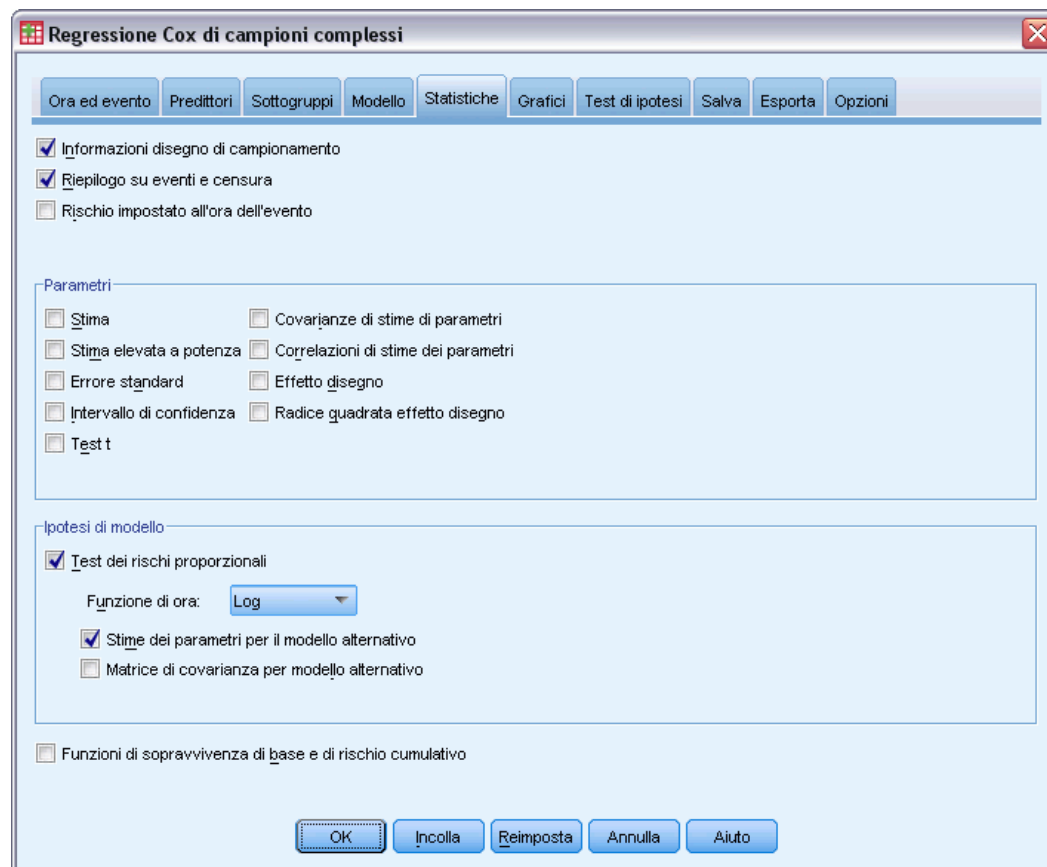
È possibile anche includere gli effetti di interazione, come i termini polinomiali che interessano la stessa covariata, oppure aggiungere più livelli di nidificazione al termine nidificato.

Limitazioni. I termini nidificati sono sottoposti alle seguenti restrizioni:

- Tutti i fattori compresi in un'interazione devono essere univoci. Quindi, se A è un fattore, non è consentito specificare $A*A$.
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Pertanto, se A è un fattore, non è consentito specificare $A(A)$.
- Nessun effetto può essere nidificato all'interno di una covariata. Quindi, se A è un fattore e X è una covariata, non è possibile specificare $A(X)$.

Statistiche

Figura 12-7
Finestra di dialogo *Regressione di Cox*, scheda *Statistiche*



Informazioni sul disegno di campionamento. Visualizza informazioni di riepilogo sul campione, compreso il conteggio non pesato e le dimensioni della popolazione.

Riepilogo su eventi e censura. Visualizza le informazioni di riepilogo sul numero e la percentuale di casi troncati.

Rischio impostato all'ora dell'evento. Visualizza il numero di eventi e il numero a rischio per ogni ora dell'evento in ogni strato di base.

Parametri. Il gruppo consente di controllare la visualizzazione delle statistiche relative ai parametri del modello.

- **Stima.** Visualizza le stime dei coefficienti.
- **Stima elevata a potenza.** Visualizza la base del logaritmo naturale elevata alla potenza delle stime dei coefficienti. Mentre la stima ha proprietà adatte alla verifica statistica, la stima elevata a potenza, o $\exp(B)$, è più facile da interpretare.
- **Errore standard.** Visualizza l'errore standard per ogni stima dei coefficienti.
- **Intervallo di confidenza.** Visualizza un intervallo di confidenza per ogni stima dei coefficienti. Il livello di confidenza per l'intervallo è impostato nella finestra di dialogo Opzioni.

- **Test t.** Visualizza un test t per ogni stima dei coefficienti. L'ipotesi nulla per ogni test corrisponde al caso in cui il valore del coefficiente è 0.
- **Covarianze di stime di parametri** Visualizza una stima della matrice di covarianza per i coefficienti del modello.
- **Correlazioni di stime di parametri** Visualizza una stima della matrice di correlazione per i coefficienti del modello.
- **Effetto disegno.** Il rapporto fra la varianza della stima e la varianza ottenuta assumendo che il campione sia un campione casuale semplice. È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più piccoli indicano effetti maggiori.
- **Radice quadrata dell'effetto disegno.** È una misura dell'effetto derivato dalla specificazione di un disegno complesso, dove valori più lontani da 1 indicano effetti maggiori.

Ipotesi di modello. Questo gruppo consente di generare un test dell'ipotesi dei rischi proporzionali. Il test confronta il modello adattato a un modello alternativo che include i predittori dipendenti dal tempo $x*_TF$ per ogni predittore x , dove $_TF$ è la funzione di ora specificata.

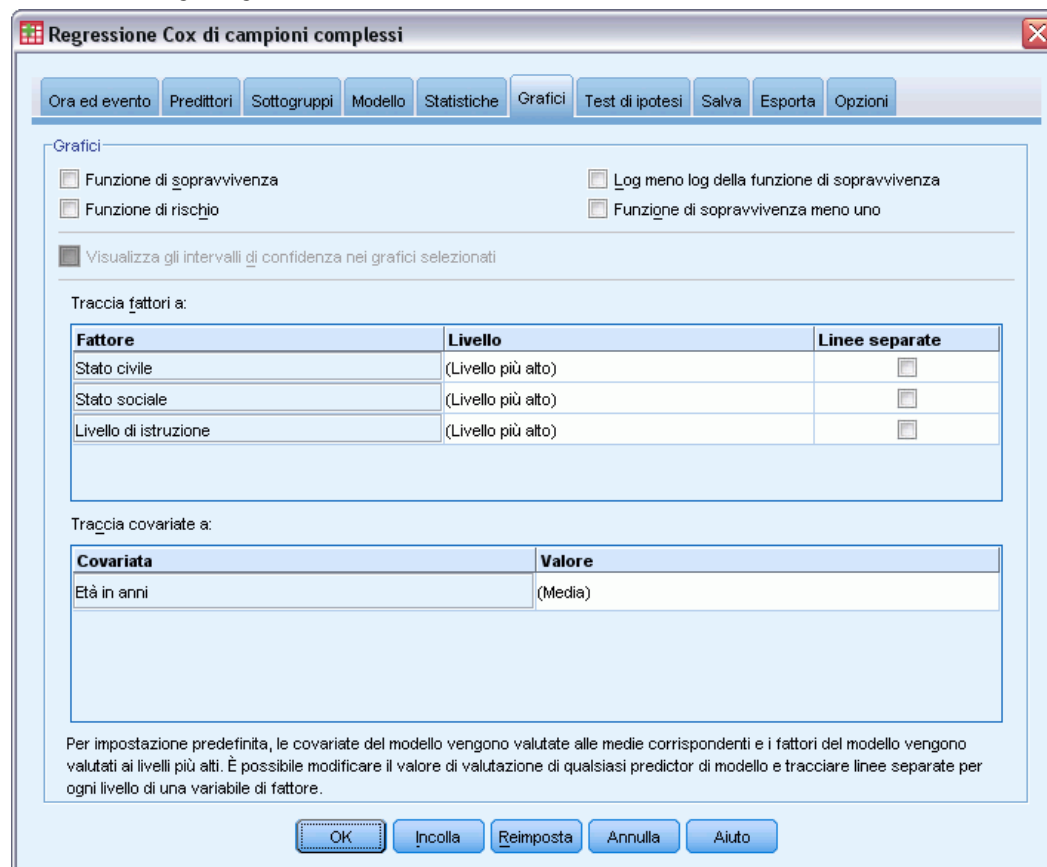
- **Funzione di ora.** Specifica la forma di $_TF$ per il modello alternativo. Per la funzione di **identità**, $_TF=T_$. Per la funzione **log**, $_TF=\log(T_)$. Per **Kaplan-Meier**, $_TF=1-S_{KM}(T_)$, dove $S_{KM}(\cdot)$ è la stima Kaplan-Meier della funzione di sopravvivenza. Per **rango**, $_TF$ è il punteggio di rango di $T_$ tra le ore finali osservate.
- **Stime dei parametri per il modello alternativo.** Visualizza la stima, l'errore standard e l'intervallo di confidenza per ogni parametro nel modello alternativo.
- **Matrice di covarianza per modello alternativo.** Visualizza la matrice delle covarianze stimate tra i parametri nel modello alternativo.

Funzioni di sopravvivenza di base e di rischio cumulativo. Visualizza la funzione di sopravvivenza di base e la funzione di rischio cumulativo di base insieme agli errori standard.

Nota: se i predittori dipendenti dal tempo definiti nella scheda Predittori vengono inclusi nel modello, questa opzione non è disponibile.

Grafici

Figura 12-8
Finestra di dialogo *Regressione di Cox*, scheda *Grafici*



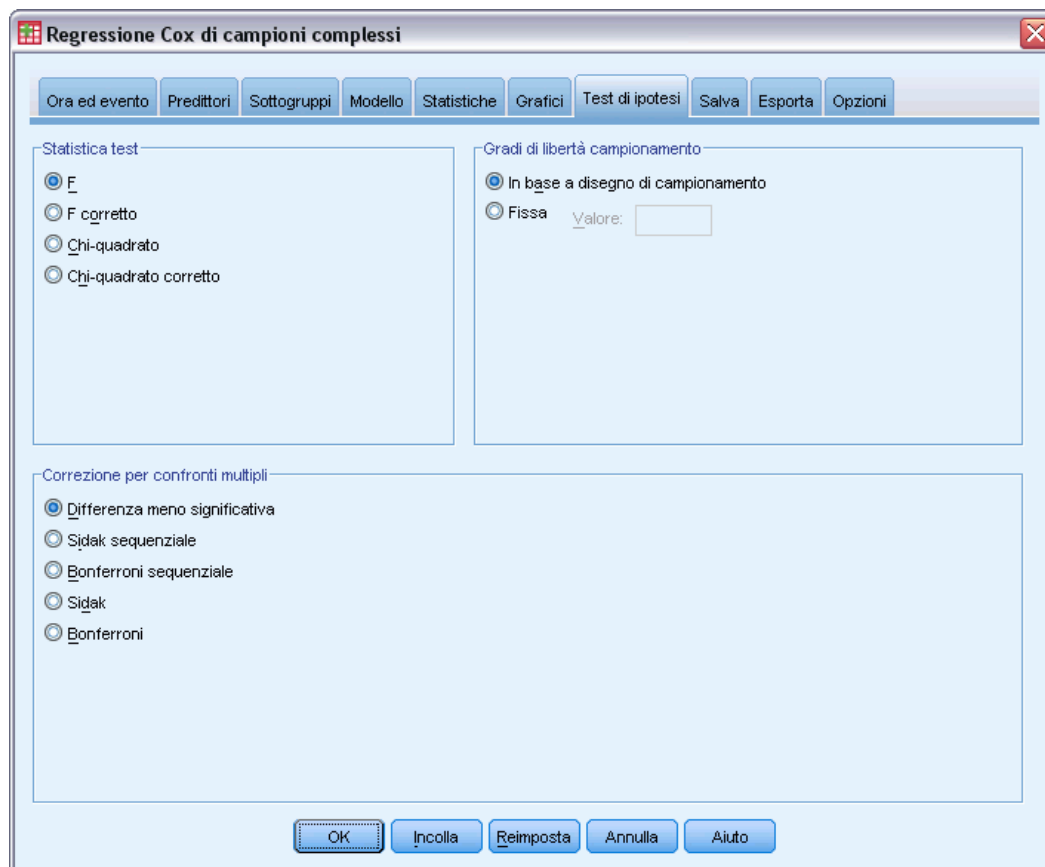
La scheda Grafici consente di richiedere i grafici delle funzioni di rischio e di sopravvivenza, della funzione log meno log della sopravvivenza e della funzione uno meno sopravvivenza. È possibile anche scegliere di tracciare intervalli di confidenza lungo le funzioni specificate; il livello di confidenza viene impostato nella scheda Opzioni.

Modelli predittore. È possibile specificare un modello di valori predittore da utilizzare per i grafici richiesti e il file di sopravvivenza esportato nella scheda Esporta. Se i predittori dipendenti dal tempo definiti nella scheda Predittori vengono inclusi nel modello, queste opzioni non sono disponibili.

- **Traccia fattori a.** Per impostazione predefinita, ogni fattore viene valutato al livello più alto. Immettere o selezionare un livello differente, se si desidera. In alternativa, è possibile scegliere di tracciare linee separate per ogni livello di un singolo fattore selezionando la casella di controllo relativa a tale fattore.
- **Traccia covariate a.** Ogni covariata viene valutata al valore medio. Immettere o selezionare un valore differente, se si desidera.

Test di ipotesi

Figura 12-9
Finestra di dialogo Regressione di Cox, scheda Test di ipotesi



Statistica test. Questo gruppo permette di selezionare il tipo di statistica da usare per i test delle ipotesi. È possibile scegliere tra F , F corretta, Chi-quadrato e Chi-quadrato corretto.

Campionamento dei gradi di libertà. Questo gruppo permette di controllare i gradi di libertà del disegno di campionamento utilizzati per calcolare i valori p di tutte le statistiche di test. Se si utilizza il disegno di campionamento, il valore è pari alla differenza tra il numero delle unità di campionamento primario e il numero di strati della prima fase di campionamento. In alternativa, è possibile impostare un valore personalizzato per i gradi di libertà specificando un intero positivo.

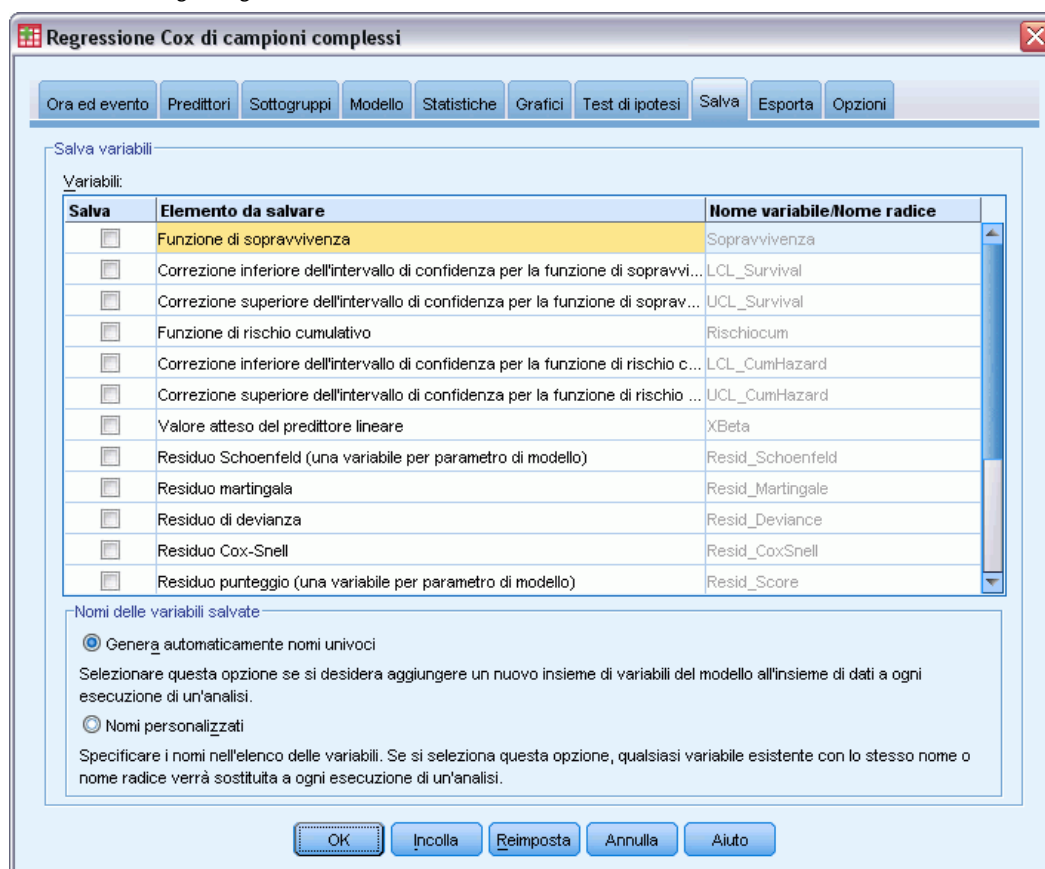
Correzione per confronti multipli. Quando si eseguono test di ipotesi con contrasti multipli, è possibile adattare il livello di significatività globale ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di adattamento.

- **Differenza meno significativa.** Questo metodo non permette di controllare la probabilità complessiva di rifiutare le ipotesi che alcuni contrasti lineari siano diversi dai valori delle ipotesi nulle.
- **Sidak sequenziale.** Una procedura di Sidak con scarti sequenzialmente discendenti, molto meno conservativa per quanto riguarda il rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.

- **Bonferroni sequenziale.** Una procedura di Bonferroni con scarti sequenzialmente decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.
- **Sidak.** Questo metodo fornisce limiti più stretti rispetto all'approccio di Bonferroni.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.

Salva

Figura 12-10
Finestra di dialogo *Regressione di Cox*, scheda *Salva*



Salva variabili. Questo gruppo consente di salvare le variabili correlate al modello nell'insieme di dati attivo per consentirne l'utilizzo successivo per la diagnostica e la creazione di report dei risultati. Nessuna di queste opzioni è disponibile se nel modello sono inclusi i predittori dipendenti dal tempo.

- **Funzione di sopravvivenza.** Salva la probabilità di sopravvivenza (il valore della funzione di sopravvivenza) al momento dell'osservazione e i valori del predittore per ogni caso.
- **Correzione inferiore dell'intervallo di confidenza per la funzione di sopravvivenza.** Salva il limite inferiore dell'intervallo di confidenza per la funzione di sopravvivenza al momento dell'osservazione e i valori predittori per ogni caso.

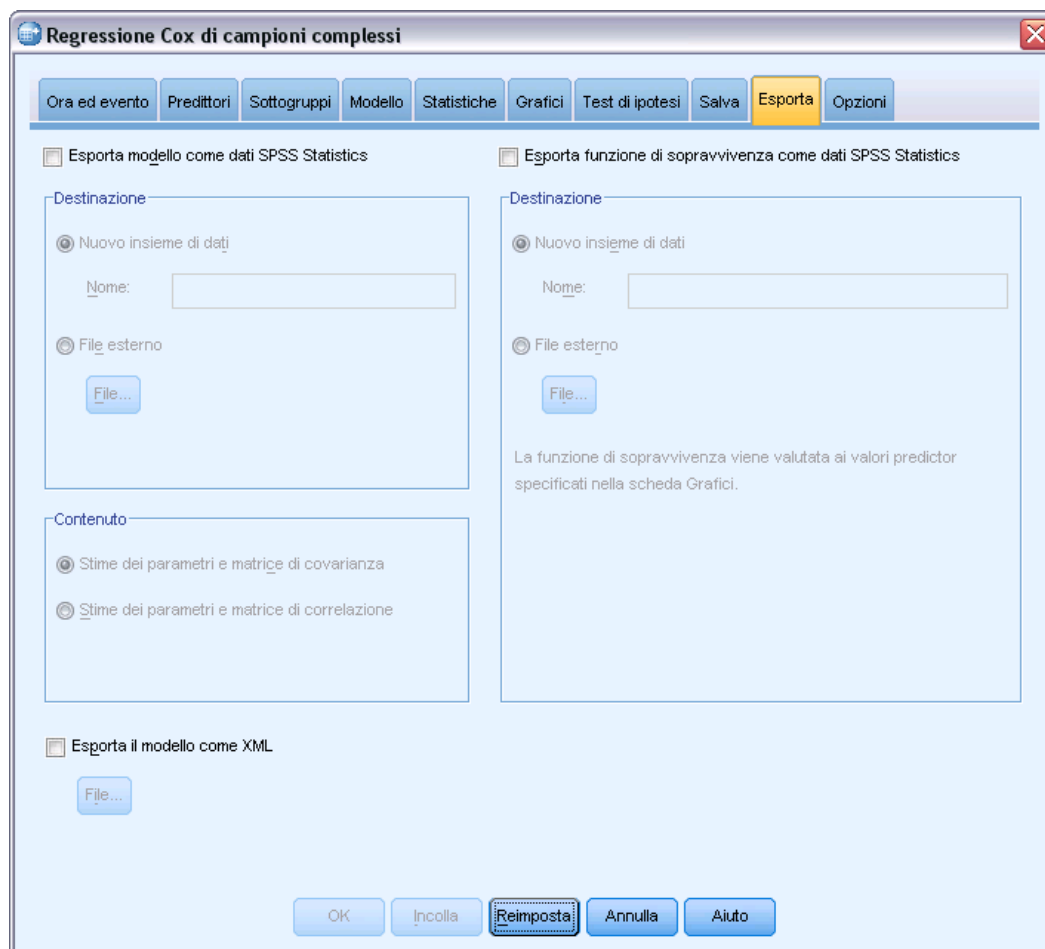
- **Correzione superiore dell'intervallo di confidenza per la funzione di sopravvivenza** Salva il limite superiore dell'intervallo di confidenza per la funzione di sopravvivenza al momento dell'osservazione e i valori del predittore per ogni caso.
- **Funzione di rischio cumulativo.** Salva il rischio cumulativo o $-\ln(\text{survival})$ al momento dell'osservazione e i valori del predittore per ogni caso.
- **Correzione inferiore dell'intervallo di confidenza per la funzione di rischio cumulativo.** Salva il limite inferiore dell'intervallo di confidenza per la funzione di rischio cumulativo al momento dell'osservazione e i valori del predittore per ogni caso.
- **Correzione superiore dell'intervallo di confidenza per la funzione di rischio cumulativo.** Salva il limite superiore dell'intervallo di confidenza per la funzione di rischio cumulativo al momento dell'osservazione e i valori del predittore per ogni caso.
- **Valore atteso del predittore lineare.** Salva la combinazione lineare dei predittori corretti del valore di riferimento per i coefficienti di regressione. Il predittore lineare è il rapporto della funzione di rischio con il rischio di base. Nel modello dei rischi proporzionali, questo valore è costante nel tempo.
- **Residuo Schoenfeld.** Per ogni caso non troncato e ogni parametro non ridondante nel modello, il residuo di Schoenfeld è la differenza tra il valore osservato del predittore associato al parametro del modello e il valore atteso del predittore per i casi nell'insieme di rischio all'ora dell'evento osservato. I residui di Schoenfeld possono essere utilizzati per valutare l'ipotesi di rischi proporzionali; ad esempio, per un predittore x , i grafici dei residui di Schoenfeld per il predittore dipendente dal tempo $x \cdot \ln(T_)$ rispetto all'ora devono mostrare una linea orizzontale in corrispondenza di 0 se i rischi proporzionali sono validi. Viene salvata una variabile separata per ogni parametro non ridondante nel modello. I residui di Schoenfeld vengono calcolati solo per i casi non troncati.
- **Residuo martingala.** Per ogni caso, il residuo martingala è la differenza tra il troncamento osservato (0 in caso di troncamento, 1 in caso contrario) e l'attesa di un evento durante il tempo di osservazione.
- **Residuo di devianza** I residui di devianza sono residui martingala "corretti" in modo da apparire più simmetrici rispetto allo 0. I grafici dei residui di devianza a fronte dei predittori non devono rivelare modelli.
- **Residuo Cox-Snell.** Per ogni caso, il residuo di Cox-Snell è l'attesa di un evento durante il tempo di osservazione o il troncamento osservato meno il residuo martingala.
- **Residuo punteggio.** Per ogni caso e ogni parametro non ridondante nel modello, il residuo di punteggio è il contributo del caso alla prima derivata della pseudo-probabilità. Viene salvata una variabile separata per ogni parametro non ridondante nel modello.
- **Residuo DFBeta.** Per ogni caso e ogni parametro non ridondante nel modello, il residuo DFBeta arrotonda la variazione di valore della stima del parametro quando il caso viene rimosso dal modello. I casi con residui DFBeta relativamente grandi possono esercitare un'influenza eccessiva sull'analisi. Viene salvata una variabile separata per ogni parametro non ridondante nel modello.
- **Residui aggregati.** Quando più casi rappresentano un singolo soggetto, il residuo aggregato per un soggetto è semplicemente la somma dei residui dei casi corrispondenti su tutti i casi appartenenti allo stesso soggetto. Per il residuo di Schoenfeld, la versione aggregata è la stessa della versione non aggregata poiché il residuo di Schoenfeld viene definito solo per i

casi non troncati. Questi residui sono disponibili solo se viene specificato un identificatore di soggetto nella scheda Ora ed evento.

Nomi delle variabili salvate. La generazione automatica del nome assicura il mantenimento di tutto il lavoro. I nomi personalizzati consentono di eliminare/sostituire i risultati di precedenti esecuzioni senza dover prima eliminare le variabili salvate nell'Editor dei dati.

Esporta

Figura 12-11
Finestra di dialogo Regressione di Cox, scheda Esporta



Esporta modello come dati SPSS Statistics. Genera un insieme di dati in formato IBM® SPSS® Statistics che contiene la correlazione dei parametri o la matrice di covarianza utilizzando le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file della matrice è come segue.

- **rowtype_.** Assume valori (ed etichette dei valori) COV (Covarianze), CORR (Correlazioni), EST (Stime dei parametri), SE (Errori standard), SIG (Livelli di significatività) e DF (Gradi di libertà del disegno di campionamento). Per ogni modello di parametro vi è un caso separato con tipo di riga COV (o CORR) e anche un caso separato per ciascuno degli altri tipi di riga.

- **varname_.** Assume i valori P1, P2, ..., che corrispondono a un elenco ordinato di tutti i parametri dei modelli per i tipi di riga COV o CORR, con etichette dei valori corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri. Le celle sono vuote per gli altri tipi di riga.
- **P1, P2, ...** Queste variabili corrispondono a un elenco ordinato di tutti i parametri dei modelli con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella relativa alle stime dei parametri e assumono i valori in base al tipo di riga. Per i parametri ridondanti, tutte le covarianze sono impostate su zero e le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate su zero e tutti gli errori standard, i livelli di significatività e i gradi di libertà dei residui sono impostati sul valore mancante di sistema.

Nota: questo file non può essere utilizzato immediatamente per ulteriori analisi in altre procedure che eseguono la lettura di un file di matrice a meno che queste procedure non accettino tutti i tipi di riga esportati qui.

Esporta funzione di sopravvivenza come dati SPSS Statistics. Scrive un insieme di dati in formato SPSS Statistics contenente la funzione di sopravvivenza, l'errore standard della funzione di sopravvivenza, i limiti superiore e inferiore dell'intervallo di confidenza della funzione di sopravvivenza e la funzione di rischio cumulativo per ogni ora dell'evento o errore, valutato alla linea base e ai modelli dei predittori specificati nella scheda Grafici. L'ordine delle variabili nel file della matrice è come segue.

- **Variabile strati di base.** Vengono generate tabelle di sopravvivenza separate per ogni valore della variabile di strati.
- **Variabile ora di sopravvivenza.** L'ora in cui si è verificato l'evento; per ogni ora univoca viene creato un caso separato.
- **Sur_0, LCL_Sur_0, UCL_Sur_0.** Funzione di sopravvivenza di base e limiti superiore e inferiore dell'intervallo di confidenza.
- **Sur_R, LCL_Sur_R, UCL_Sur_R.** Funzione di sopravvivenza valutata al modello di "riferimento" (vedere la tabella dei valori del modello nell'output) e limiti inferiore e superiore del relativo intervallo di confidenza.
- **Sur_##, LCL_Sur_##, UCL_Sur_##, ...** Funzione di sopravvivenza valutata a ognuno dei modelli del predittore specificati nella scheda Grafici e limiti inferiore e superiore dei relativi intervalli di confidenza. Vedere la tabella dei valori dei modelli nell'output per la corrispondenza con il numero ##.
- **Haz_0, LCL_Haz_0, UCL_Haz_0.** Funzione di rischio cumulativo di base e limiti superiore e inferiore del relativo intervallo di confidenza.
- **Haz_R, LCL_Haz_R, UCL_Haz_R.** Funzione di rischio cumulativo valutata al modello di "riferimento" (vedere la tabella dei valori dei modelli nell'output) e limiti inferiore e superiore del relativo intervallo di confidenza.
- **Haz_##, LCL_Haz_##, UCL_Haz_##, ...** Funzione di rischio cumulativo valutata a ognuno dei modelli del predittore specificati nella scheda Grafici e limiti superiore e inferiore dei relativi intervalli di confidenza. Vedere la tabella dei valori dei modelli nell'output per la corrispondenza con il numero ##.

Esporta modello come XML. Salva tutte le informazioni necessarie per prevedere la funzione di sopravvivenza, incluse le stime dei parametri e la funzione di sopravvivenza di base, nel formato XML (PMML). È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

Opzioni

Figura 12-12
Finestra di dialogo Regressione di Cox, scheda Opzioni

Stima. Questi controlli specificano i criteri per la stima dei coefficienti di regressione.

- **Massimo numero di iterazioni.** Il numero massimo di iterazioni che verranno eseguite nell'algoritmo. Specifica un intero non negativo.
- **Massimo numero di dimezzamenti.** Per ciascuna iterazione, la dimensione di passo è ridotta in base al fattore 0,5 finché la verosimiglianza non aumenta o non viene raggiunto il dimezzamento massimo. Specificare un intero positivo.
- **Limitare le iterazioni in base alla variazione nelle stime dei parametri.** Se l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nelle stime del parametro è minore del valore specificato, che deve essere positivo.

- **Limitare le iterazioni in base alla variazione nel log verosimiglianza.** Se l'opzione è selezionata, l'algoritmo si arresta dopo un'iterazione in cui la variazione assoluta o relativa nel log verosimiglianza è minore del valore specificato, che deve essere positivo.
- **Visualizza cronologia iterazioni.** Visualizza la cronologia delle iterazioni per le stime dei parametri e la pseudo-verosimiglianza; stampa l'ultima valutazione della variazione nelle stime dei parametri e nella pseudo-probabilità. La tabella della cronologia delle iterazioni stampa ogni n iterazioni a partire dall' 0 -esima iterazione (stime iniziali), dove n è il valore dell'incremento. Se la cronologia delle iterazioni è richiesta, l'ultima iterazione viene sempre visualizzata indipendentemente da n .
- **Metodo di interruzione dei casi pari merito per la stima dei parametri.** Se esistono ore pari merito di errore osservate, viene utilizzato uno di questi metodi per interrompere i casi pari merito. Il metodo Efron richiede un'elaborazione più impegnativa.

Funzioni di sopravvivenza. Questi controlli specificano i criteri relativi ai calcoli che interessano la funzione di sopravvivenza.

- **Metodo per stimare le funzioni di sopravvivenza di base.** Il metodo **Breslow** (o Nelson-Aalan o empirico) stima il rischio cumulativo di base tramite una funzione di passo non decrescente con passi alle ore di errore osservate, quindi calcola la sopravvivenza di base per la relazione $\text{sopravvivenza} = \exp(-\text{rischio cumulativo})$. Il metodo **Efron** richiede calcoli più impegnativi ed equivale al metodo Breslow quando non vi sono casi di pari merito. Il metodo **prodotto-limite** stima la sopravvivenza di base tramite una funzione continua a destra non crescente; quando non vi sono predittori nel modello, questo metodo equivale alla stima Kaplan-Meier.
- **Intervalli di confidenza delle funzioni di sopravvivenza.** L'intervallo di confidenza può essere calcolato in tre modi: in unità originali, mediante una trasformazione logaritmica o una trasformazione log meno log. Solo la trasformazione log meno log garantisce che i limiti dell'intervallo di confidenza saranno compresi tra 0 e 1, ma la trasformazione logaritmica sembra garantire risultati "migliori".

Valori mancanti definiti dall'utente. Tutte le variabili devono contenere valori validi per un caso per essere incluse nell'analisi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente vengono considerati come validi nei modelli categoriali (inclusi fattori, variabili di evento, di stratificazione e di sottopopolazione).

Intervallo di confidenza (%). Livello dell'intervallo di confidenza utilizzato per le stime dei coefficienti, dei coefficienti elevate a potenza, della funzione di sopravvivenza e della funzione di rischio cumulativo. Specificare un valore maggiore o uguale a 0 e minore di 100.

Opzioni aggiuntive del comando CSCOXREG

Il linguaggio a comandi consente inoltre di:

- Eseguire test di ipotesi personalizzati (utilizzando il sottocomando `CUSTOM` e `/PRINT LMATRIX`).
- Specifica della tolleranza (tramite `/CRITERIA SINGULAR`).
- Tabella della funzione stimabile generalizzata (tramite `/PRINT GEF`).

- Più modelli di predittore (con più sottocomandi `PATTERN`).
- Numero massimo di variabili salvate quando viene specificato un nome radice (tramite il sottocomando `SAVE`). La finestra di dialogo rispetta il valore predefinito `CSCOXREG` di 25 variabili.

Vedere *Command Syntax Reference* per informazioni dettagliate sulla sintassi.

Parte II: Esempi

Campionamento guidato di Campioni complessi

Campionamento guidato consente di eseguire le varie fasi di creazione, modifica o esecuzione di un file del piano di campionamento. Per usare la procedura guidata, è necessario definire correttamente una popolazione di destinazione, disporre di un elenco di unità di campionamento e avere in mente un disegno di campionamento.

Ottenere un campione da una struttura di campionamento completa

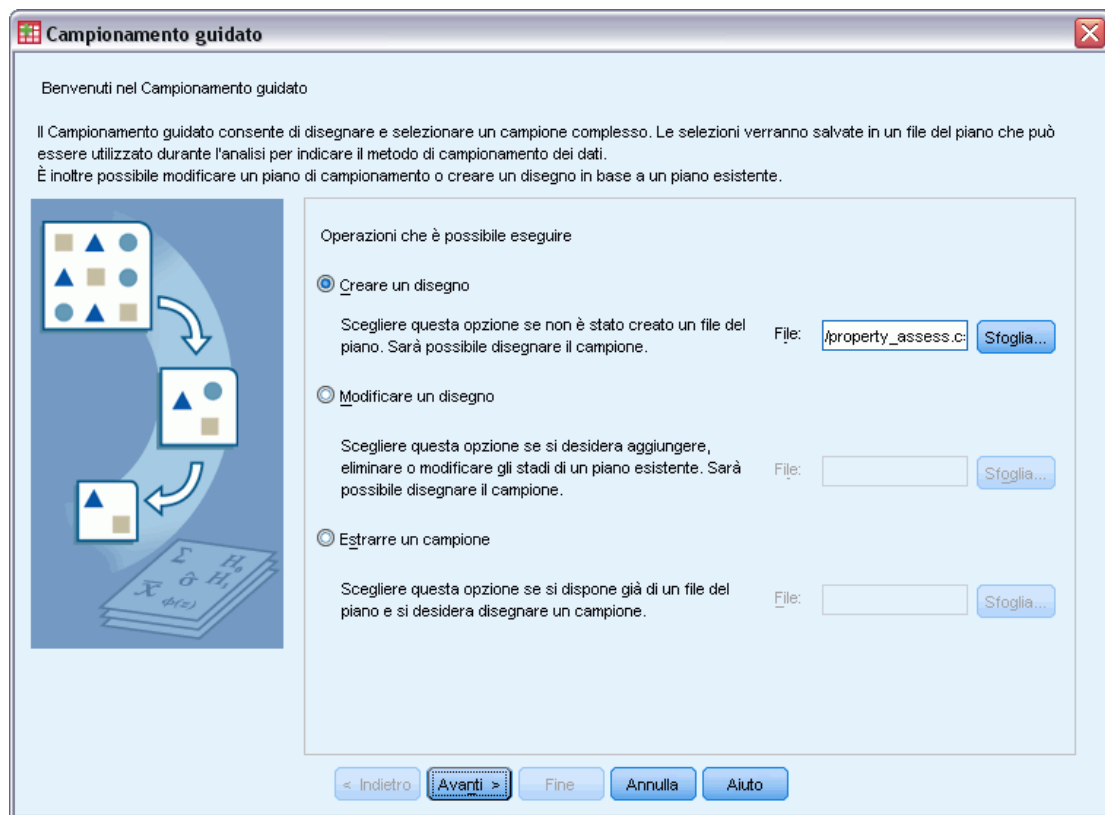
Un'agenzia statale deve assicurare imposte sulla proprietà corrette nelle diverse zone. Le imposte si basano sul valore stimato della proprietà, quindi l'agenzia desidera realizzare un sondaggio su un campione di proprietà per zona per essere certa che i record di ogni zona siano aggiornati allo stesso modo. Tuttavia, le risorse che consentono di ottenere le attuali stime sono limitate, quindi è importante utilizzare ciò che è disponibile. L'agenzia decide di utilizzare una metodologia di campionamento complesso per selezionare un campione di proprietà.

L'elenco delle proprietà si trova nel file *property_assess_cs.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Utilizzare Campionamento guidato di Campioni complessi per selezionare un campione.

Utilizzo della procedura guidata

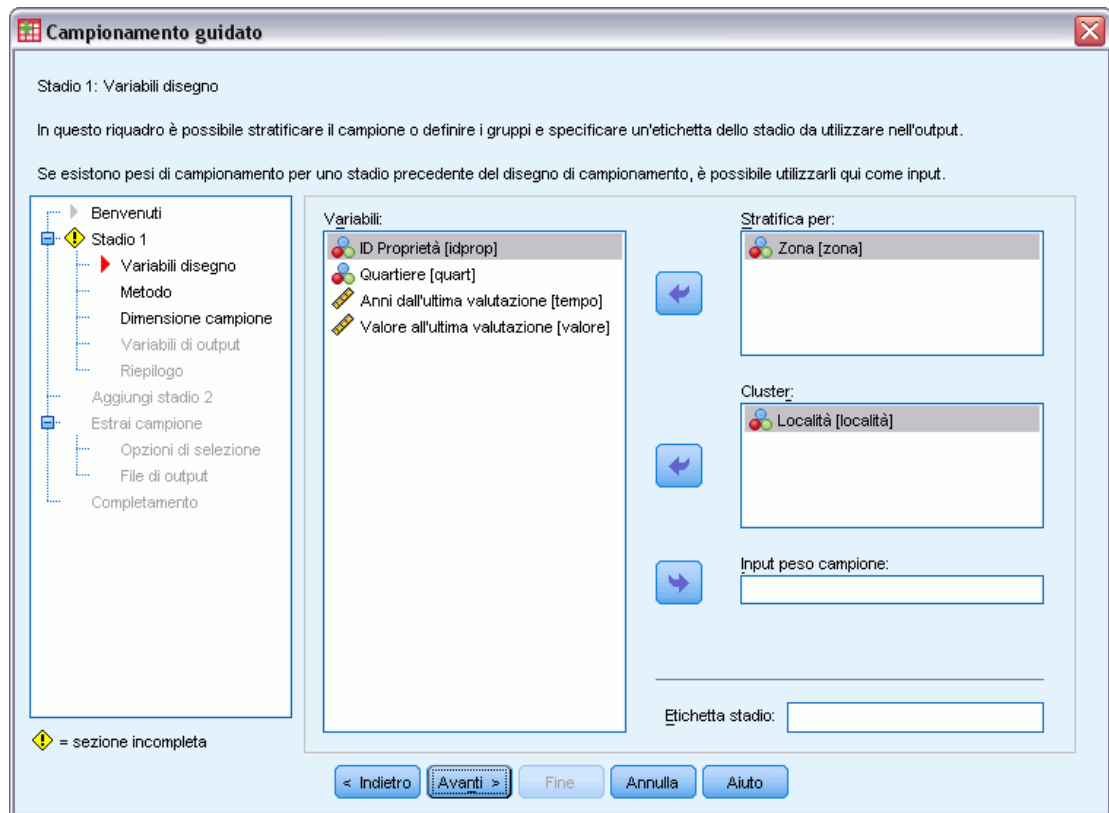
- ▶ Per eseguire Campionamento guidato di Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Seleziona un campione...

Figura 13-1
Campionamento guidato, Finestra di benvenuto



- ▶ Selezionare Creare un disegno, individuare la posizione in cui si desidera salvare il file e digitare property_assess.csplan come nome del file del piano.
- ▶ Fare clic su Avanti.

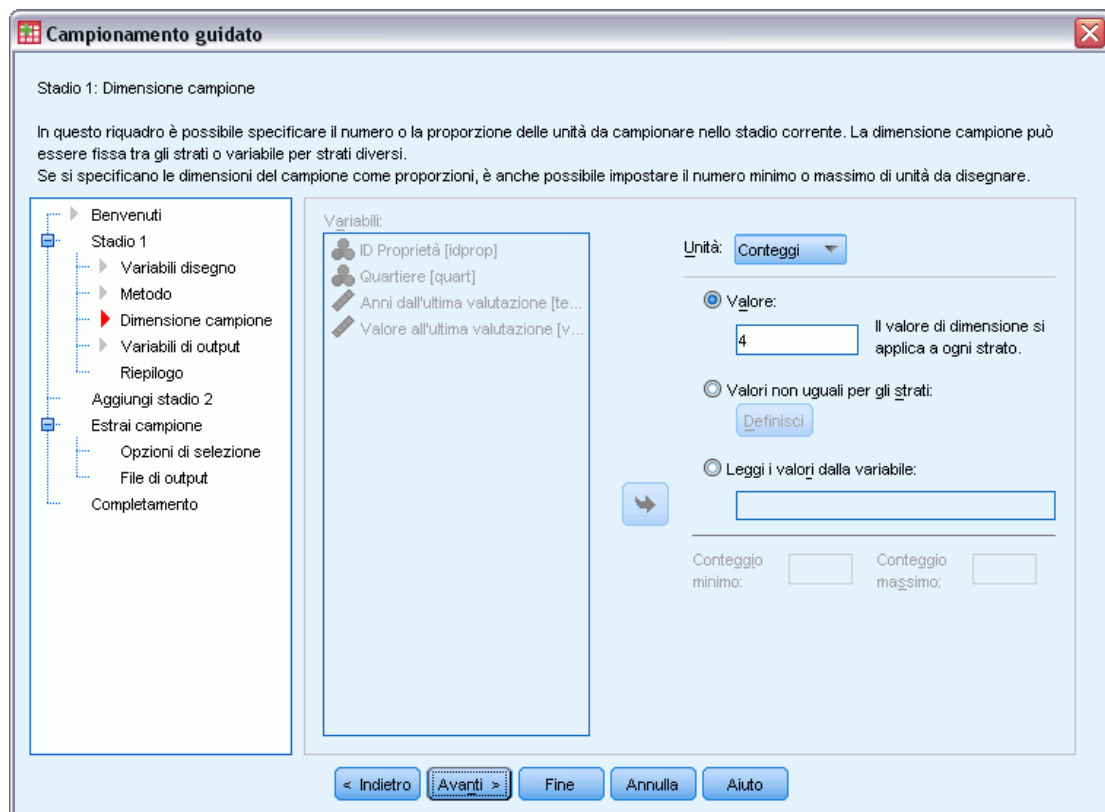
Figura 13-2
Campionamento guidato, Passo Variabili disegno (stadio 1)



- ▶ Selezionare *Zona* come variabile di stratificazione.
- ▶ Selezionare *Area* come variabile di raggruppamento.
- ▶ Fare clic su *Avanti*, quindi ancora su *Avanti* nel passo *Metodo* di campionamento.

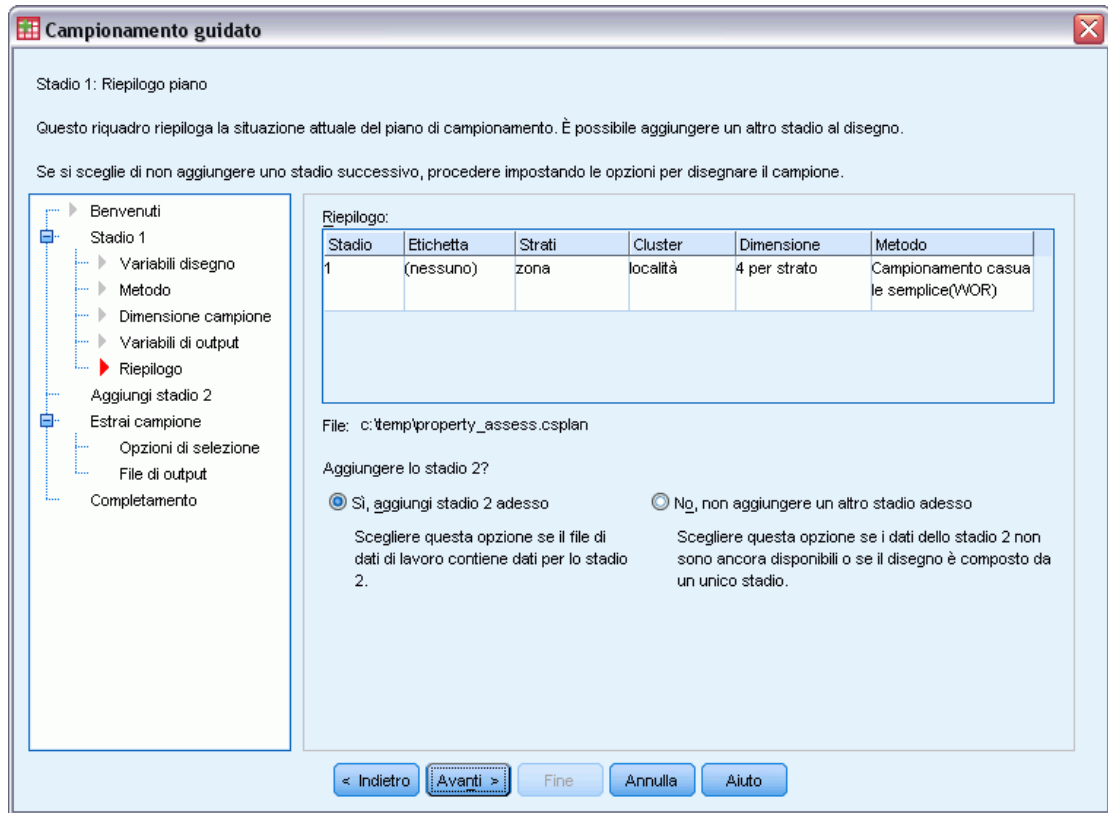
Tale struttura di disegno comporta l'estrazione di campioni indipendenti da ogni zona. In questo stadio, le aree vengono estratte come unità di campionamento principali utilizzando il metodo predefinito, Campionamento casuale semplice.

Figura 13-3
Campionamento guidato, Passo Dimensione campione (stadio 1)



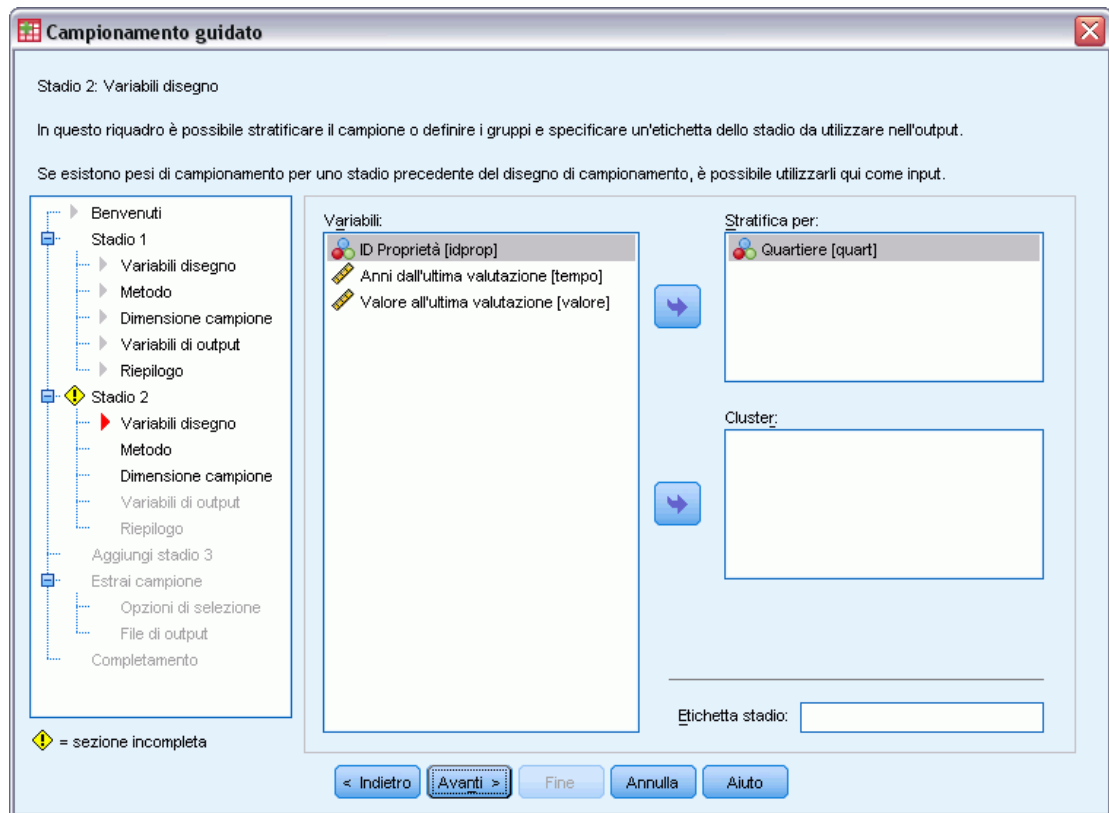
- ▶ Selezionare Conteggi dall'elenco a discesa Unità.
- ▶ Digitare 4 come valore per il numero di unità da selezionare in questo stadio.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-4
Campionamento guidato, Passo Riepilogo del piano (stadio 1)



- ▶ Selezionare Sì, aggiungere lo stadio 2 ora.
- ▶ Fare clic su Avanti.

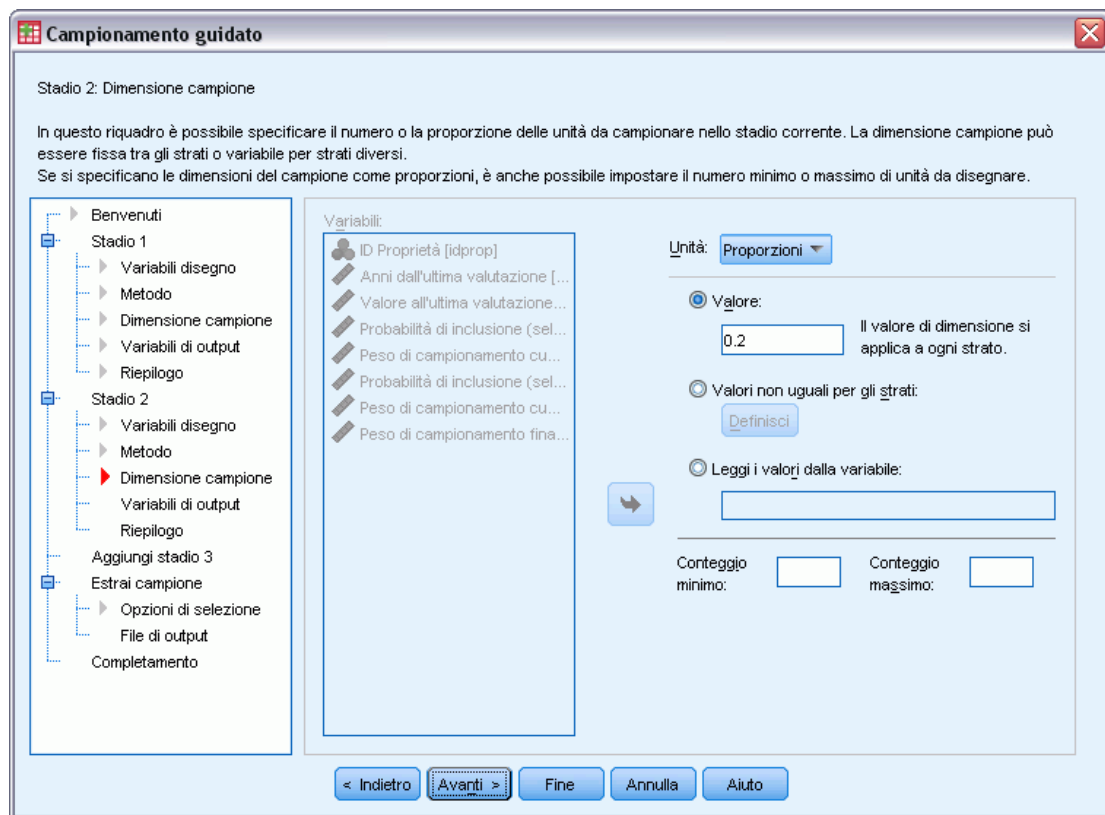
Figura 13-5
Campionamento guidato, Passo Variabili disegno (stadio 2)



- ▶ Selezionare *Quartiere* come variabile di stratificazione.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Metodo di campionamento.

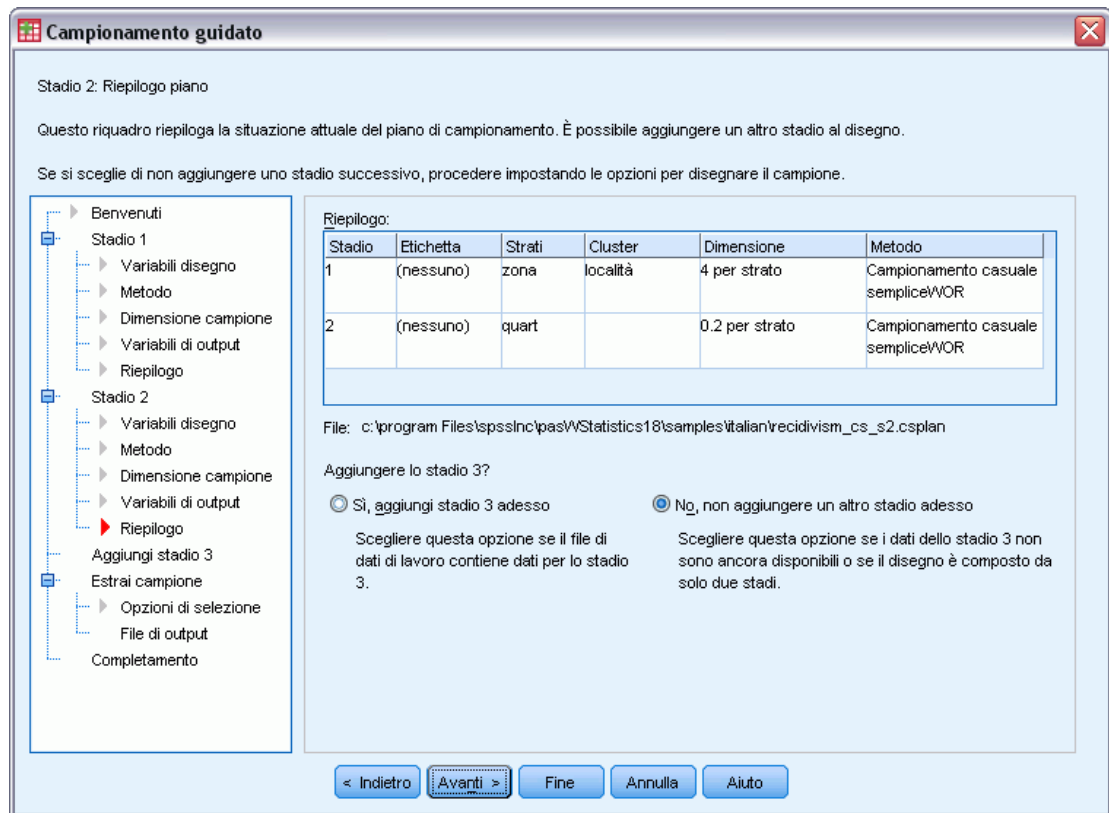
Tale struttura di disegno comporta l'estrazione di campioni indipendenti per ogni quartiere delle aree estratte nella fase 1. In questo stadio, le proprietà vengono estratte come unità di campionamento principali utilizzando il metodo predefinito, Campionamento casuale semplice.

Figura 13-6
Campionamento guidato, Passo Dimensione campione (stadio 2)



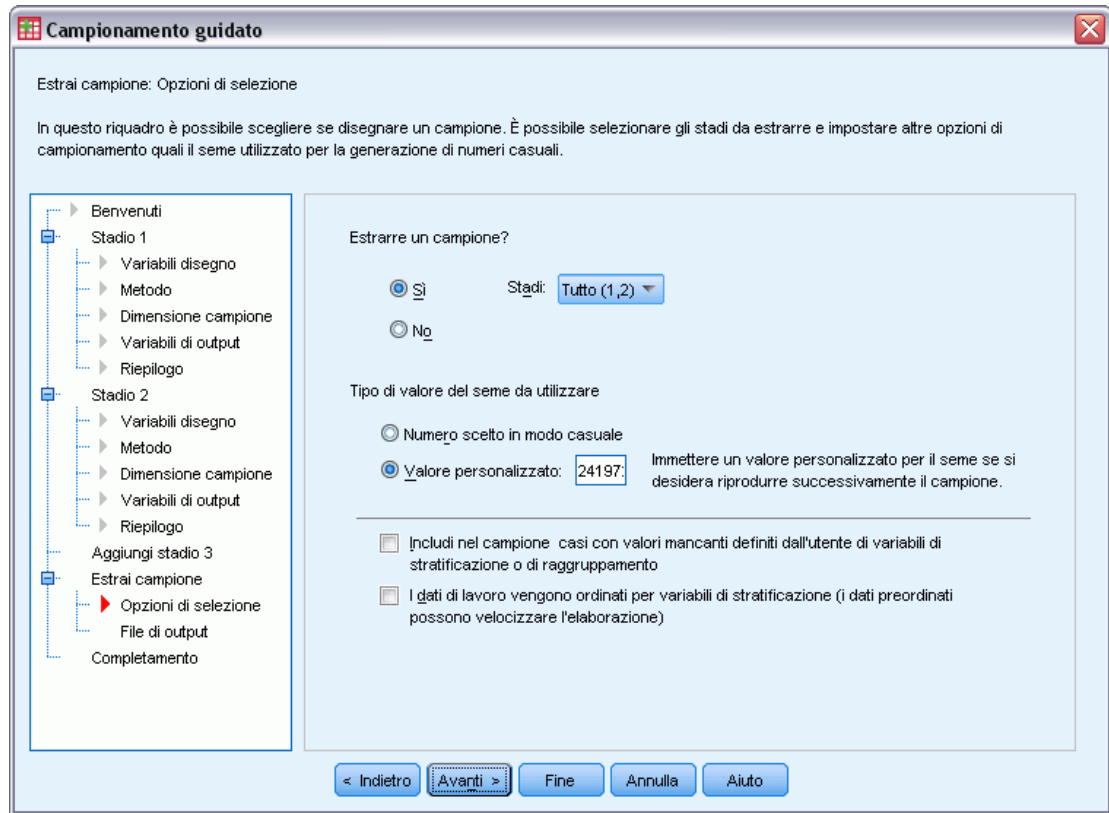
- ▶ Selezionare Proporzioni dall'elenco a discesa Unità.
- ▶ Digitare 0,2 come valore della proporzione di unità da campionare da ogni strato.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-7
Campionamento guidato, Passo Riepilogo del piano (stadio 2)



- Controllare il disegno di campionamento, quindi fare clic su Avanti.

Figura 13-8
Campionamento guidato, Passo Estrai campione - Opzioni di selezione

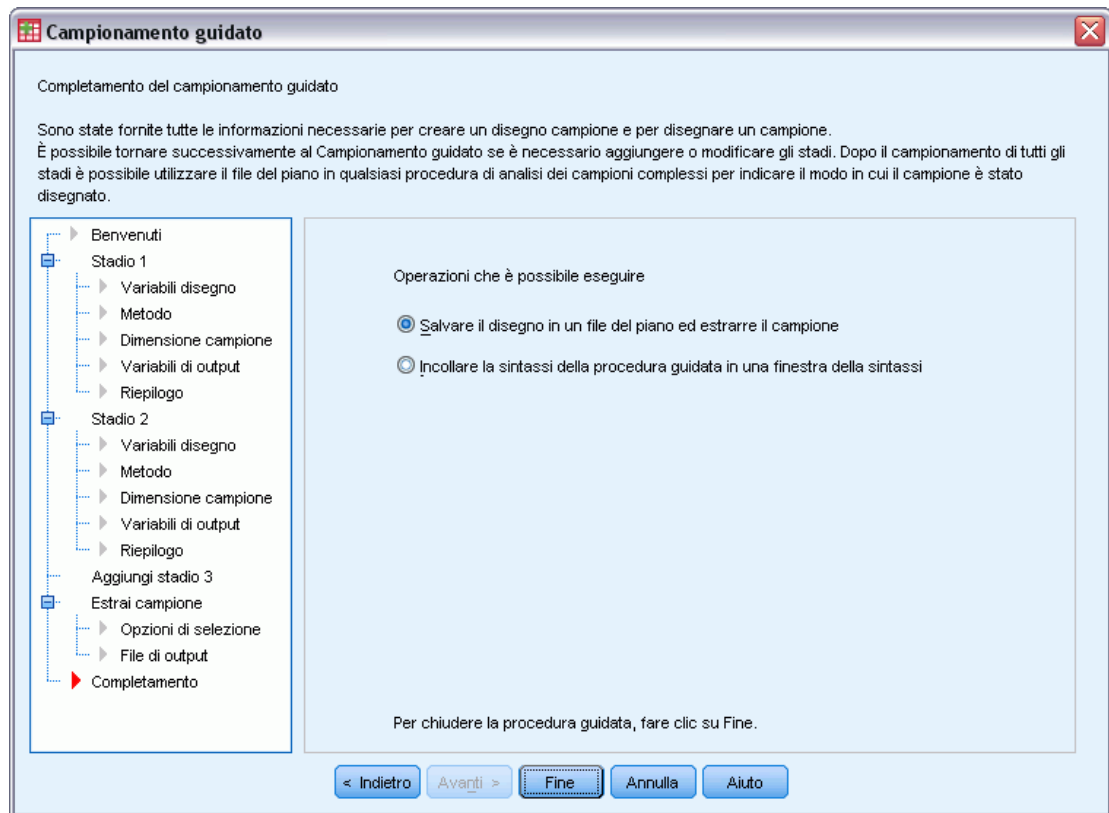


- Selezionare Valore personalizzato per il tipo di seme casuale da utilizzare e digitare 241972 come valore.

L'utilizzo di un valore personalizzato consente di replicare esattamente i risultati dell'esempio.

- Fare clic su Avanti, quindi ancora su Avanti nel passo Estrai campione - Opzioni di selezione.

Figura 13-9
Campionamento guidato, Passo Fine



- Fare clic su Fine.

Tali selezioni consentono di generare il file del piano di campionamento *property_assess.csplan* e di estrarre un campione in base a tale piano.

Riepilogo del piano

Figura 13-10
Riepilogo del piano

			Stadio 1	Stadio 2
Variabili disegno	Stratificazione	1	Zona	Quartiere
	Raggruppamento	1	Località	
Informazioni sul campione	Metodo di selezione		Campionamento casuale semplice senza sostituzione	Campionamento casuale semplice senza sostituzione
	Numero di unità campionate		4	
	Variabili create o modificate	Probabilità di inclusione (selezione) a stadi	Inclusion Probability_1_	Inclusion Probability_2_
		Peso di campionamento cumulato a stadi	Sample Weight Cumulative_1_	Sample Weight Cumulative_2_
	Proporzione di unità campionate		.2	
Informazioni sull'analisi	Ipotesi stimatore		Campionamento con probabilità uguale senza sostituzione	Campionamento con probabilità uguale senza sostituzione
	Probabilità di inclusione		Ottenuto dalla variabile Inclusion Probability_1_	Ottenuto dalla variabile Inclusion Probability_2_

File del piano: C:\property_assess.csplan
Variabile peso: SampleWeight_Final_

La tabella riassuntiva consente di verificare il piano di campionamento ed è utile per assicurarsi che il piano rappresenti le intenzioni dell'utente.

Riepilogo del campionamento

Figura 13-11
Riepilogo stadio

Zona	Numero di unità campionate		Proporzione di unità campionate	
	Requested	Actual	Requested	Actual
Orientale	4	4	44.4%	44.4%
Centrale	4	4	57.1%	57.1%
Occidentale	4	4	25.0%	25.0%
Settentrionale	4	4	44.4%	44.4%
Meridionale	4	4	50.0%	50.0%

File del piano: \property_assess.csplan

La tabella riassuntiva consente di verificare il primo stadio di campionamento ed è utile per controllare che il campionamento sia stato eseguito in base al piano. Da ogni zona sono state estratte quattro aree, come richiesto.

Figura 13-12
Riepilogo stadio

Zona	Località	Quartiere	Numero di unità campionate		Proporzione di unità campionate	
			Requested	Actual	Requested	Actual
Orientale	2	8	4	4	20,0%	19,0%
		9	14	14	20,0%	20,6%
		10	7	7	20,0%	18,9%
		11	14	14	20,0%	20,0%
	6	36	13	13	20,0%	20,3%
		37	14	14	20,0%	20,6%
		38	13	13	20,0%	20,6%
	7	43	12	12	20,0%	20,7%
		44	11	11	20,0%	19,6%
		45	11	11	20,0%	20,8%
		46	13	13	20,0%	20,0%
	9	57	13	13	20,0%	20,6%
		58	5	5	20,0%	18,5%
		59	11	11	20,0%	19,3%
60		13	13	20,0%	19,4%	
Centrale	22	148	9	9	20,0%	19,6%
		149	8	8	20,0%	20,0%

La tabella riassuntiva, di cui viene visualizzata la parte superiore, consente di verificare il secondo stadio di campionamento. È inoltre utile per controllare che il campionamento sia stato eseguito in base al piano. Da ogni quartiere di ogni area campionata nel primo stadio è stato estratto approssimativamente il 20% delle proprietà, come richiesto.

Risultati del campione

Figura 13-13
Editor dei dati con i risultati del campione.

	idprop	quart	località	zona	tempo	valore	InclusionPr obability_1	SampleWei ghtCumulat ve_1	InclusionPr obability_2	SampleWei ghtCumulat ve_2	SampleWei ght_Final
289 :											
273	577,00	8	2	1	4	181,70
274	578,00	8	2	1	5	189,60
275	579,00	8	2	1	4	200,10
276	580,00	8	2	1	5	211,50
277	581,00	8	2	1	4	181,50
278	641,00	9	2	1	7	192,40
279	642,00	9	2	1	6	236,70	,44	2,25	,21	10,93	10,93
280	643,00	9	2	1	6	150,40	,44	2,25	,21	10,93	10,93
281	644,00	9	2	1	8	204,80
282	645,00	9	2	1	6	225,40
283	646,00	9	2	1	7	180,80	,44	2,25	,21	10,93	10,93
284	647,00	9	2	1	5	176,90

Visualizzazione dati Visualizzazione variabili

Vengono visualizzati i risultati del campionamento nell'Editor dei dati. Nel file di lavoro sono state salvate cinque nuove variabili che rappresentano le probabilità di inclusione e i pesi di campionamento cumulato per ogni stadio, nonché i pesi di campionamento finali.

- I casi con valori per queste variabili sono stati selezionati per il campione.
- Non sono stati selezionati i casi con valori mancanti di sistema per le variabili.

L'agenzia ricorrerà alle proprie risorse per raccogliere le valutazioni attuali relative alle proprietà selezionate nel campione. Quando tali valutazioni sono disponibili, è possibile elaborare il campione con le procedure di analisi di Campioni complessi, utilizzando il piano di campionamento *property_assess.csplan* per fornire le specifiche di campionamento.

Ottenere un campione da una struttura di campionamento parziale

Una società è interessata alla compilazione e alla vendita di un database di informazioni ricavate da sondaggi di alta qualità. Il campione del sondaggio deve essere rappresentativo ma realizzato con efficienza. Pertanto verranno utilizzati metodi di campionamento complessi. Il disegno di campionamento completo richiede la struttura seguente:

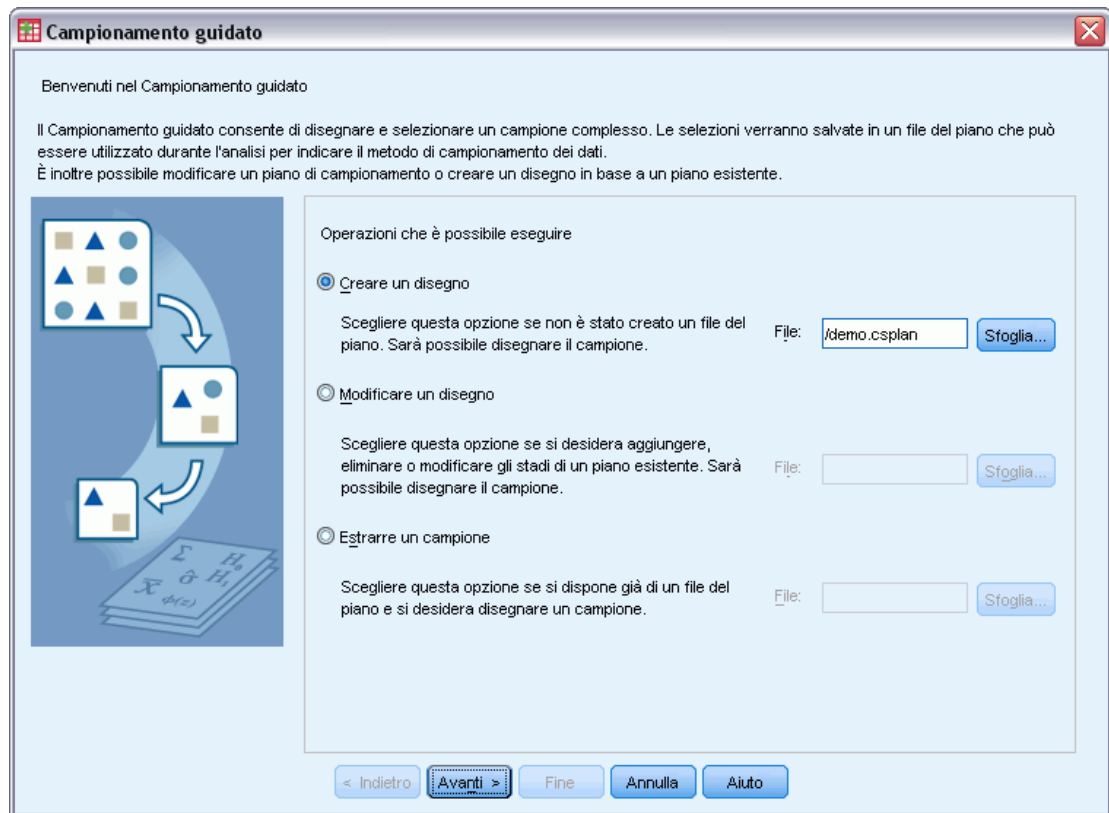
Stadio	Strati	Raggrup- pamenti
1	Regione	Provincia
2	Distretto	Città
3	Suddivisione	

Nel terzo stadio, le abitazioni rappresentano le unità di campionamento principali e pertanto verrà esaminata una selezione di abitazioni. Tuttavia, poiché le informazioni sono facilmente accessibili solo a livello di città, la società intende eseguire i primi due stadi del disegno adesso e raccogliere in un secondo momento il numero di suddivisioni e abitazioni dalle città campionate. Le informazioni disponibili a livello di città sono contenute nel file *demo_cs_1.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Notare che il file contiene la variabile *Suddivisione* contenente una serie di 1. Si tratta del segnaposto della variabile effettiva, i cui valori verranno raccolti dopo l'esecuzione dei primi due stadi del disegno e consente di specificare in questo momento il disegno di campionamento in tre stadi. Utilizzare Campionamento guidato di Campioni complessi per specificare il disegno di campionamento complesso completo e per estrarre i primi due stadi.

Utilizzo della procedura guidata per campionare dalla prima struttura parziale

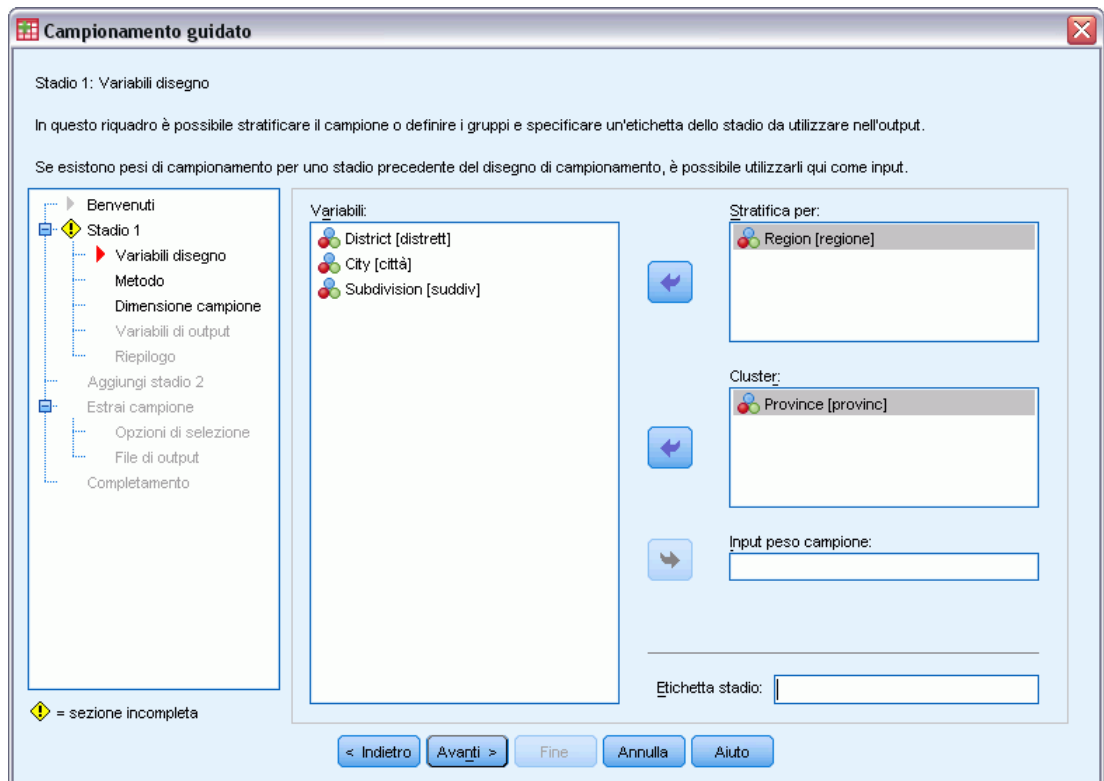
- Per eseguire Campionamento guidato di Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Seleziona un campione...

Figura 13-14
Campionamento guidato, Finestra di benvenuto



- ▶ Selezionare Creare un disegno, individuare la posizione in cui si desidera salvare il file e digitare demo.csplan come nome del file del piano.
- ▶ Fare clic su Avanti.

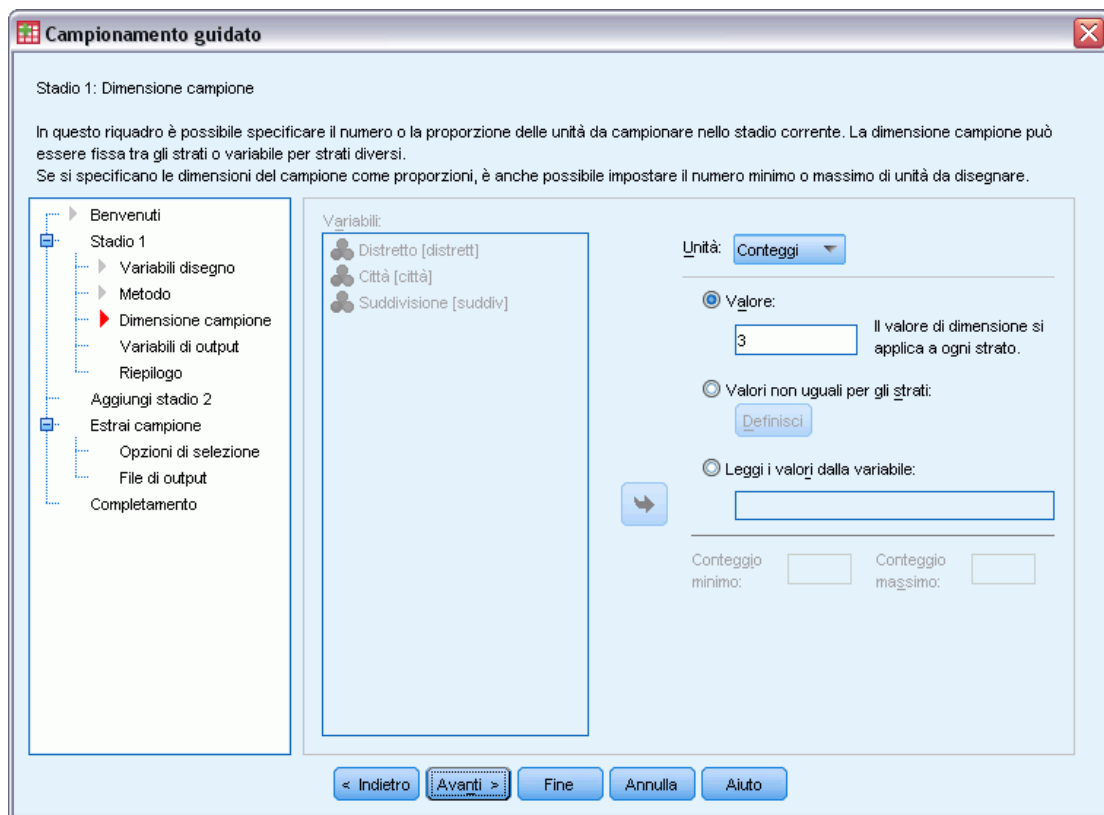
Figura 13-15
 Campionamento guidato, Passo Variabili disegno (stadio 1)



- ▶ Selezionare *Regione* come variabile di stratificazione.
- ▶ Selezionare *Provincia* come variabile di raggruppamento.
- ▶ Fare clic su *Avanti*, quindi ancora su *Avanti* nel passo *Metodo* di campionamento.

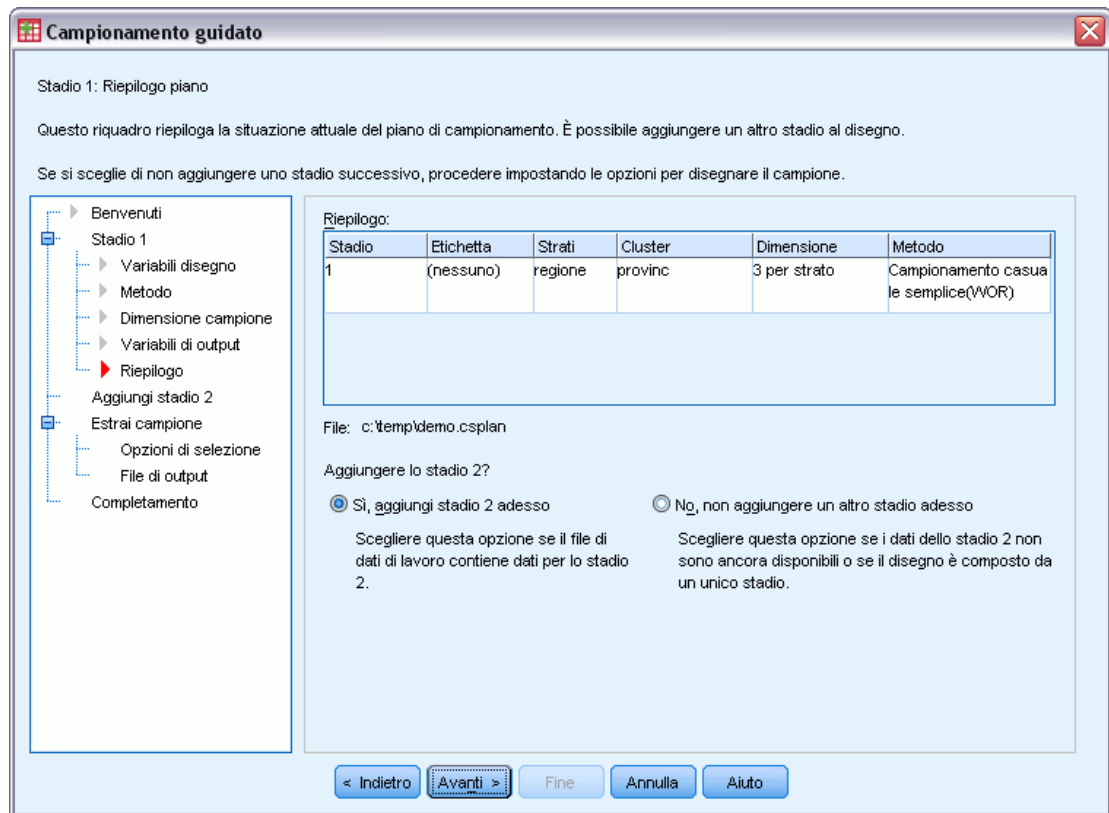
Tale struttura di disegno comporta l'estrazione di campioni indipendenti da ogni regione. In questo stadio, le province vengono estratte come unità di campionamento principali utilizzando il metodo predefinito, Campionamento casuale semplice.

Figura 13-16
Campionamento guidato, Passo Dimensione campione (stadio 1)



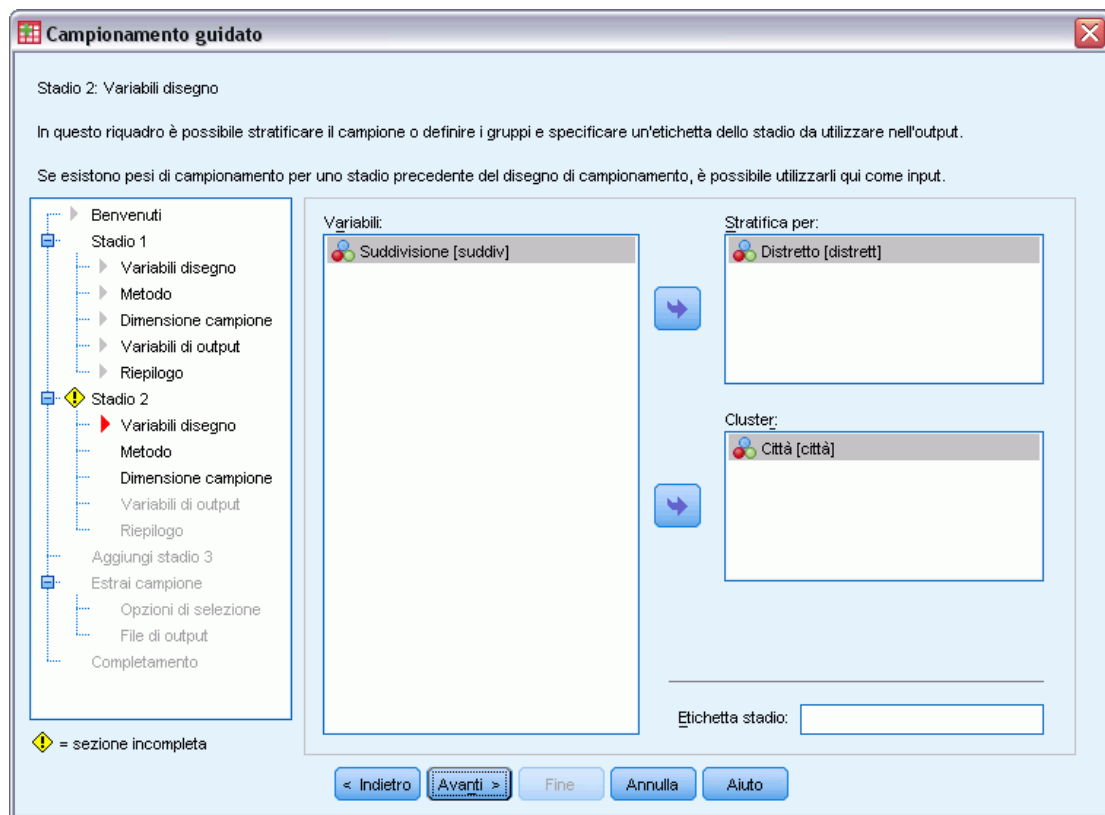
- ▶ Selezionare Conteggi dall'elenco a discesa Unità.
- ▶ Digitare 3 come valore per il numero di unità da selezionare in questo stadio.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-17
Campionamento guidato, Passo Riepilogo del piano (stadio 1)



- ▶ Selezionare Sì, aggiungere lo stadio 2 ora.
- ▶ Fare clic su Avanti.

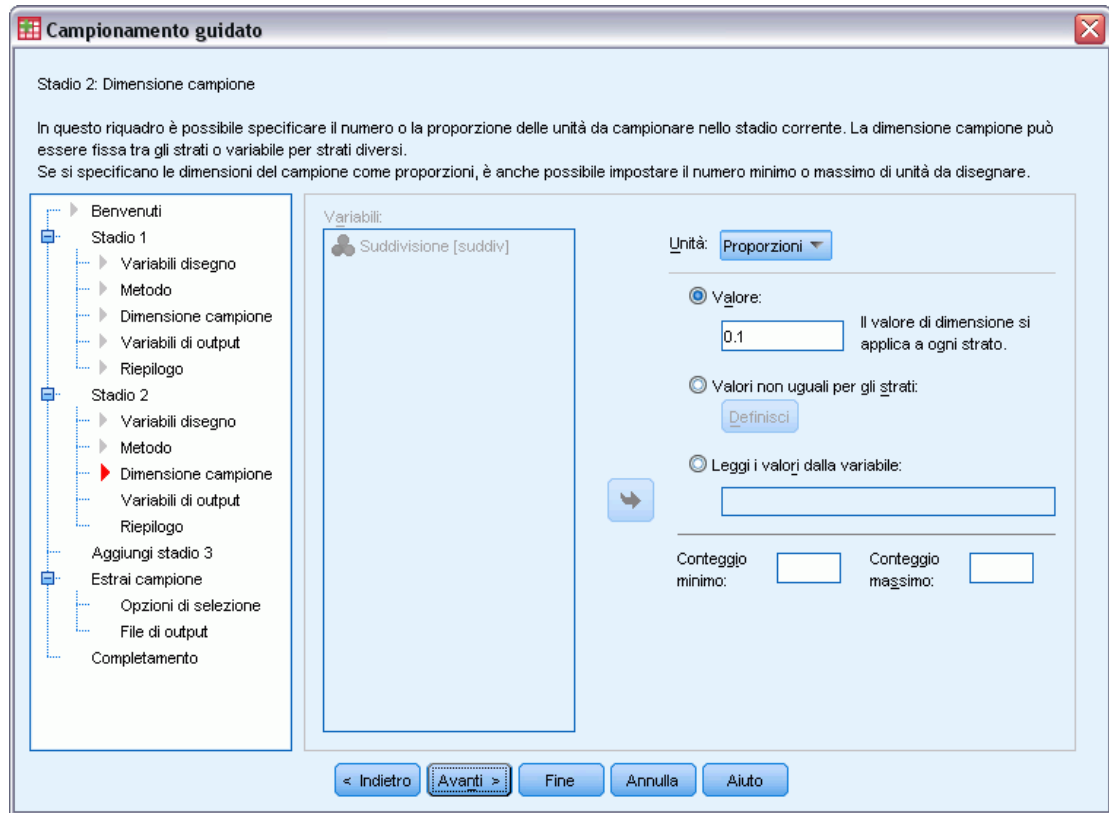
Figura 13-18
Campionamento guidato, Passo Variabili disegno (stadio 2)



- ▶ Selezionare *Distretto* come variabile di stratificazione.
- ▶ Selezionare *Città* come variabile di raggruppamento.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Metodo di campionamento.

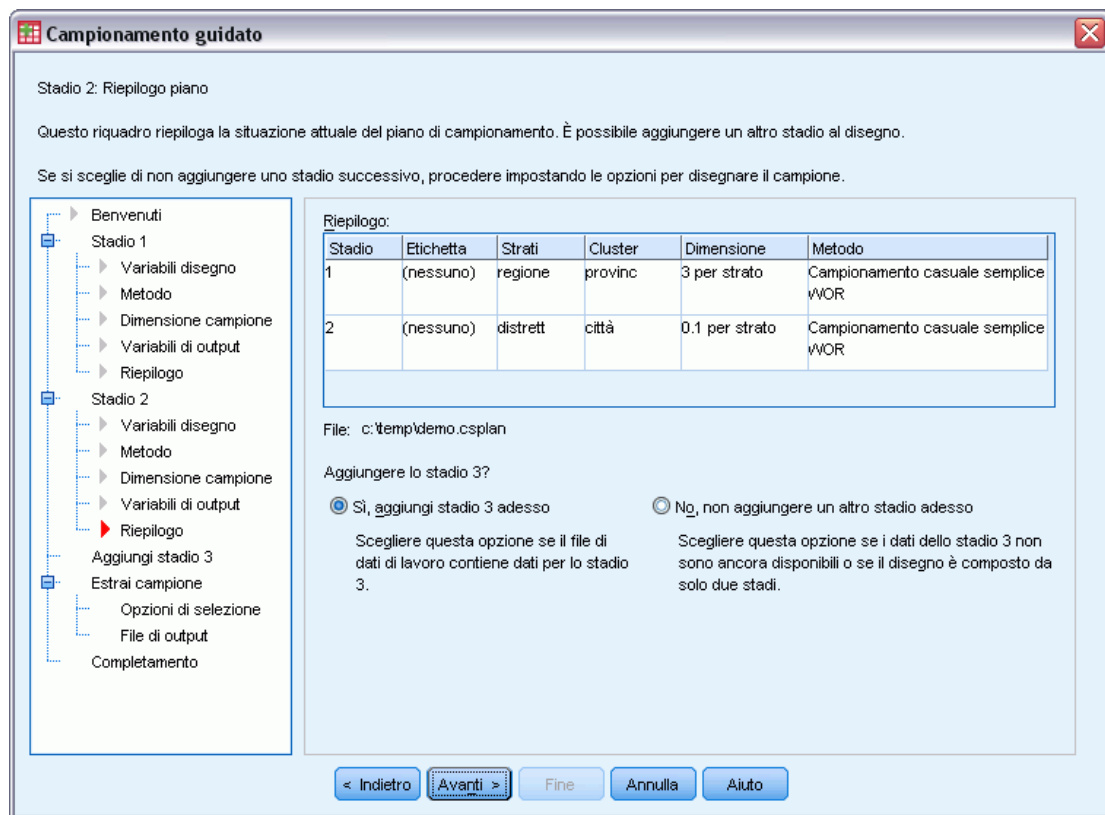
Tale struttura di disegno comporta l'estrazione di campioni indipendenti da ogni distretto. In questo stadio, le città vengono estratte come unità di campionamento principali utilizzando il metodo predefinito, Campionamento casuale semplice.

Figura 13-19
 Campionamento guidato, Passo Dimensione campione (stadio 2)



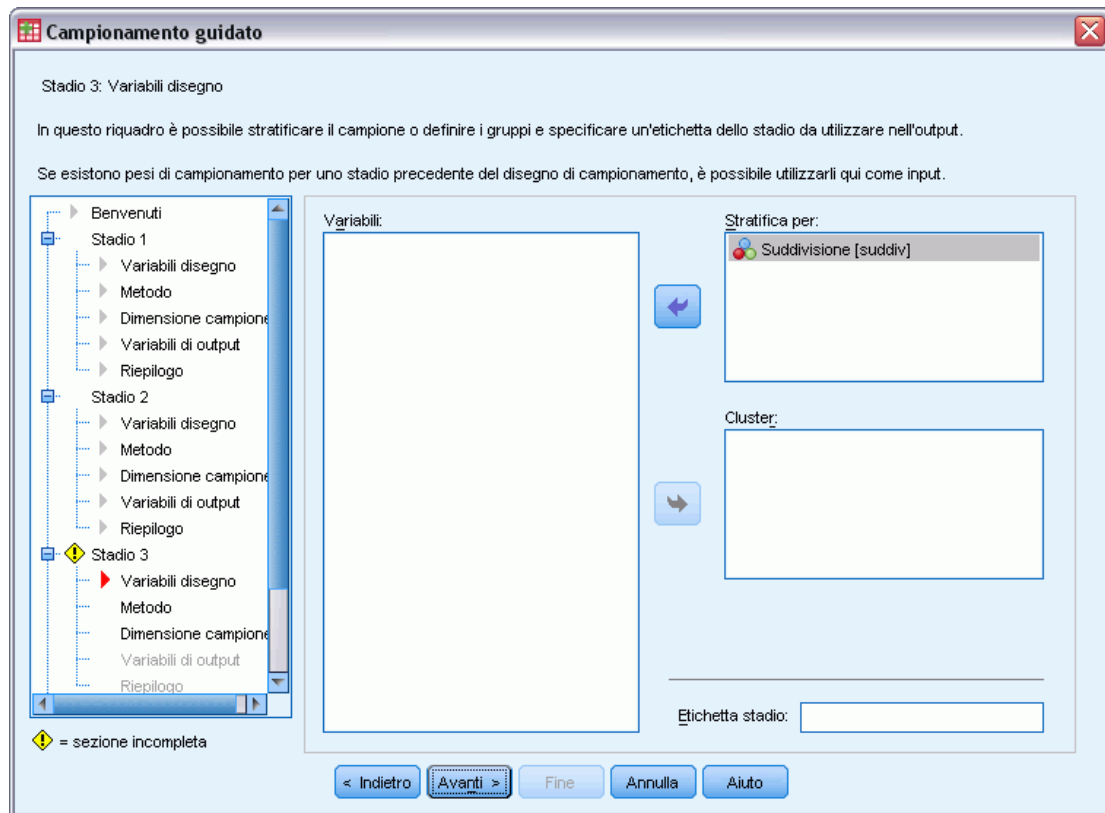
- ▶ Selezionare Proporzioni dall'elenco a discesa Unità.
- ▶ Digitare 0,1 come valore della proporzione di unità da campionare da ogni strato.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-20
Campionamento guidato, Passo Riepilogo del piano (stadio 2)



- Selezionare Sì, aggiungere lo stadio 3 ora.
- Fare clic su Avanti.

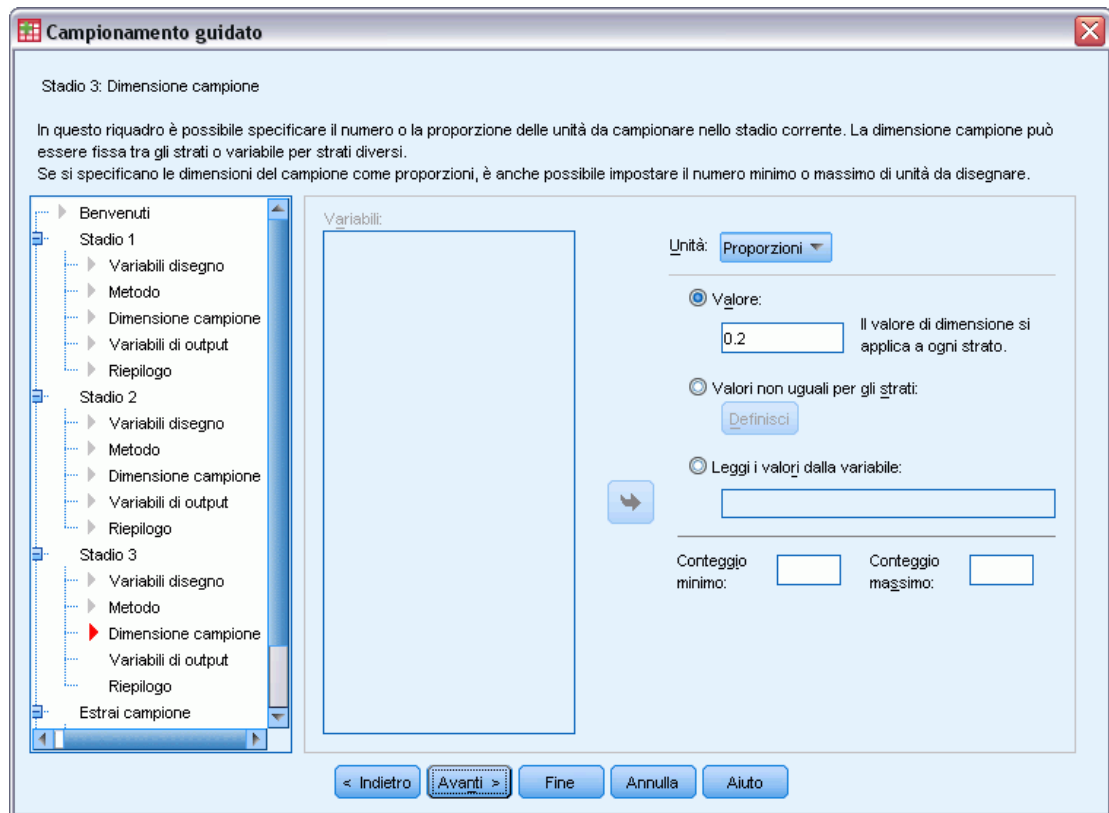
Figura 13-21
Campionamento guidato, Passo Variabili disegno (stadio 3)



- ▶ Selezionare *Suddivisione* come variabile di stratificazione.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Metodo di campionamento.

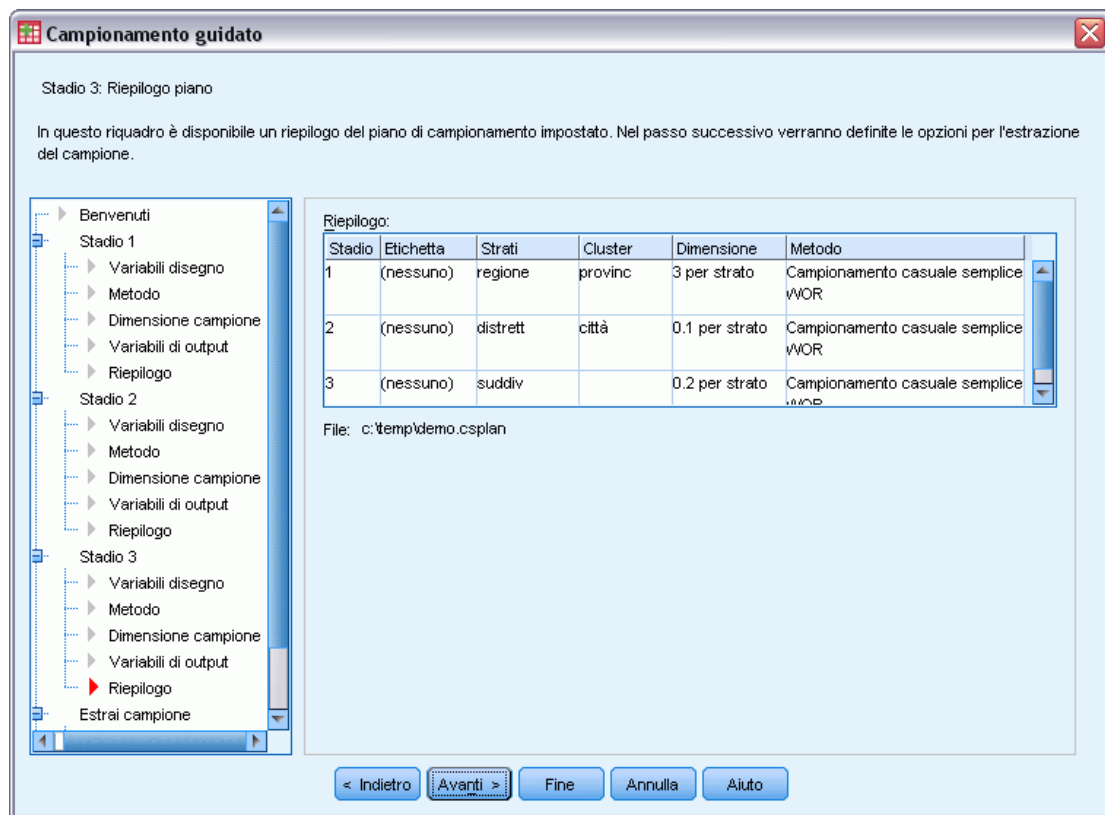
Tale struttura di disegno comporta l'estrazione di campioni indipendenti da ogni suddivisione. In questo stadio, le unità di abitazione vengono estratte come unità di campionamento principali utilizzando il metodo predefinito, Campionamento casuale semplice.

Figura 13-22
Campionamento guidato, Passo Dimensione campione (stadio 3)



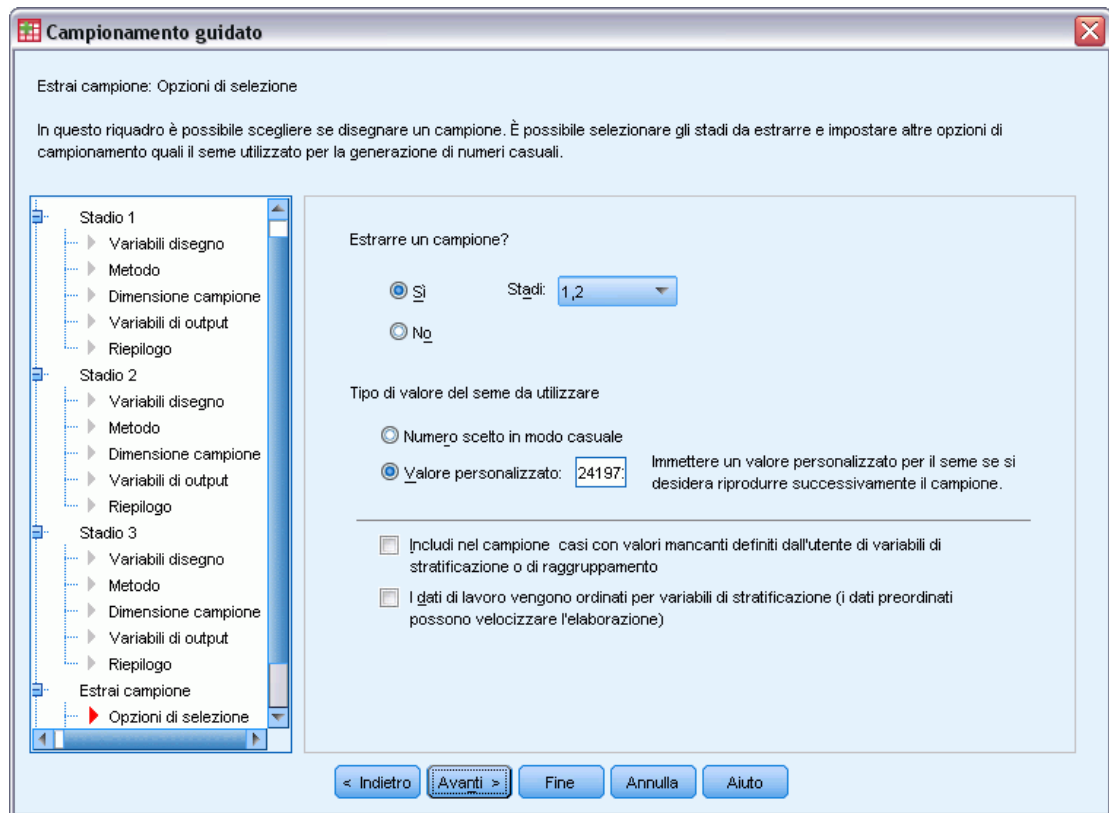
- ▶ Selezionare Proporzioni dall'elenco a discesa Unità.
- ▶ Digitare 0,2 come valore per la proporzione di unità da selezionare in questo stadio.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-23
Campionamento guidato, Passo Riepilogo del piano (stadio 3)



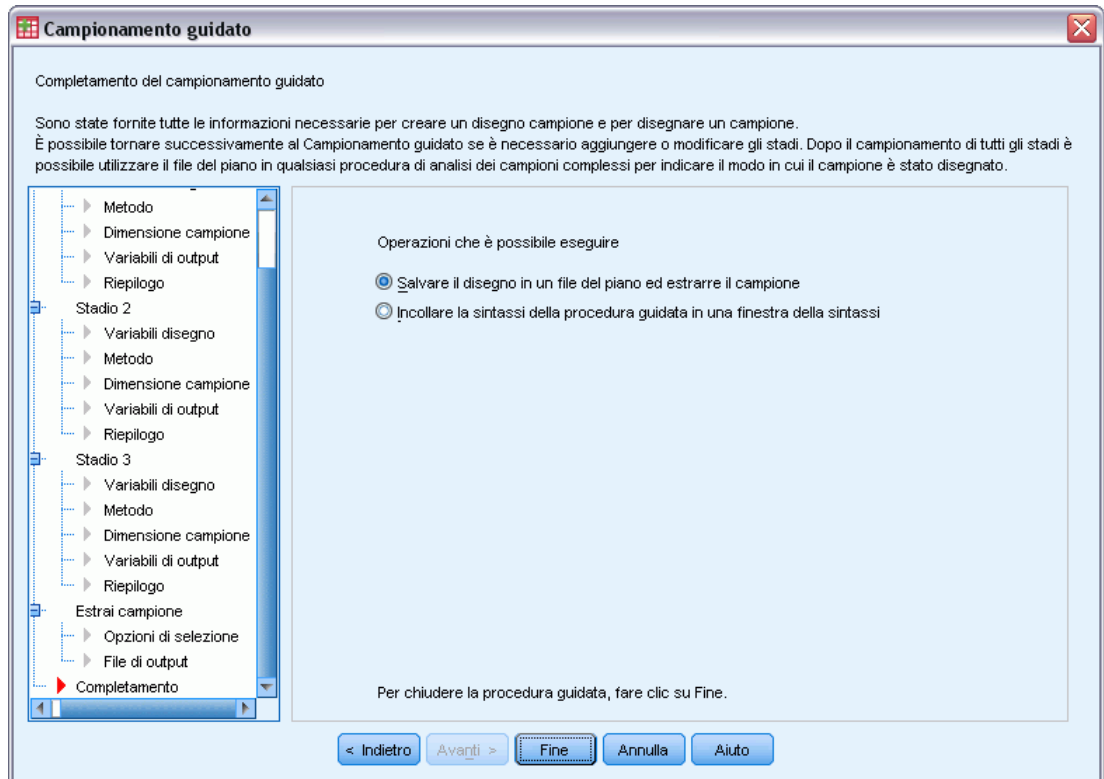
- Controllare il disegno di campionamento, quindi fare clic su Avanti.

Figura 13-24
Campionamento guidato, Passo Estrai campione - Opzioni di selezione



- ▶ Selezionare 1, 2 come stadi da campionare in questo momento.
- ▶ Selezionare Valore personalizzato per il tipo di seme casuale da utilizzare e digitare 241972 come valore.
L'utilizzo di un valore personalizzato consente di replicare esattamente i risultati dell'esempio.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Estrai campione - Opzioni di selezione.

Figura 13-25
Campionamento guidato, Passo Fine



- Fare clic su Fine.

Tali selezioni consentono di generare il file del piano di campionamento *demo.csplan* e di estrarre un campione in base ai primi due stadi di tale piano.

Risultati del campione

Figura 13-26
Editor dei dati con i risultati del campione.

	regione	provinc	distrett	città	InclusionPr obability_1_	SampleWei ghtCumulat ve_1_	InclusionPr obability_2_	SampleWei ghtCumulat ve_2_	SampleWei ght_Final_
295	1	2	10	295
296	1	2	10	296
297	1	2	10	297
298	1	2	10	298	,20	5,00	,10	50,00	50,00
299	1	2	10	299
300	1	2	10	300	,20	5,00	,10	50,00	50,00
301	1	2	11	301
302	1	2	11	302
303	1	2	11	303
304	1	2	11	304
305	1	2	11	305
306	1	2	11	306
307	1	2	11	307	,20	5,00	,10	50,00	50,00
308	1	2	11	308

Vengono visualizzati i risultati del campionamento nell'Editor dei dati. Nel file di lavoro sono state salvate cinque nuove variabili che rappresentano le probabilità di inclusione e i pesi di campionamento cumulato per ogni stadio, nonché i pesi di campionamento finali dei primi due stadi.

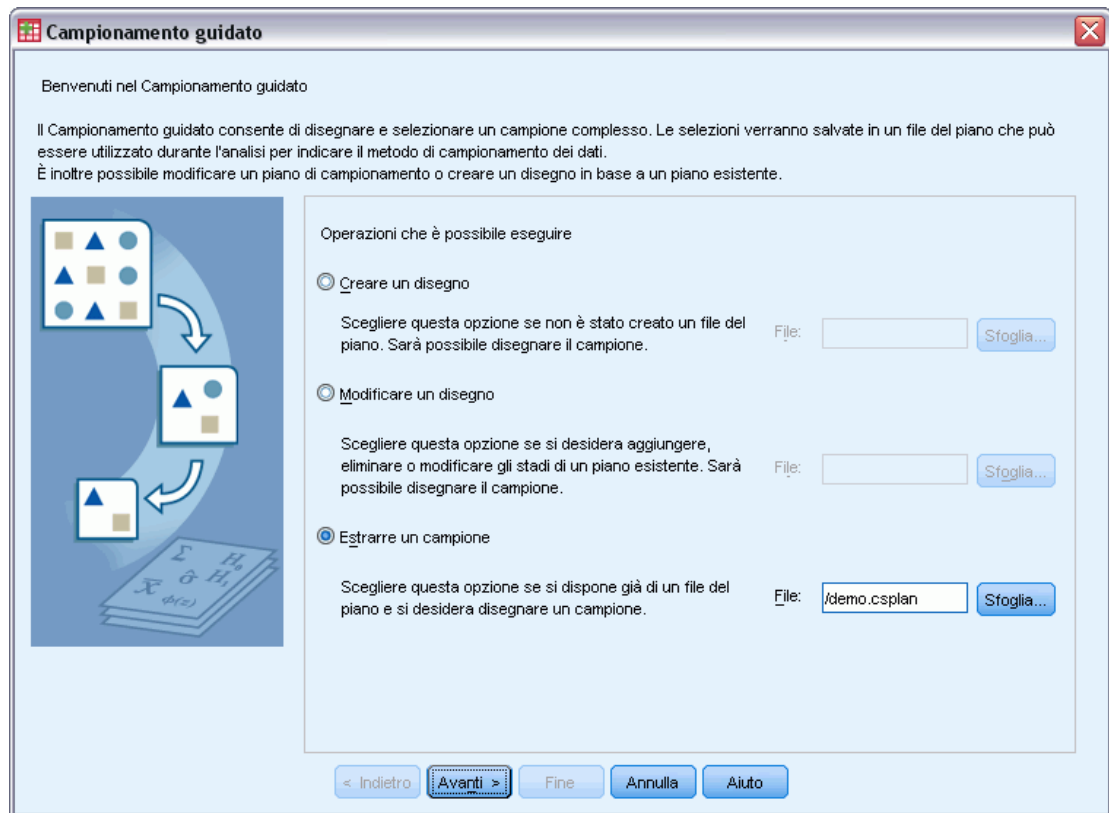
- Per il campione sono state selezionate le città con valori relativi a queste variabili.
- Non sono state selezionate le città con valori mancanti di sistema per le variabili.

Per ogni città selezionata, la società ha acquisito informazioni sulle unità abitazione e suddivisione, che sono state inserite nel file *demo_cs_2.sav*. Utilizzare questo file e Campionamento guidato per campionare il terzo stadio del disegno.

Utilizzo della procedura guidata per campionare dalla seconda struttura parziale

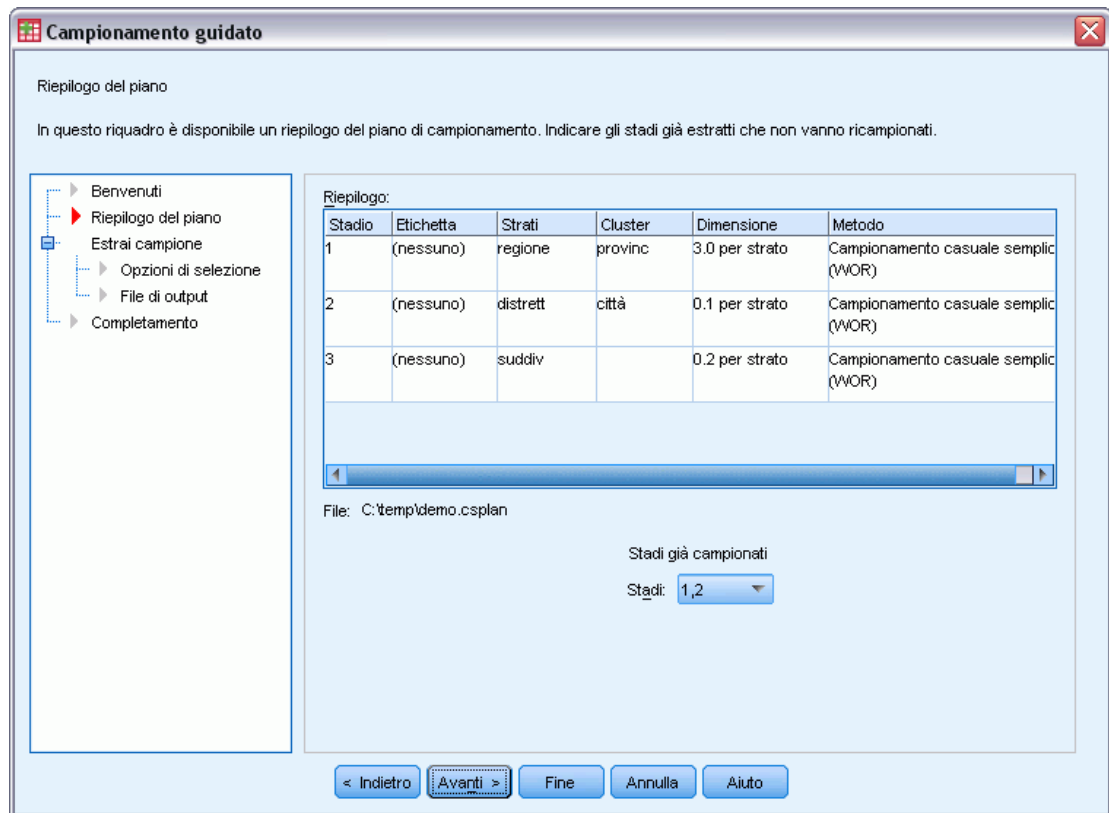
- Per eseguire Campionamento guidato di Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Seleziona un campione...

Figura 13-27
Campionamento guidato, Finestra di benvenuto



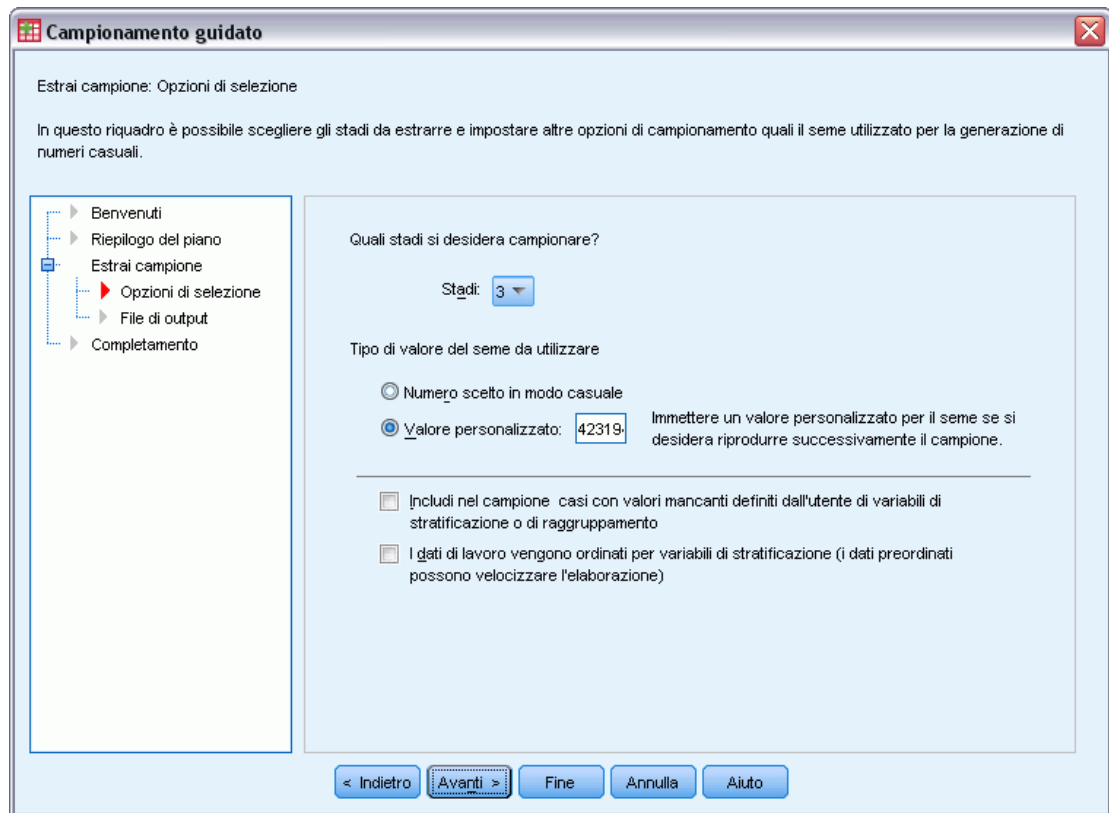
- ▶ Selezionare Estrarre un campione, individuare la posizione in cui è stato salvato il file del piano e selezionare il file del piano demo.csplan che è stato creato.
- ▶ Fare clic su Avanti.

Figura 13-28
Campionamento guidato, Passo Riepilogo del piano (stadio 3)



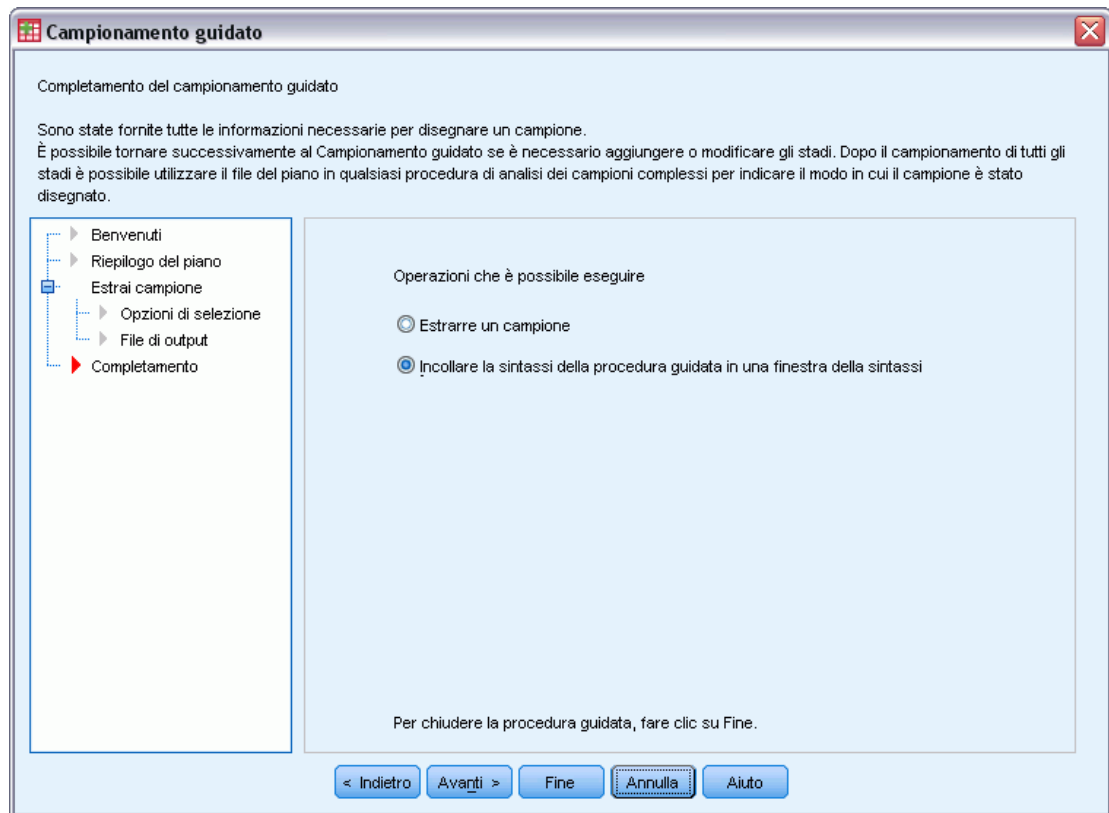
- ▶ Selezionare 1, 2 come stadi già campionati.
- ▶ Fare clic su Avanti.

Figura 13-29
Campionamento guidato, Passo Estrai campione - Opzioni di selezione



- ▶ Selezionare Valore personalizzato per il tipo di seme casuale da utilizzare e digitare 4231946 come valore.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Estrai campione - Opzioni di selezione.

Figura 13-30
Campionamento guidato, Passo Fine



- ▶ Selezionare Incollare la sintassi della procedura guidata in una finestra della sintassi.
- ▶ Fare clic su Fine.

Verrà generata la sintassi seguente:

```
* Campionamento guidato.
CSSELECT
/PLAN FILE='demo.csplan'
/CRITERIA STAGES = 3 SEED = 4231946
/CLASSMISSING EXCLUDE
/DATA RENAMEVARS
/PRINT SELECTION.
```

In tal caso se si stampa il riepilogo del campionamento viene prodotta una tabella complessa che crea problemi nel visualizzatore dell'output. Per disattivare la visualizzazione del riepilogo del campionamento, sostituire `SELECTION` con `CPS` nel sottocomando `PRINT`. Quindi eseguire la sintassi nella finestra di sintassi.

Le selezioni effettuate estraggono un campione in base al terzo stadio del piano di campionamento *demo.csplan*.

Risultati del campione

Figura 13-31
Editor dei dati con i risultati del campione.

	città	suddiv	unità	InclusionPr obability_1	SampleWei ghtCumulat ve_1	InclusionPr obability_2	SampleWei ghtCumulat ve_2	InclusionPr obability_3	SampleWei ghtCumulat ve_3	SampleWei ght_Final_
14	190	946	94514	,20	5,00	,10	50,00	.	.	.
15	190	946	94515	,20	5,00	,10	50,00	.	.	.
16	190	946	94516	,20	5,00	,10	50,00	,20	244,44	244,44
17	190	946	94517	,20	5,00	,10	50,00	.	.	.
18	190	946	94518	,20	5,00	,10	50,00	.	.	.
19	190	946	94519	,20	5,00	,10	50,00	.	.	.
20	190	946	94520	,20	5,00	,10	50,00	.	.	.
21	190	946	94521	,20	5,00	,10	50,00	.	.	.
22	190	946	94522	,20	5,00	,10	50,00	.	.	.
23	190	946	94523	,20	5,00	,10	50,00	.	.	.
24	190	946	94524	,20	5,00	,10	50,00	,20	244,44	244,44
25	190	946	94525	,20	5,00	,10	50,00	.	.	.
26	190	946	94526	,20	5,00	,10	50,00	.	.	.
27	190	946	94527	,20	5,00	,10	50,00	.	.	.
28	190	946	94528	,20	5,00	,10	50,00	.	.	.
29	190	946	94529	,20	5,00	,10	50,00	,20	244,44	244,44
30	190	946	94530	,20	5,00	,10	50,00	.	.	.

Vengono visualizzati i risultati del campionamento nell'Editor dei dati. Nel file di lavoro sono state salvate tre nuove variabili che rappresentano le probabilità di inclusione e i pesi di campionamento cumulato per il terzo stadio, nonché i pesi di campionamento finali. I nuovi pesi prendono in considerazione i pesi calcolati durante il campionamento dei primi due stadi.

- Per il campione sono state selezionate le unità con valori relativi a queste variabili.
- Non sono state selezionate le unità con valori mancanti di sistema per queste variabili.

La società ricorrerà alle proprie risorse per ottenere le informazioni del sondaggio relative alle unità abitative selezionate nel campione. Dopo aver raccolto i sondaggi, è possibile elaborare il campione con le procedure di analisi di Campioni complessi, utilizzando il piano di campionamento *demo.csplan* per fornire le specifiche di campionamento.

Campionamento con PPS (Probability Proportional to Size)

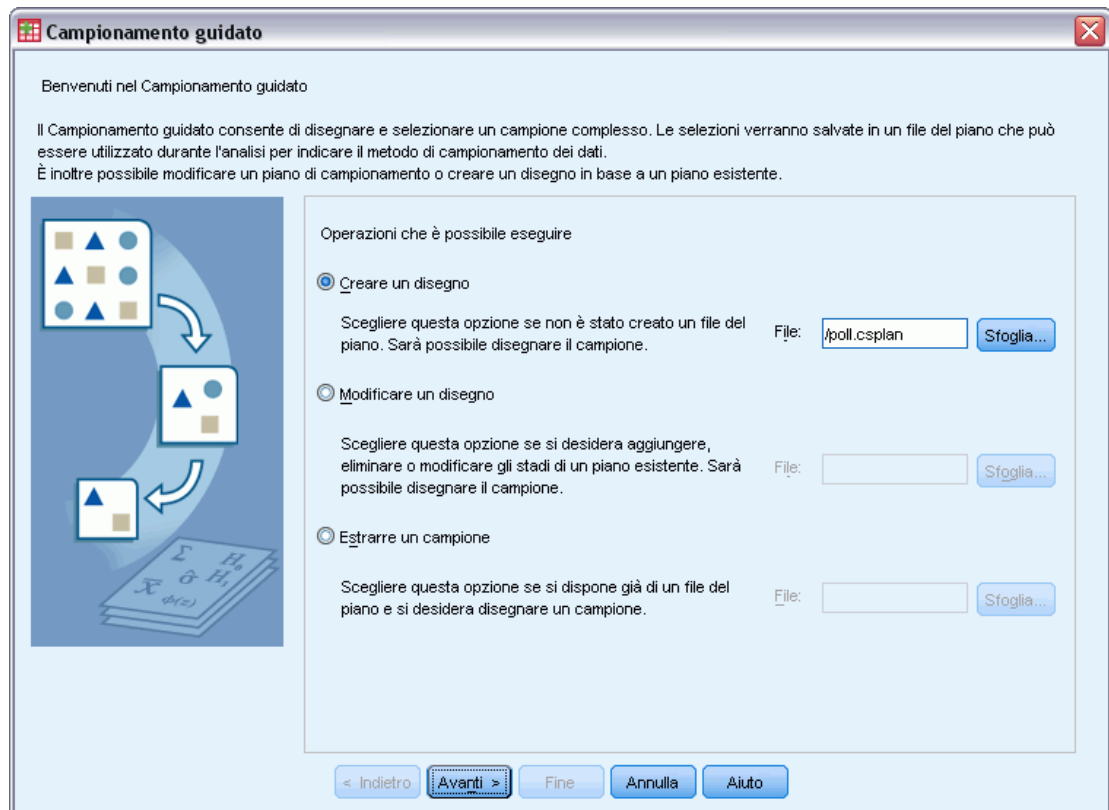
I deputati che discutono una legge in parlamento hanno l'interesse di sapere se la legge gode del sostegno pubblico e il modo in cui tale sostegno è correlato ai dati demografici degli elettori. Gli esperti in sondaggi di opinione sviluppano ed eseguono interviste in base a un disegno di campionamento complesso.

L'elenco degli elettori registrati si trova nel file *poll_cs.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20.](#) Utilizzare Campionamento guidato di Campioni complessi per selezionare un campione per un'ulteriore analisi.

Utilizzo della procedura guidata

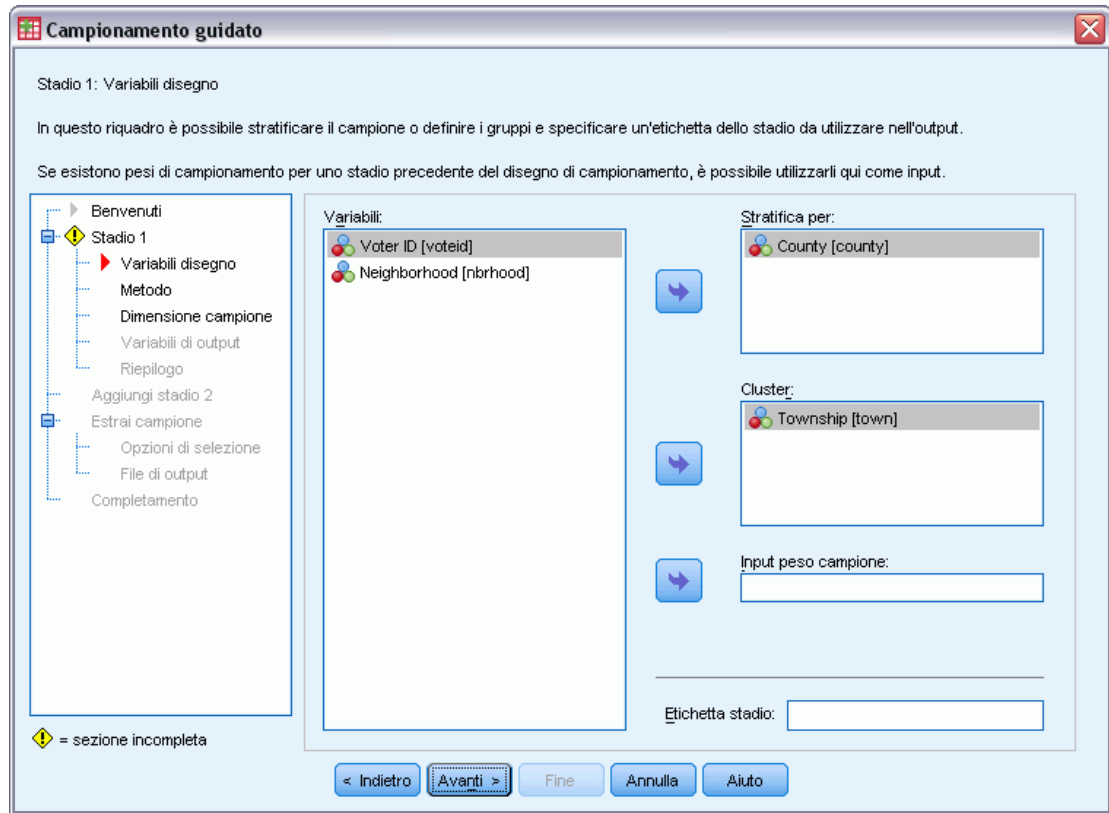
- Per eseguire Campionamento guidato di Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Seleziona un campione...

Figura 13-32
Campionamento guidato, Finestra di benvenuto



- Selezionare Creare un disegno, individuare la posizione in cui si desidera salvare il file e digitare poll.csplan come nome del file del piano.
- Fare clic su Avanti.

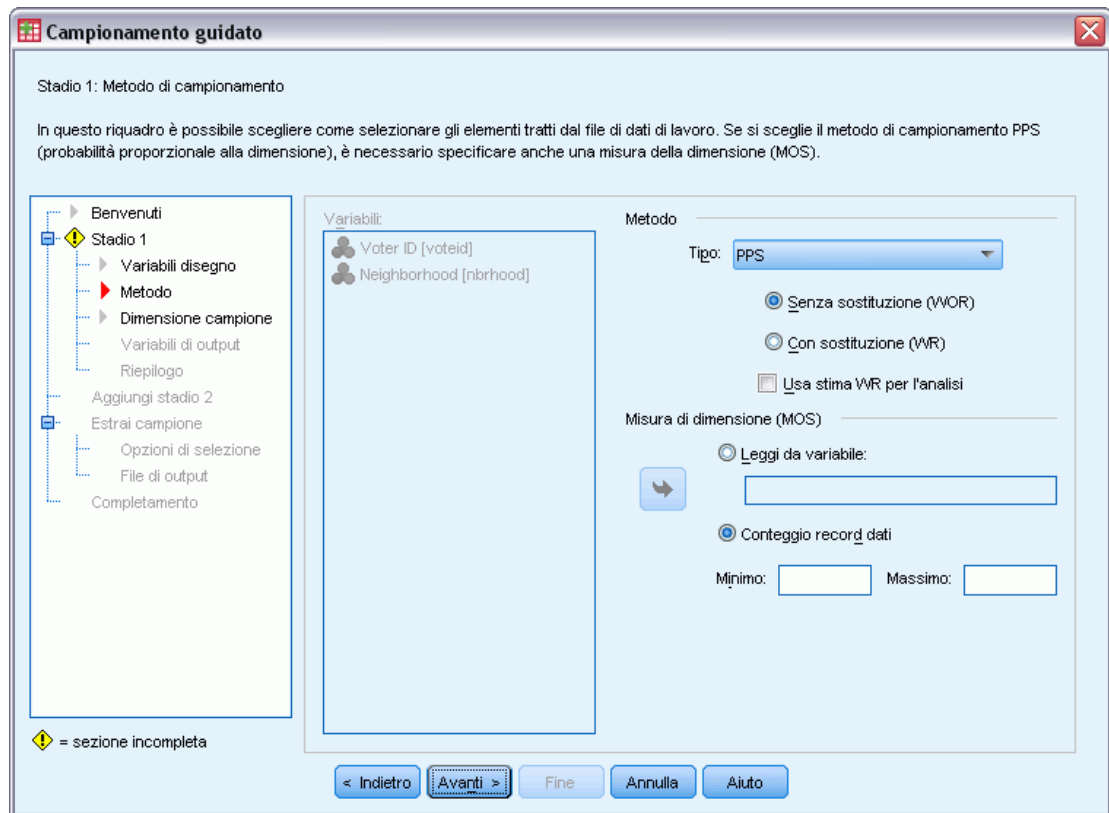
Figura 13-33
Campionamento guidato, Passo Variabili disegno (stadio 1)



- ▶ Selezionare *Zona* come variabile di stratificazione.
- ▶ Selezionare *Area* come variabile di raggruppamento.
- ▶ Fare clic su Avanti.

Tale struttura di disegno comporta l'estrazione di campioni indipendenti da ogni zona. In questo stadio le aree della città vengono estratte come unità di campionamento principale.

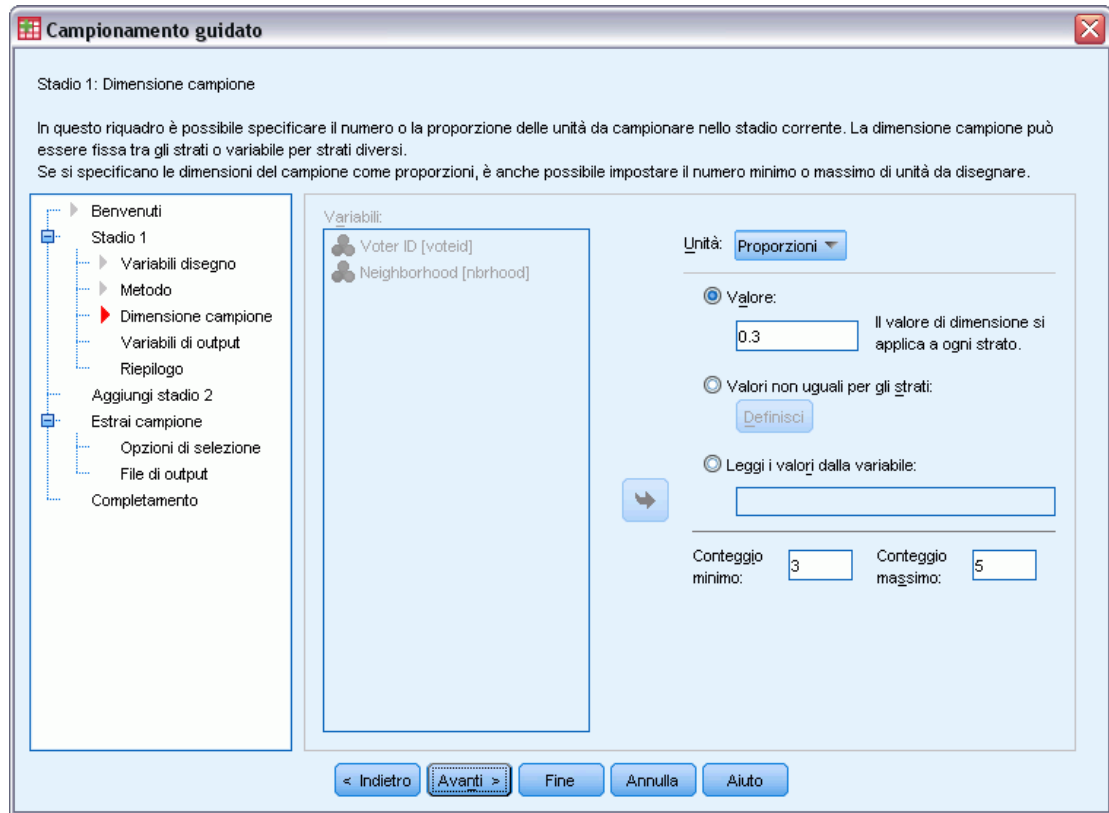
Figura 13-34
Campionamento guidato, Passo Metodo di campionamento (stadio 1)



- ▶ Selezionare il metodo di campionamento PPS.
- ▶ Selezionare la misura di dimensione Conteggio record dati.
- ▶ Fare clic su Avanti.

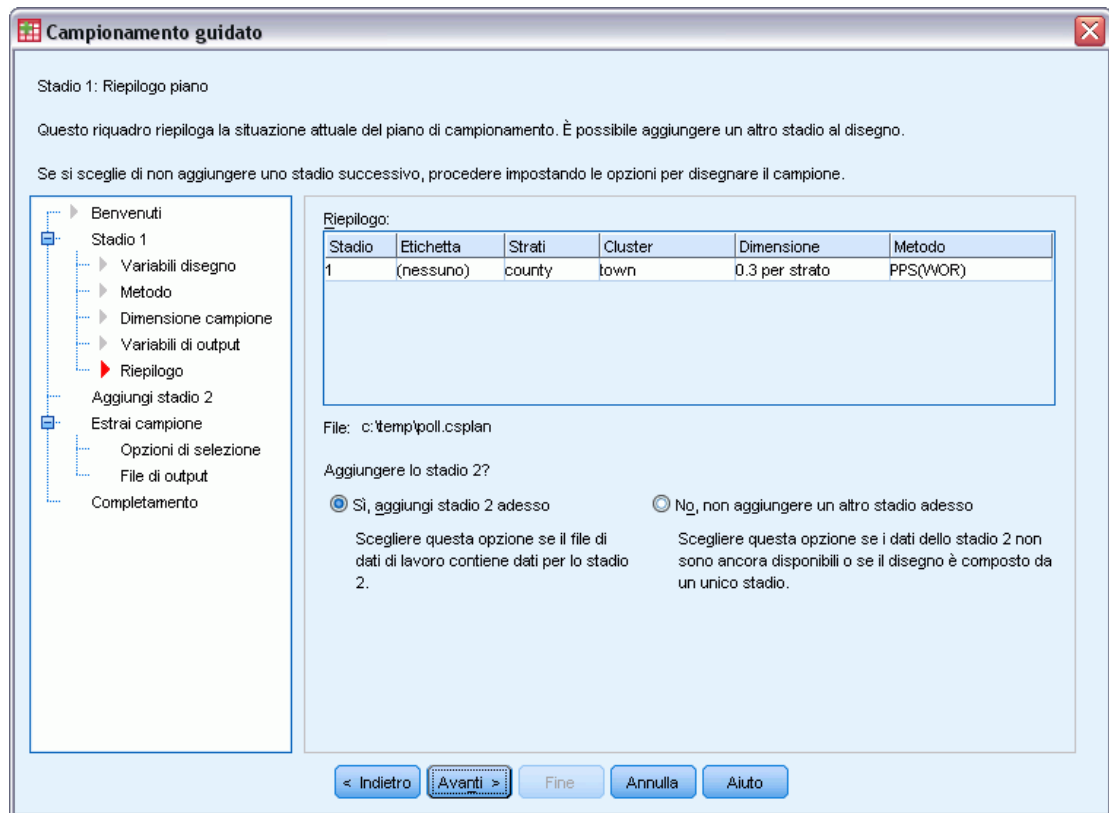
In ogni zona, le aree vengono estratte senza sostituzione con probabilità proporzionale al numero di record per ogni area. L'utilizzo di un metodo PPS genera probabilità di campionamento congiunte per le aree. Specificare la posizione in cui salvare questi valori nel passo File di output.

Figura 13-35
 Campionamento guidato, Passo Dimensione campione (stadio 1)



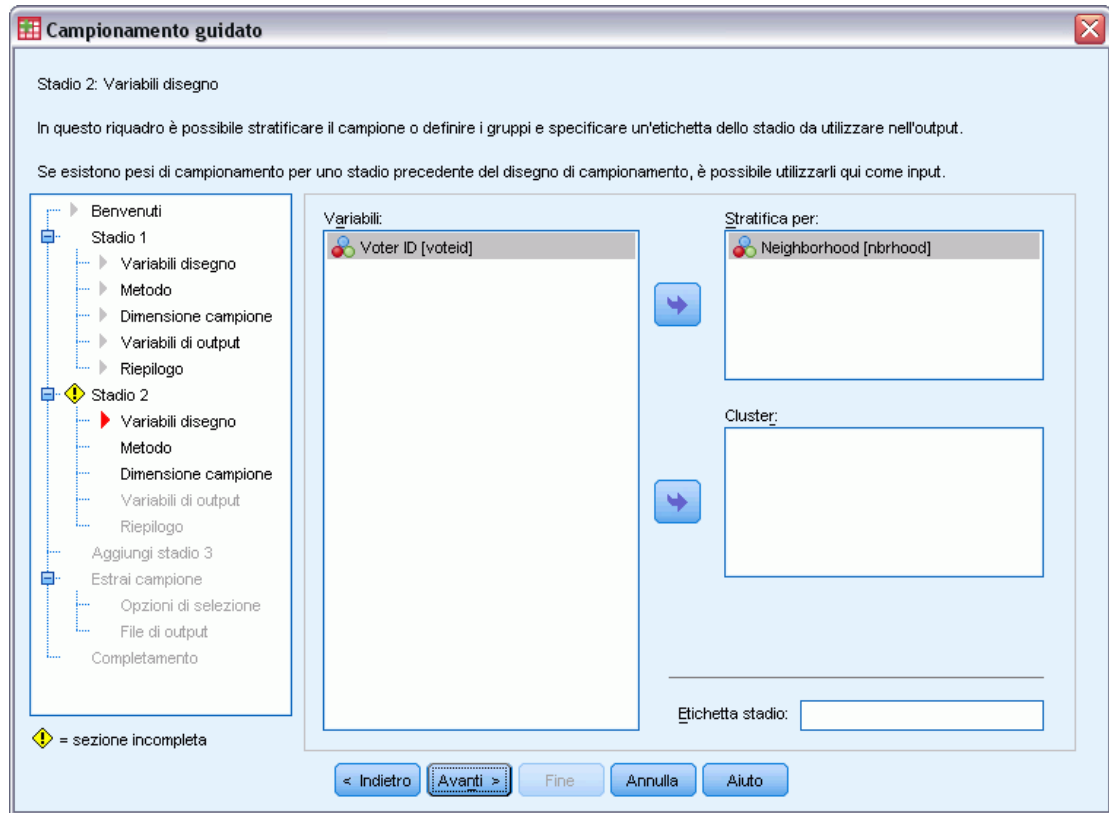
- ▶ Selezionare Proporzioni dall'elenco a discesa Unità.
- ▶ Digitare 0,3 come valore per la proporzione di aree da selezionare per zona in questo stadio.
 I legislatori provenienti dalla zona occidentale sottolineano che in quella zona esistono meno aree che in altre. Per garantire una rappresentazione adeguata, vorrebbero stabilire un minimo di 3 aree campionate da ogni zona.
- ▶ Digitare 3 come numero minimo di aree da selezionare e 5 come numero massimo.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-36
Campionamento guidato, Passo Riepilogo del piano (stadio 1)



- ▶ Selezionare Sì, aggiungere lo stadio 2 ora.
- ▶ Fare clic su Avanti.

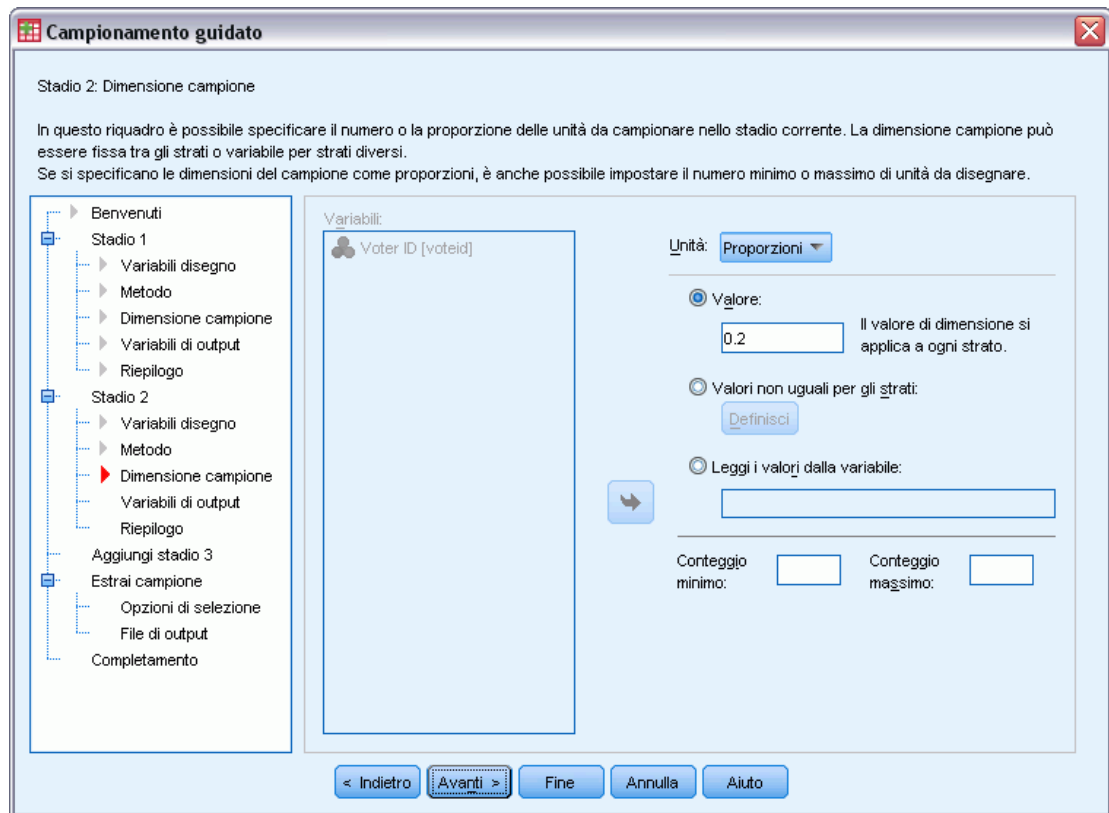
Figura 13-37
 Campionamento guidato, Passo Variabili disegno (stadio 2)



- ▶ Selezionare *Quartiere* come variabile di stratificazione.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Metodo di campionamento.

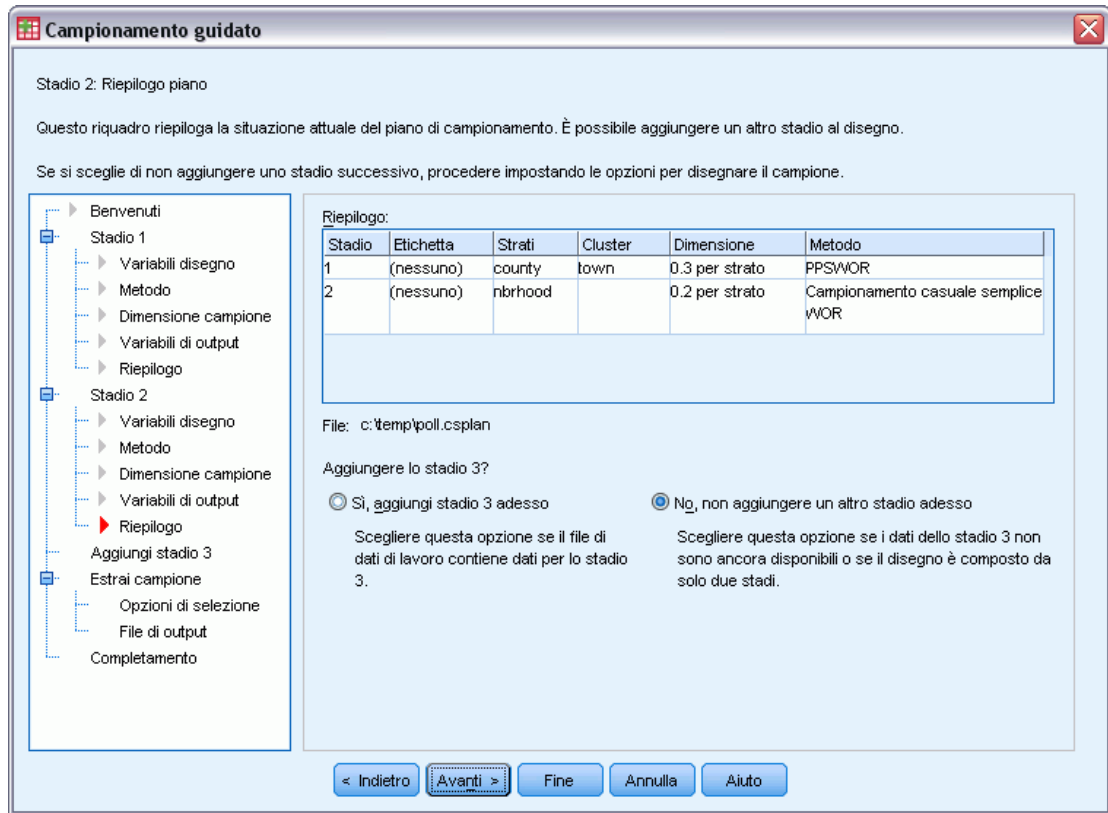
Tale struttura di disegno comporta l'estrazione di campioni indipendenti per ogni quartiere delle aree estratte nella fase 1. In questo stadio, gli elettori vengono estratti come unità di campionamento principale utilizzando il metodo di campionamento casuale semplice senza sostituzione.

Figura 13-38
Campionamento guidato, Passo Dimensione campione (stadio 2)



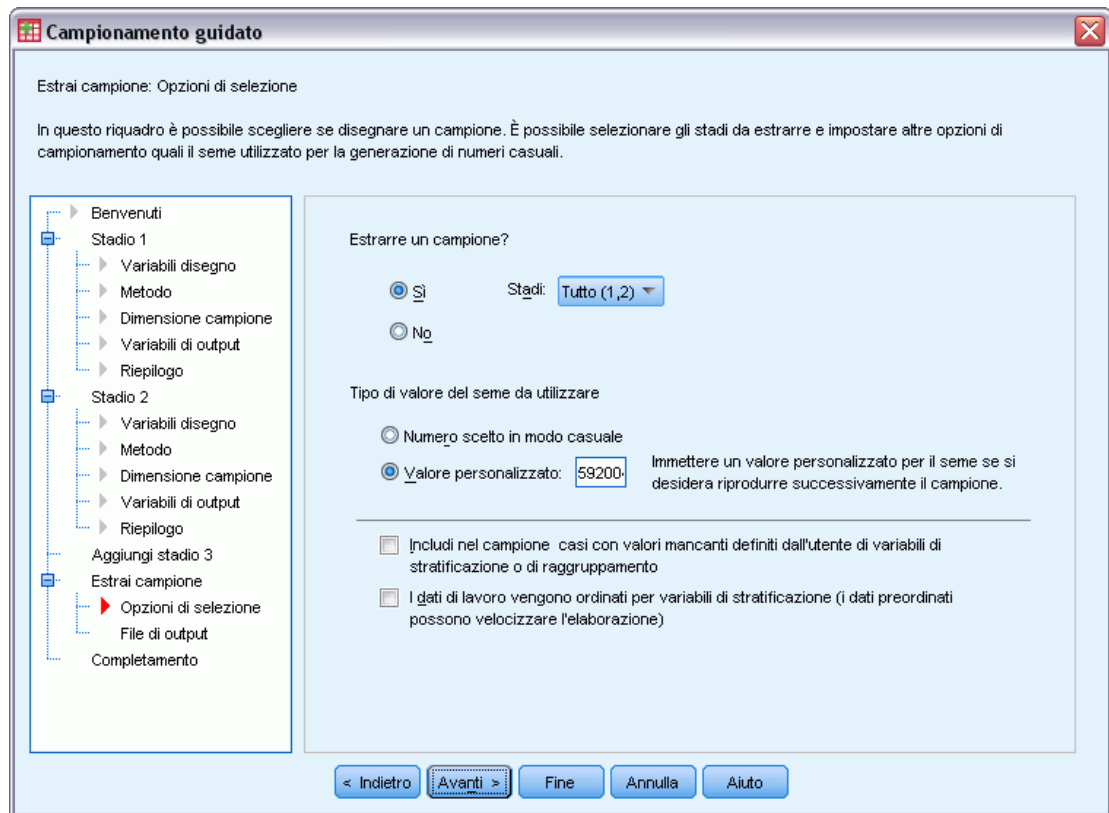
- ▶ Selezionare Proporzioni dall'elenco a discesa Unità.
- ▶ Digitare 0,2 come valore della proporzione di unità da campionare da ogni strato.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili di output.

Figura 13-39
Campionamento guidato, Passo Riepilogo del piano (stadio 2)



- Controllare il disegno di campionamento, quindi fare clic su Avanti.

Figura 13-40
Campionamento guidato, Passo Estrai campione - Opzioni di selezione

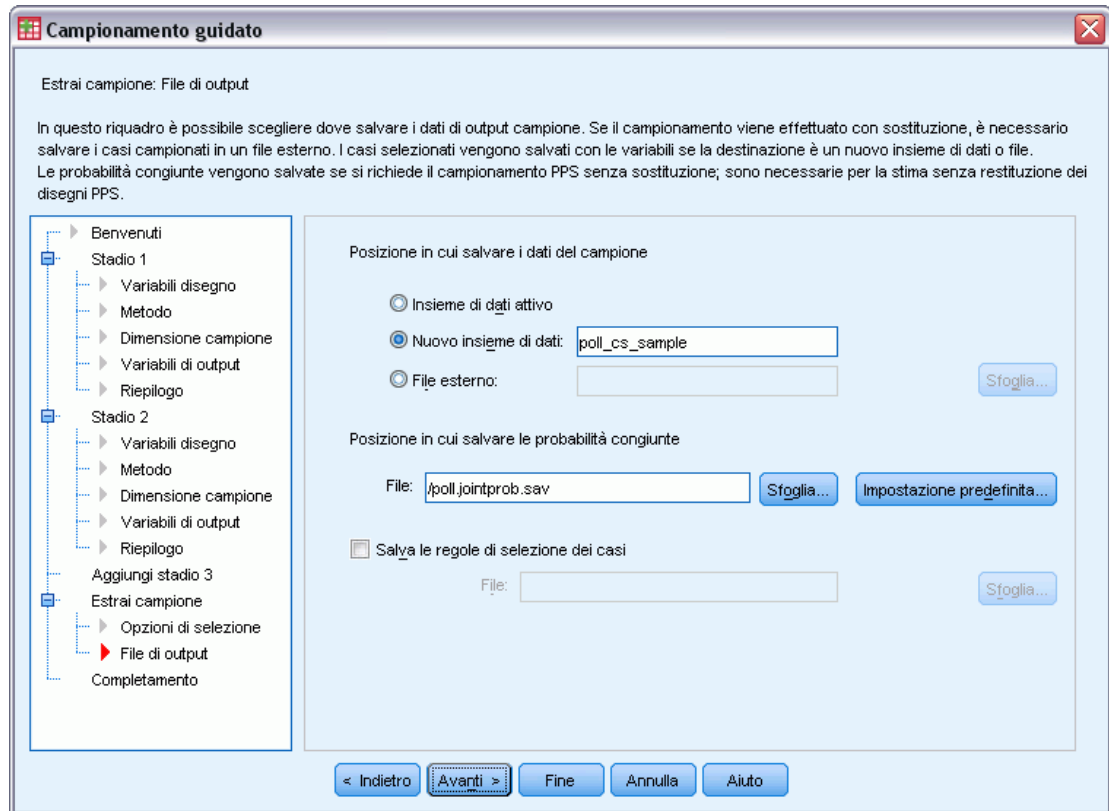


- ▶ Selezionare Valore personalizzato per il tipo di seme casuale da utilizzare e digitare 592004 come valore.

L'utilizzo di un valore personalizzato consente di replicare esattamente i risultati dell'esempio.

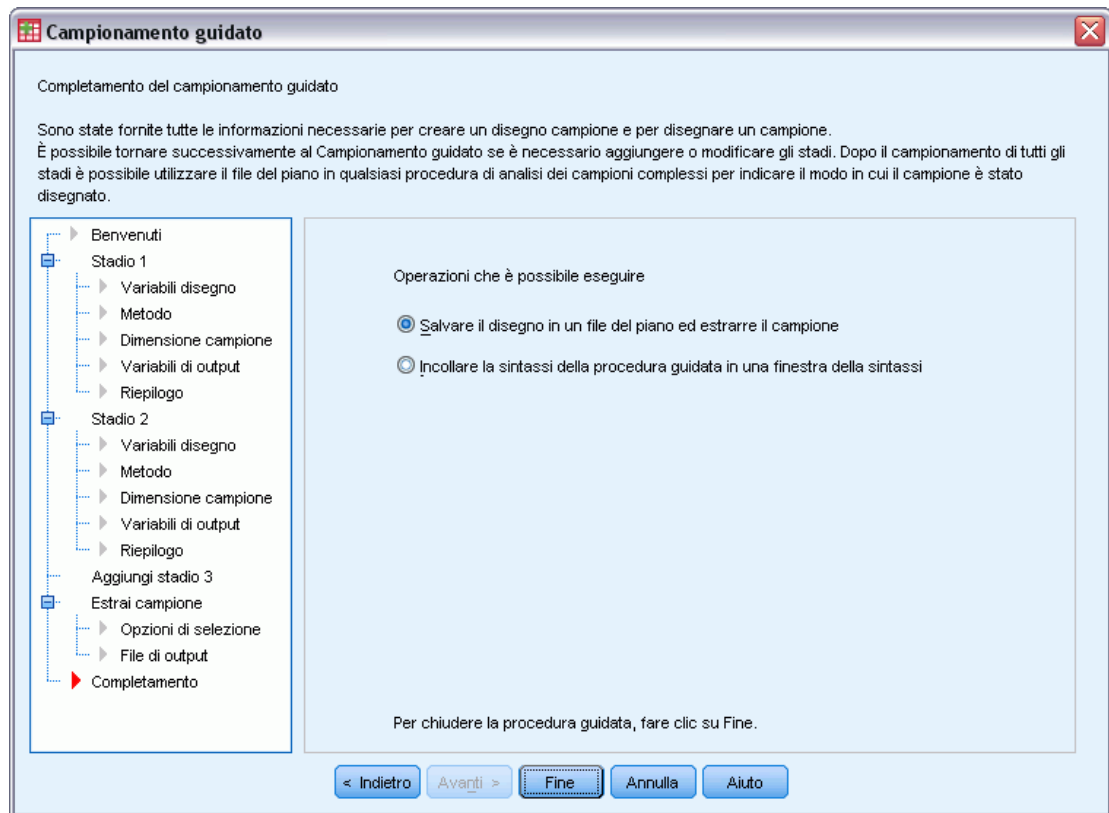
- ▶ Fare clic su Avanti.

Figura 13-41
Campionamento guidato, Passo Estrai campione - Opzioni di selezione



- ▶ Scegliere di salvare il campione in un nuovo insieme di dati e digitare poll_cs_sample come nome dell'insieme di dati.
- ▶ Individuare la posizione in cui si desidera salvare le probabilità congiunte e digitare poll_jointprob.sav come nome del file delle probabilità congiunte.
- ▶ Fare clic su Avanti.

Figura 13-42
Campionamento guidato, Passo Fine



- Fare clic su Fine.

Tali selezioni consentono di generare il file del piano di campionamento *poll.csplan* e di estrarre un campione in base al piano, salvare i risultati del campione nel nuovo insieme di dati *poll_cs_sample* e salvare il file delle probabilità congiunte nel file di dati esterno *poll_jointprob.sav*.

Riepilogo del piano

Figura 13-43
Riepilogo del piano

			Stadio 1	Stadio 2
Variabili disegno	Stratificazione	1	County	Neighborhood
	Raggruppamento	1	Township	
Informazioni sul campione	Metodo di selezione		Sample PPS_WOR	Sample SIMPLE_WOR
	Misura di dimensione		Ottenuto dai dati	
	Proporzione di unità campionate		,3	,2
	Numero minimo di unità campionate		3	
	Numero massimo di unità campionate		5	
	Variabili create o modificate	Probabilità di inclusione (selezione) a stadi	InclusionProbability_1_	InclusionProbability_2_
Informazioni sull'analisi		Peso di campionamento cumulato a stadi	SampleWeight Cumulative_1_	SampleWeight Cumulative_2_
		Ipotesi stimatore	Campionamento con probabilità non uguale senza sostituzione (utilizzando le probabilità di inclusione congiunta)	Campionamento con probabilità uguale senza sostituzione
		Probabilità di inclusione	Ottenuto dalla variabile InclusionProbability_1_	Ottenuto dalla variabile InclusionProbability_2_
Variabili disegno				

File del piano: c:\poll.csplan
Variabile peso: Sampleweight_f_ina_

La tabella riassuntiva consente di verificare il piano di campionamento ed è utile per assicurarsi che il piano rappresenti le intenzioni dell'utente.

Riepilogo del campionamento

Figura 13-44
Riepilogo stadio

County	Numero di unità campionate		Proporzione di unità campionate	
	Richiesti	Effettivi	Richiesti	Effettivi
Eastern	4	4	30,0%	30,8%
Central	4	4	30,0%	30,8%
Western	3	3	30,0%	50,0%
Northern	5	5	30,0%	33,3%
Southern	3	3	30,0%	50,0%

File del piano: c:\poll.csplan

La tabella riassuntiva consente di verificare il primo stadio di campionamento ed è utile per controllare che il campionamento sia stato eseguito in base al piano. Inizialmente è stato richiesto un campione al 30% delle aree per zona. Le proporzioni campionate effettive sono prossime al 30%, fatta eccezione per la zona occidentale e quella meridionale. Ciò è dovuto al fatto che queste zone presentano solo sei aree ciascuna ed è stato specificato il requisito di selezionare almeno tre aree per ogni zona.

Figura 13-45
Riepilogo stadio

County	Township	Neighborhood	Numero di unità campionate		Proporzione di unità campionate		
			Richiesti	Effettivi	Richiesti	Effettivi	
Eastern	9	1	49	49	20,0%	19,9%	
		2	143	143	20,0%	20,0%	
		3	113	113	20,0%	20,0%	
		4	77	77	20,0%	20,0%	
		5	139	139	20,0%	20,0%	
		6	120	120	20,0%	20,0%	
	10	1	149	149	20,0%	20,1%	
		2	117	117	20,0%	20,0%	
		3	116	116	20,0%	20,0%	
		4	69	69	20,0%	19,9%	
	11	1	65	65	20,0%	19,9%	
		2	72	72	20,0%	19,9%	
		3	109	109	20,0%	20,0%	
		4	140	140	20,0%	20,0%	
		5	42	42	20,0%	19,8%	
		6	142	142	20,0%	20,0%	
	12	1	145	145	20,0%	20,1%	
		2	69	69	20,0%	20,1%	
		3	98	98	20,0%	20,1%	
		4	134	134	20,0%	20,0%	
		5	114	114	20,0%	20,0%	
		6	137	137	20,0%	19,9%	
	Central	2	1	119	119	20,0%	20,1%
			2	153	153	20,0%	19,9%
3			101	101	20,0%	20,0%	
4			52	52	20,0%	19,8%	
5			144	144	20,0%	20,0%	

File del piano: c:\poll2.csplan

La tabella riassuntiva, di cui viene visualizzata la parte superiore, consente di verificare il secondo stadio di campionamento. È inoltre utile per controllare che il campionamento sia stato eseguito in base al piano. Da ogni quartiere di ogni area campionata nel primo stadio è stato estratto approssimativamente il 20% degli elettori, come richiesto.

Risultati del campione

Figura 13-46
Editor dei dati con i risultati del campione.

	voteid	nrhood	town	county	InclusionPr obability_1_	SampleWei ghtCumulat ve_1	InclusionPr obability_2_	SampleWei ghtCumulat ve_2	SampleWei ght_Final_
376	368	4	9	1	,44	2,26	,20	11,28	11,28
377	369	4	9	1	,44	2,26	,20	11,28	11,28
378	374	4	9	1	,44	2,26	,20	11,28	11,28
379	376	4	9	1	,44	2,26	,20	11,28	11,28
380	379	4	9	1	,44	2,26	,20	11,28	11,28
381	380	4	9	1	,44	2,26	,20	11,28	11,28
382	382	4	9	1	,44	2,26	,20	11,28	11,28
383	13	5	9	1	,44	2,26	,20	11,26	11,26
384	18	5	9	1	,44	2,26	,20	11,26	11,26
385	23	5	9	1	,44	2,26	,20	11,26	11,26
386	38	5	9	1	,44	2,26	,20	11,26	11,26
387	39	5	9	1	,44	2,26	,20	11,26	11,26
388	40	5	9	1	,44	2,26	,20	11,26	11,26
389	41	5	9	1	,44	2,26	,20	11,26	11,26

È stato possibile vedere i risultati di campionamento nell'insieme di dati appena creato. Nel file di lavoro sono state salvate cinque nuove variabili che rappresentano le probabilità di inclusione e i pesi di campionamento cumulato per ogni stadio, nonché i pesi di campionamento finali. Gli elettori non selezionati nel campione sono stati esclusi dall'insieme di dati.

I pesi di campionamento finali sono identici per gli elettori nello stesso quartiere perché selezionati in base a un metodo di campionamento casuale semplice all'interno dei quartieri. Sono tuttavia diversi tra quartieri della stessa area perché le proporzioni campionate non sono esattamente il 20% in tutti i quartieri.

Figura 13-47

Editor dei dati con i risultati del campione.

	voteid	nbrhood	town	county	InclusionPr obability_1	SampleWei ghtCumulat ve_1	InclusionPr obability_2	SampleWei ghtCumulat ve_2	SampleWei ght_Final_
635	577	6	9	1	,44	2,26	,20	11,30	11,30
636	578	6	9	1	,44	2,26	,20	11,30	11,30
637	582	6	9	1	,44	2,26	,20	11,30	11,30
638	590	6	9	1	,44	2,26	,20	11,30	11,30
639	594	6	9	1	,44	2,26	,20	11,30	11,30
640	597	6	9	1	,44	2,26	,20	11,30	11,30
641	600	6	9	1	,44	2,26	,20	11,30	11,30
642	4	1	10	1	,31	3,21	,20	16,00	16,00
643	5	1	10	1	,31	3,21	,20	16,00	16,00
644	9	1	10	1	,31	3,21	,20	16,00	16,00
645	10	1	10	1	,31	3,21	,20	16,00	16,00
646	12	1	10	1	,31	3,21	,20	16,00	16,00
647	16	1	10	1	,31	3,21	,20	16,00	16,00
648	17	1	10	1	,31	3,21	,20	16,00	16,00

Diversamente dagli elettori nel secondo stadio, i pesi di campionamento del primo stadio non sono identici per aree nella stessa zona perché selezionati con il metodo PPS (Probability Proportional to Size, probabilità proporzionale alla dimensione).

Figura 13-48

File di probabilità congiunta

	county	town	Unit_No_	Joint_Prob 1	Joint_Prob 2	Joint_Prob 3	Joint_Prob 4	Joint_Prob 5
1	1	10	1	,31	,10	,11	,12	.
2	1	11	2	,10	,39	,15	,16	.
3	1	9	3	,11	,15	,44	,21	.
4	1	12	4	,12	,16	,21	,48	.
5	2	12	1	,22	,04	,07	,08	.
6	2	6	2	,04	,23	,07	,08	.
7	2	7	3	,07	,07	,41	,19	.
8	2	2	4	,08	,08	,19	,45	.
9	3	5	1	,58	,31	,32	.	.
10	3	3	2	,31	,61	,36	.	.
11	3	4	3	,32	,36	,63	.	.
12	4	14	1	,26	,06	,06	,07	,09
13	4	8	2	,06	,29	,07	,08	,10
14	4	4	3	,06	,07	,29	,08	,10
15	4	2	4	,07	,08	,08	,33	,12
16	4	13	5	,09	,10	,10	,12	,43
17	5	3	1	,74	,25	,27	.	.
18	5	6	2	,25	,41	,13	.	.
19	5	4	3	,27	,13	,43	.	.

Il file *poll_jointprob.sav* contiene le probabilità del primo stadio per le aree selezionate nelle contee. La *contea* rappresenta una variabile di stratificazione di primo stadio e *area* è una variabile di raggruppamento. Le combinazioni di queste variabili identificano tutti i valori PSU del primo stadio in modo univoco. *Unit_No_* assegna ai valori PSU un'etichetta nell'ambito di ciascun

strato e viene usato per la corrispondenza con *Joint_Prob_1_*, *Joint_Prob_2_*, *Joint_Prob_3_*, *Joint_Prob_4_* e *Joint_Prob_5_*. I primi due strati hanno ciascuno 4 valori PSU, quindi le matrici di probabilità di inclusione congiunta sono 4×4 per questi strati e la colonna *Joint_Prob_5_* viene lasciata vuota per queste righe. Allo stesso modo, gli strati 3 e 5 hanno 3×3 matrici di probabilità di inclusione congiunta, mentre lo strato 4 ne ha 5×5.

La necessità di usare un file di probabilità congiunta risulta evidente se si osservano i valori delle matrici di probabilità di inclusione congiunta. Se il metodo di campionamento non è un metodo PPS WOR, la scelta di un valore PSU è indipendente dalla scelta di un altro valore PSU, quindi la probabilità di inclusione congiunta risultante è semplicemente il prodotto delle rispettive probabilità di inclusione. Diversamente, la probabilità di inclusione congiunta per le aree 9 e 10 della contea 1 è pari a circa 0,11 (vedere il primo caso di *Joint_Prob_3_* o il terzo caso di *Joint_Prob_1_*) o comunque inferiore al prodotto delle singole probabilità di inclusione (il prodotto del primo caso di *Joint_Prob_1_* e il terzo caso di *Joint_Prob_3_* è $0,31 \times 0,44 = 0,1364$).

A questo punto il campione selezionato verrà intervistato. Quando i risultati sono disponibili, è possibile elaborare il campione con le procedure di analisi di Campioni complessi, utilizzando il piano di campionamento *poll.csplan* per fornire le specifiche di campionamento e *poll_jointprob.sav* per fornire le probabilità di inclusione congiunte necessarie.

Procedure correlate

La procedura Campionamento guidato di Campioni complessi è uno strumento utile per creare un file del piano di campionamento ed estrarre un campione.

- Per leggere un campione ai fini dell'analisi quando non è possibile accedere al file del piano di campionamento, utilizzare [Preparazione guidata dell'analisi](#).

Preparazione guidata dell'analisi Campioni complessi

Preparazione guidata dell'analisi consente di eseguire le varie fasi di creazione o modifica di un piano analisi da usare con le diverse procedure di analisi di Campioni complessi. Si tratta di una procedura molto utile quando non è possibile accedere al file del piano di campionamento utilizzato per estrarre il campione.

Utilizzo della Preparazione guidata dell'analisi Campioni complessi per leggere i dati pubblici NHIS

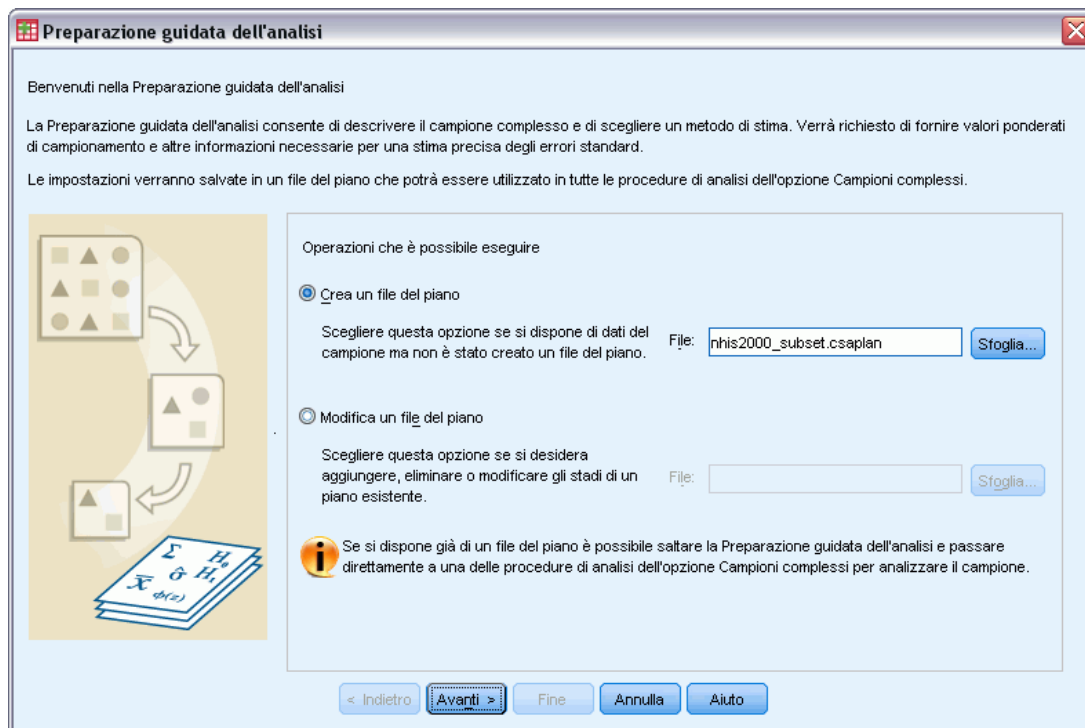
Il National Health Interview Survey (NHIS) è un sondaggio di grandi dimensioni condotto sulla popolazione civile americana. Le interviste vengono realizzate di persona e si basano su un campione rappresentativo di famiglie a livello nazionale. Per ogni membro di una famiglia vengono raccolte osservazioni e informazioni di carattere demografico relative allo stato di salute.

Un sottoinsieme del sondaggio realizzato nel 2000 è contenuto nel file *nhis2000_subset.sav*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Utilizzare Preparazione guidata dell'analisi Campioni complessi per creare un piano analisi per questo file di dati in modo da elaborarlo tramite le procedure di analisi di Complex Samples.

Utilizzo della procedura guidata

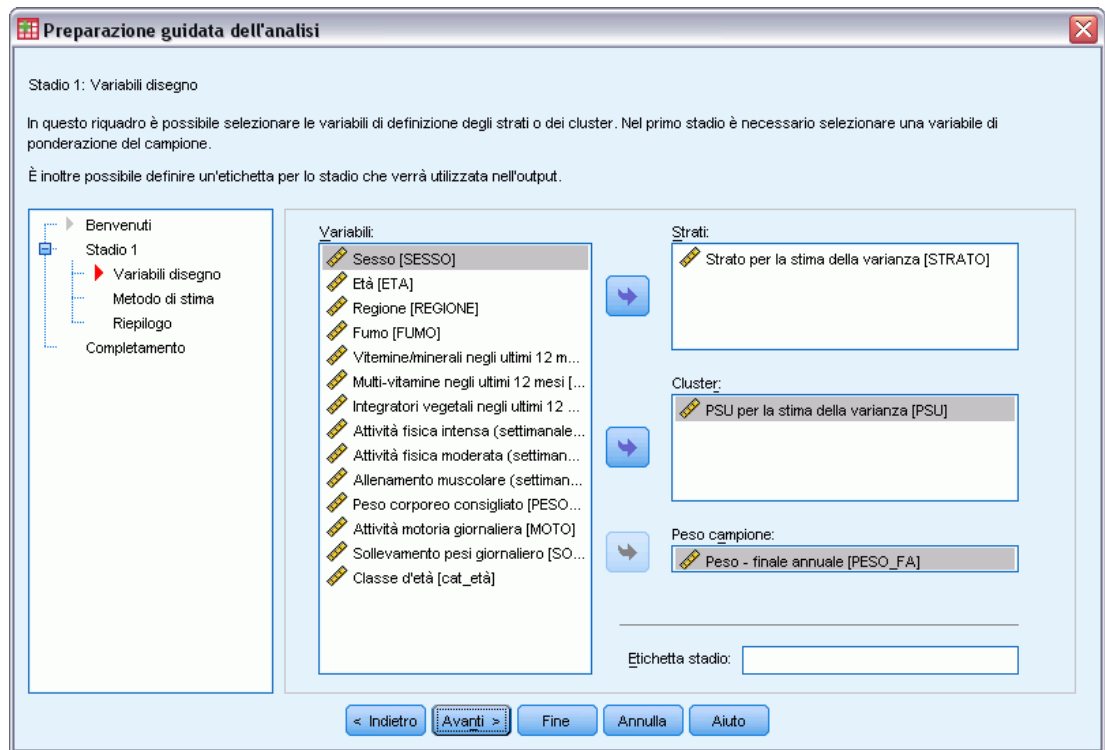
- ▶ Per preparare un campione tramite la procedura guidata Preparazione guidata dell'analisi Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Prepara per l'analisi...

Figura 14-1
Preparazione guidata dell'analisi, Finestra di benvenuto



- ▶ Individuare la posizione in cui si desidera salvare il file del piano e digitare `nhis2000_subset.csaplan` come nome del file del piano di analisi.
- ▶ Fare clic su Avanti.

Figura 14-2
Preparazione guidata dell'analisi - Passo Variabili disegno (stadio 1)



I dati vengono ottenuti utilizzando un campione complesso a più stadi. Tuttavia, per gli utenti finali, le variabili di progettazione NHIS originali sono trasformate in un insieme semplificato di disegno e variabili ponderate i cui risultati assomigliano a quelli delle strutture di disegno originali.

- ▶ Selezionare *Strato per stima della varianza* come variabile di stratificazione.
- ▶ Selezionare *PSU per stima della varianza* come variabile di raggruppamento.
- ▶ Selezionare *Peso - Annuale finale* come variabile di ponderazione del campione.
- ▶ Fare clic su Fine.

Riepilogo

Figura 14-3
Riepilogo

			Stadio 1
Variabili disegno	Stratificazione	1	STRATUM
	Raggruppamento	1	PSU per la stima della varianza
Informazioni sul campione	Metodo di selezione		Metodo di selezione
Informazioni sull'analisi	Ipotesi stimatore		Campionamento con sostituzione

File del piano: c:\nhis2000_subset.csaplan
Variabile peso: Peso - finale annuale
Stimatore SRS: Campionamento senza sostituzione

La tabella riassuntiva consente di verificare il piano analisi. Il piano è formato da uno stadio caratterizzato dal disegno di una variabile di stratificazione e di una variabile di raggruppamento. Viene utilizzata la stima con sostituzione (WR) e il piano viene salvato in *c:\nhis2000_subset.csaplan*. A questo punto è possibile utilizzare il file del piano per elaborare il file *nhis2000_subset.sav* tramite le procedure di analisi di Campioni complessi.

Preparazione per l'analisi quando i pesi di campionamento non sono inclusi nel file di dati

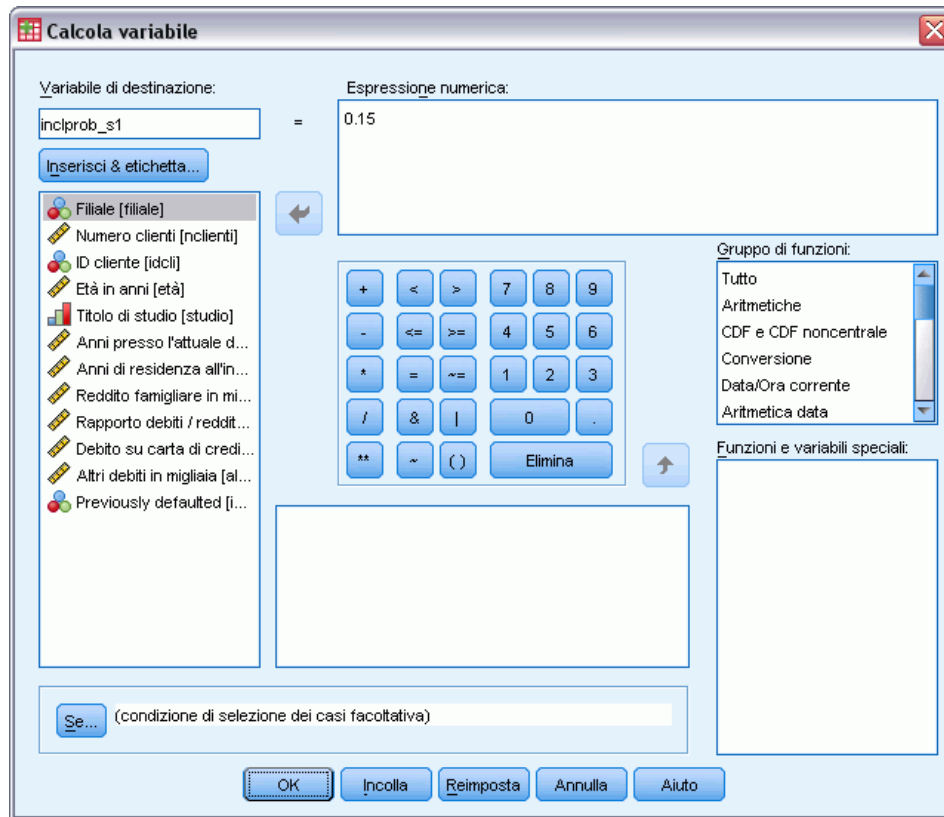
Un funzionario mutui ha raccolto una serie di dati relativi a clienti, in base a un disegno complesso; tuttavia, i pesi di campionamento non sono inclusi nel file. Tali informazioni sono contenute nel file *bankloan_cs_noweights.sav*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Partendo dalle proprie conoscenze relativamente al disegno di campionamento, il funzionario intende utilizzare Preparazione guidata dell'analisi Campioni complessi per creare un piano analisi per questo file di dati in modo da elaborarlo tramite le procedure di analisi Campioni complessi.

Il funzionario sa che i documenti sono stati selezionati in due stadi, con 15 filiali bancarie su 100 selezionate con uguale probabilità e senza sostituzione nel primo stadio. 100 clienti sono stati selezionati da ognuna di tali banche con uguale probabilità e senza sostituzione nel secondo stadio; il file di dati include inoltre informazioni sul numero di clienti di ciascuna banca. Il primo passaggio per creare un piano analisi è calcolare le probabilità di inclusione per stadi e i pesi di campionamento finali.

Calcolo delle probabilità di inclusione e dei pesi di campionamento

- Per calcolare le probabilità di inclusione per il primo stadio, dai menu scegliere:
Trasforma > Calcola variabile...

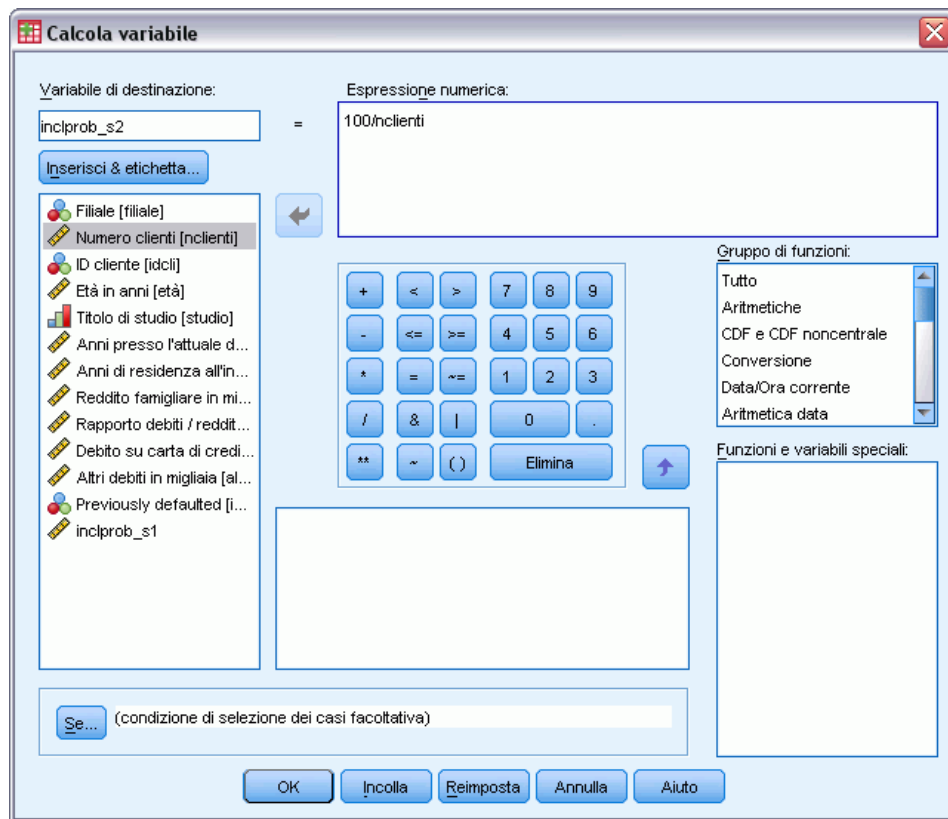
Figura 14-4
Finestra di dialogo Calcola variabile



15 filiali su 100 sono state selezionate senza sostituzione nel primo stadio; di conseguenza, la probabilità di una determinata banca di essere selezionata è $15/100 = 0,15$.

- ▶ Digitare inclprob_s1 come variabile di destinazione.
- ▶ Digitare 0,15 come espressione numerica.
- ▶ Fare clic su OK.

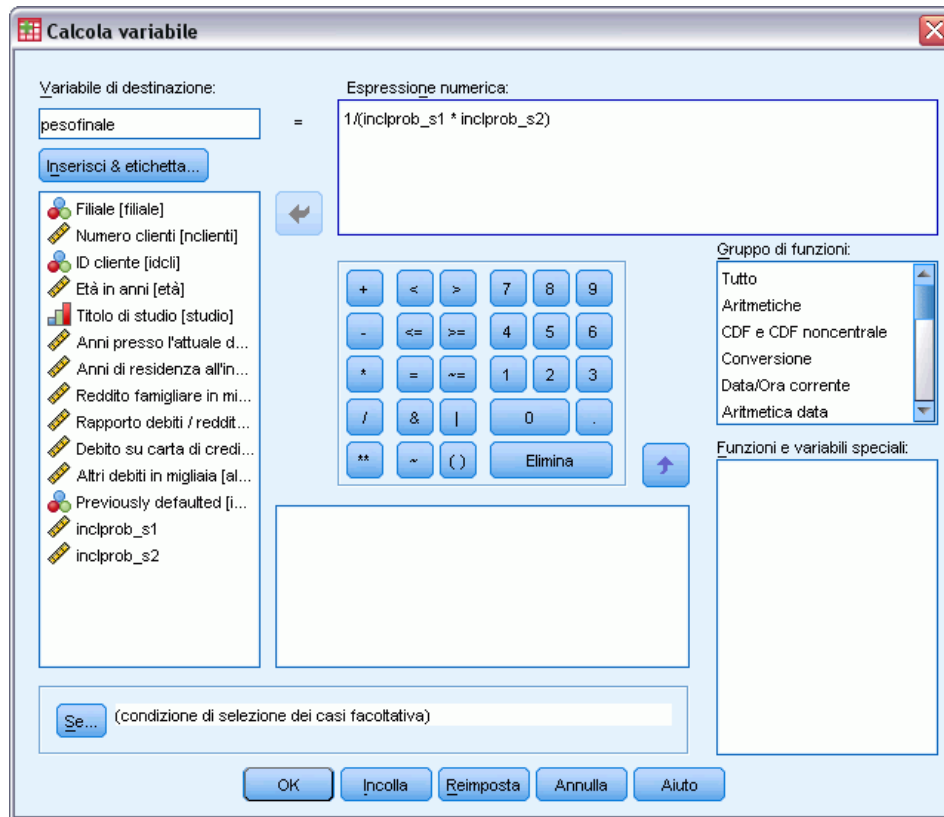
Figura 14-5
Finestra di dialogo Calcola variabile



100 clienti sono stati selezionati da ogni filiale nel secondo stadio; di conseguenza, la probabilità di inclusione nello stadio 2 per un determinato cliente di una certa banca è $100/\text{il numero dei clienti di tale banca}$.

- ▶ Richiamare la finestra di dialogo Calcola variabile.
- ▶ Digitare inclprob_s2 come variabile di destinazione.
- ▶ Digitare 100/ncl come espressione numerica.
- ▶ Fare clic su OK.

Figura 14-6
Finestra di dialogo Calcola variabile



Ora che sono note le probabilità di inclusione per ogni stadio, il calcolo dei pesi di campionamento finali è semplice.

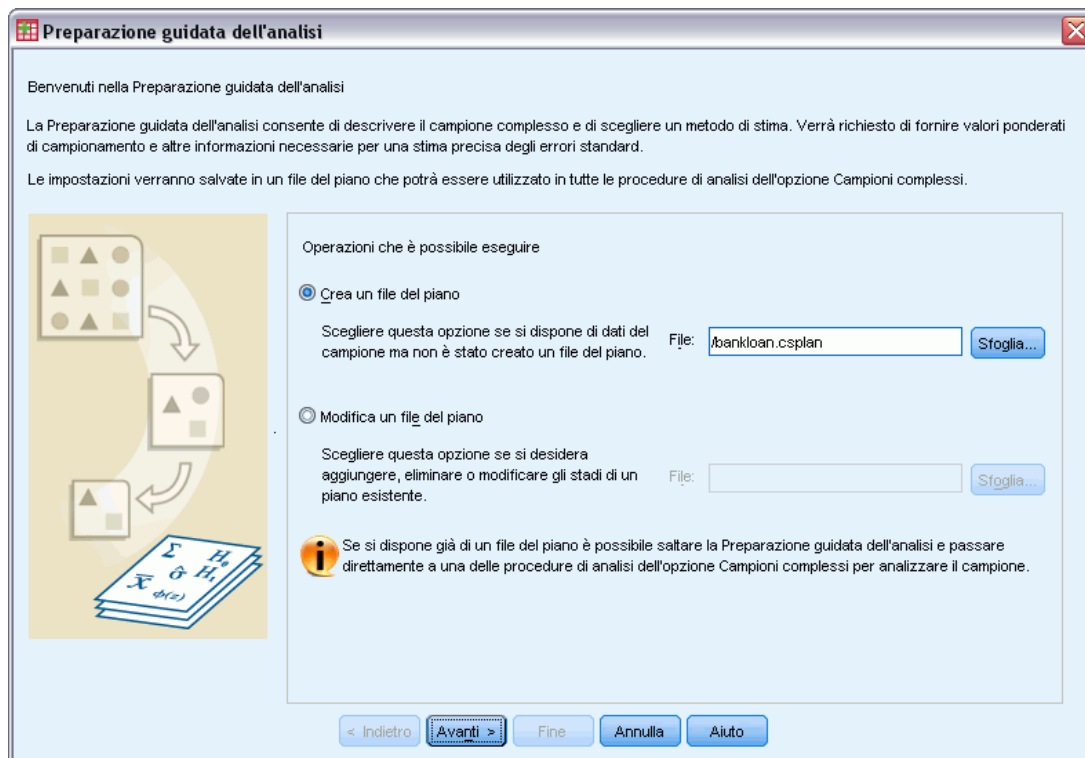
- ▶ Richiamare la finestra di dialogo Calcola variabile.
- ▶ Digitare pesofinale come variabile di destinazione.
- ▶ Digitare $1/(\text{probincl_s1} * \text{probincl_s2})$ come espressione numerica.
- ▶ Fare clic su OK.

È ora possibile creare il piano analisi.

Utilizzo della procedura guidata

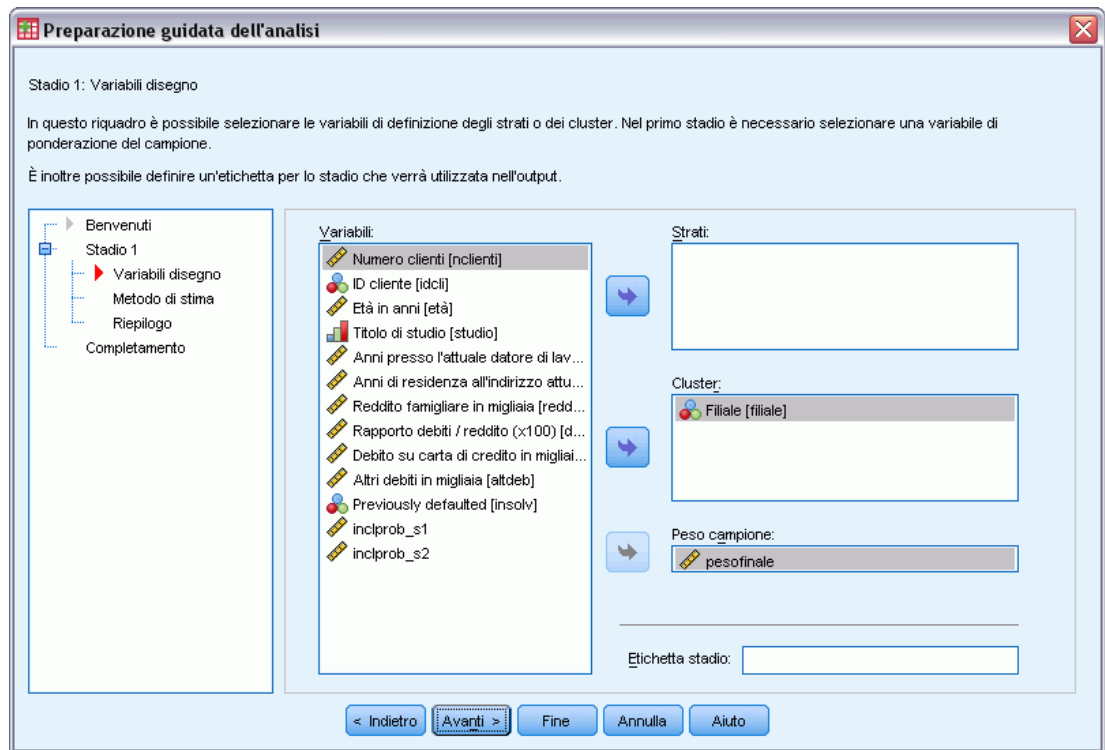
- ▶ Per preparare un campione tramite la procedura guidata Preparazione guidata dell'analisi Campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Prepara per l'analisi...

Figura 14-7
Preparazione guidata dell'analisi, Finestra di benvenuto



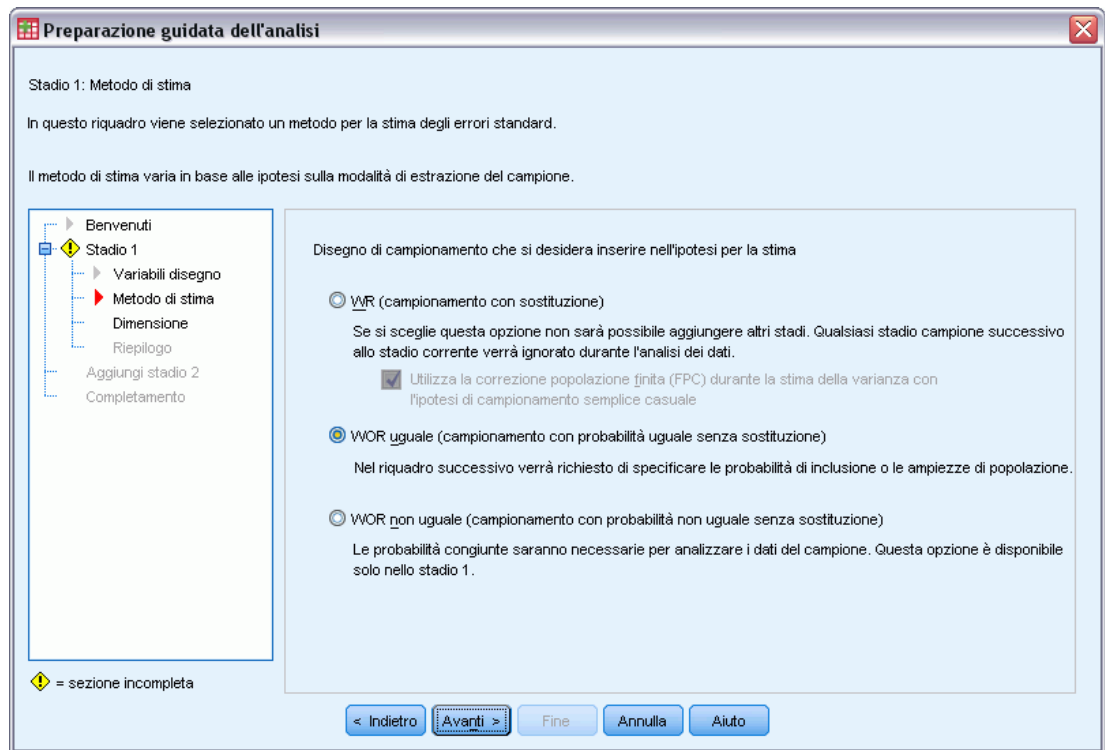
- ▶ Individuare la posizione in cui si desidera salvare il file del piano e digitare bankloan.csplan come nome del file del piano di analisi.
- ▶ Fare clic su Avanti.

Figura 14-8
Preparazione guidata dell'analisi - Passo Variabili disegno (stadio 1)



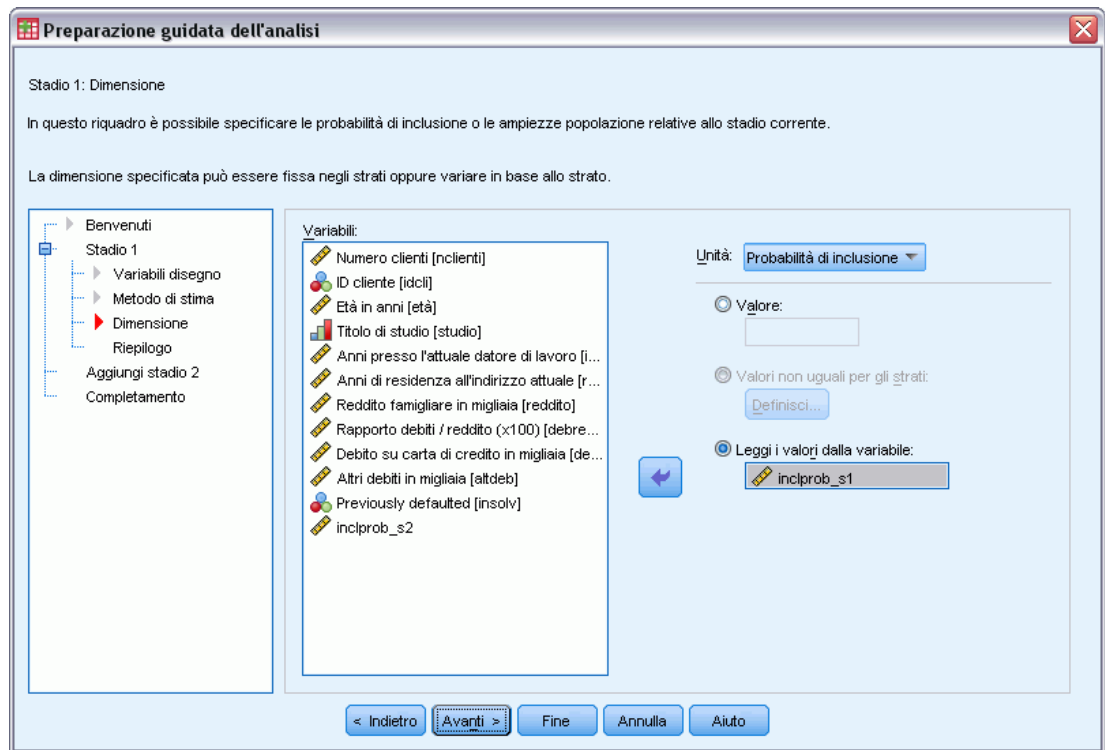
- ▶ Selezionare *Filiale* come variabile di raggruppamento.
- ▶ Selezionare *pesofinale* come variabile di ponderazione del campione.
- ▶ Fare clic su Avanti.

Figura 14-9
Preparazione guidata dell'analisi - Passo Metodo di stima (stadio 1)



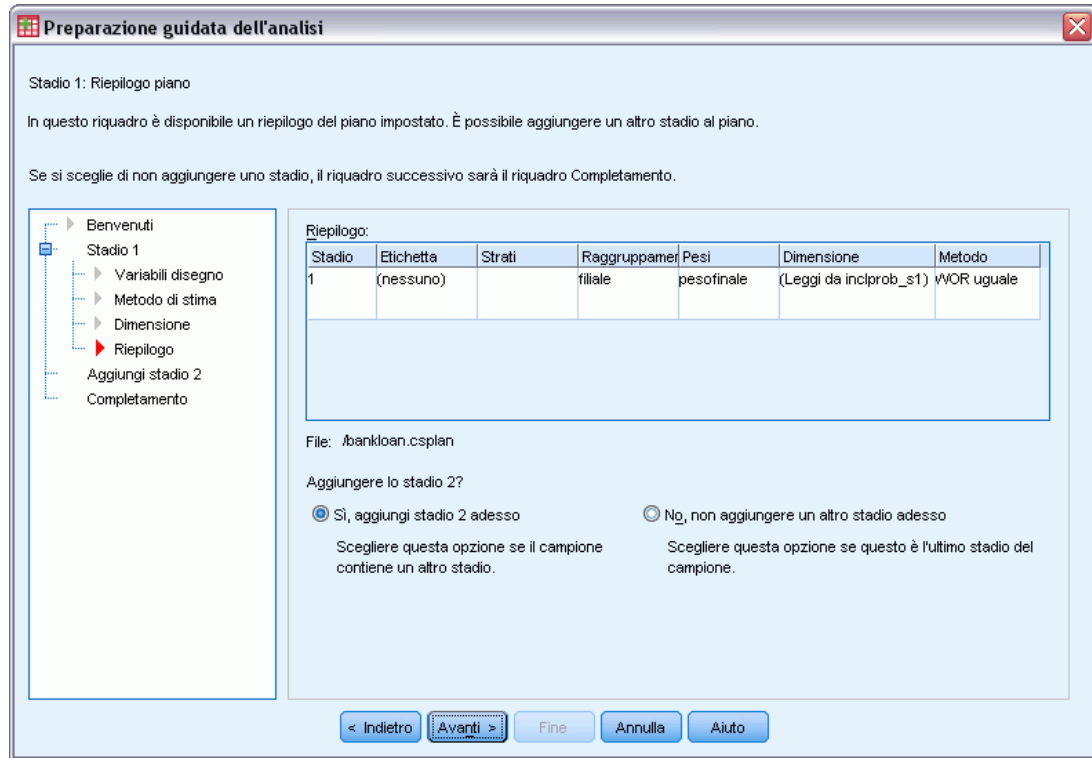
- ▶ Selezionare WOR uguale come metodo di stima per il primo stadio.
- ▶ Fare clic su Avanti.

Figura 14-10
Preparazione guidata dell'analisi - Passo Dimensione (stadio 1)



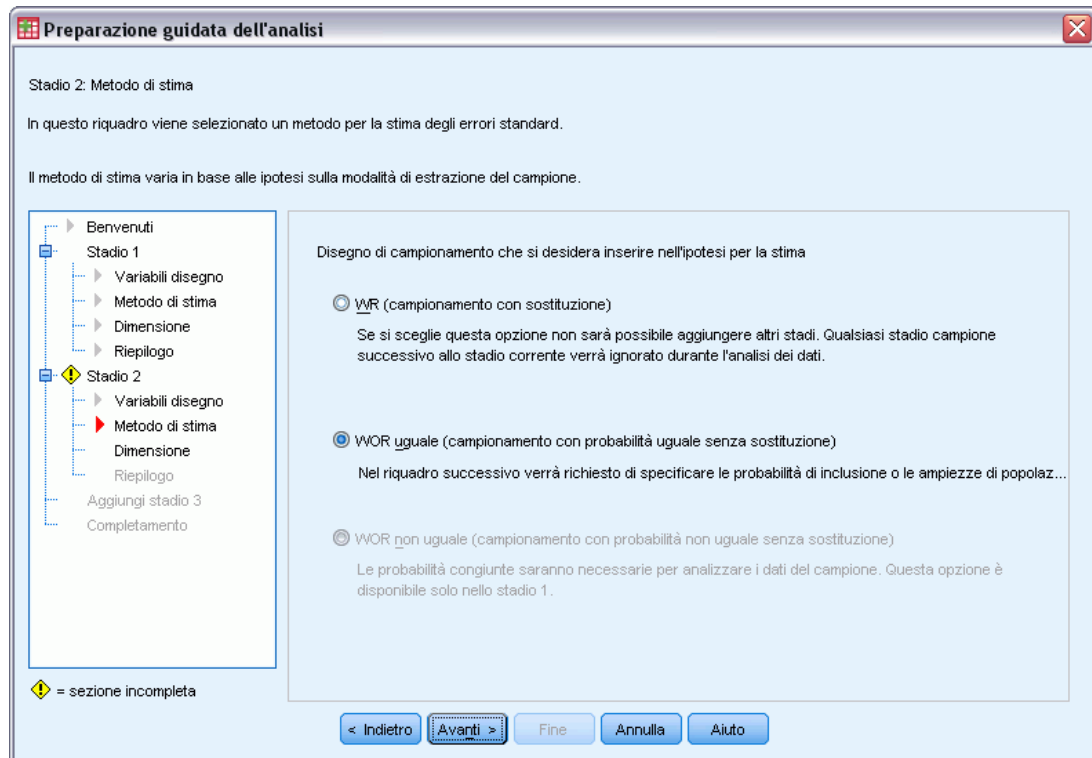
- ▶ Selezionare Leggi i valori dalla variabile e selezionare *inclprob_s1* come variabile che contiene le probabilità di inclusione per il primo stadio.
- ▶ Fare clic su Avanti.

Figura 14-11
Preparazione guidata dell'analisi - Passo Riepilogo del piano (stadio 1)



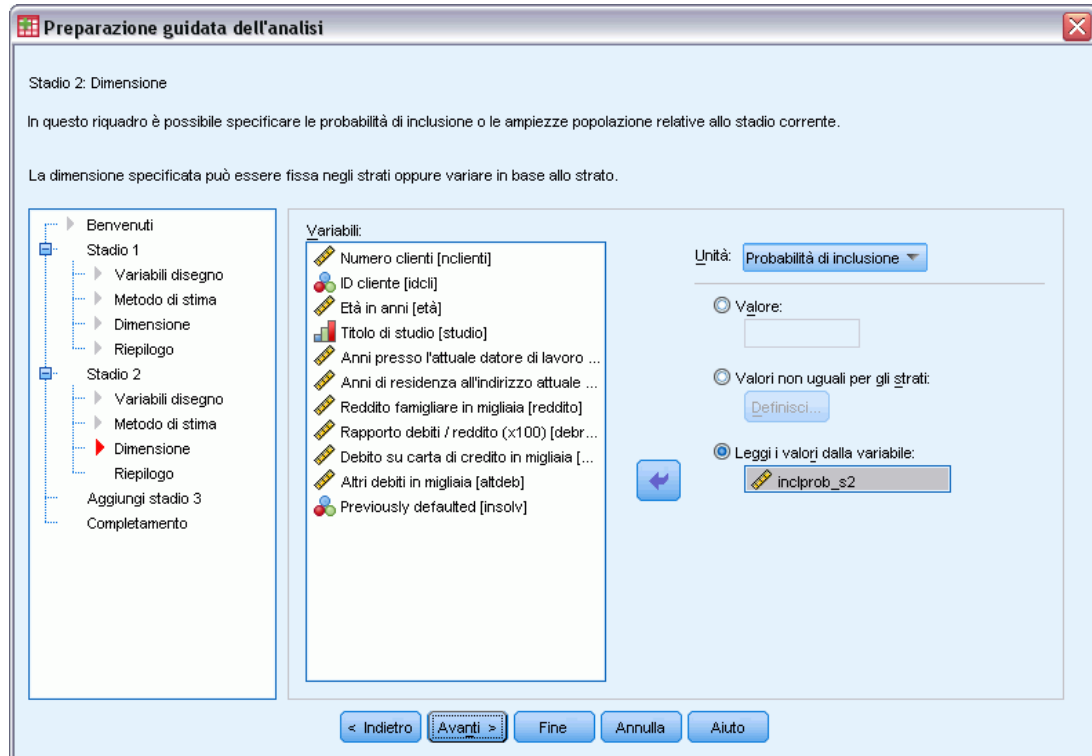
- ▶ Selezionare Sì, aggiungere lo stadio 2 ora.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Variabili disegno.

Figura 14-12
Preparazione guidata dell'analisi - Passo Metodo di stima (stadio 2)



- ▶ Selezionare WOR uguale come metodo di stima per il secondo stadio.
- ▶ Fare clic su Avanti.

Figura 14-13
Preparazione guidata dell'analisi - Passo Dimensione (stadio 2)



- ▶ Selezionare Leggi i valori dalla variabile e selezionare *inclprob_s2* come variabile che contiene le probabilità di inclusione per il secondo stadio.
- ▶ Fare clic su Fine.

Riepilogo

Figura 14-14
Tabella riassuntiva

			Stadio 1	Stadio 2
Variabili disegno	Raggruppamento	1	Filiale	
Informazioni sull'analisi	Raggruppamento	Ipotesi stimatore	Campionamento con probabilità uguale senza sostituzione	Campionamento con probabilità uguale senza sostituzione
		Probabilità di inclusione	Ottenuto dalla variabile <i>inclprob_s1</i>	Ottenuto dalla variabile <i>inclprob_s2</i>

File del piano: C:\bankloan.csaplan
 Variabile peso: finalweight
 Stimatore SRS: Campionamento senza sostituzione

La tabella riassuntiva consente di verificare il piano analisi. Il piano è formato da due stadi caratterizzate dal disegno di una variabile di raggruppamento. Viene utilizzata la stima senza sostituzione (WOR) con probabilità uguale e il piano viene salvato in *c:\bankloan.csaplan*. A questo punto è possibile utilizzare il file del piano per elaborare il file *bankloan_noweights.sav* (con le probabilità di inclusione e i pesi di campionamento calcolati) tramite le procedure di analisi di Campioni complessi.

Procedure correlate

La procedura Preparazione guidata dell'analisi Campioni complessi è uno strumento utile per la lettura di un campione ai fini dell'analisi quando non è possibile accedere al file del piano di campionamento.

- Per creare un file del piano di campionamento ed estrarre un campione, utilizzare [Campionamento guidato](#).

Frequenze campioni complessi

La procedura Frequenze campioni complessi consente di ottenere tabelle di frequenza per le variabili selezionate e di visualizzare statistiche univariate. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Utilizzo di Frequenze campioni complessi per analizzare il consumo dei supplementi nutrizionali

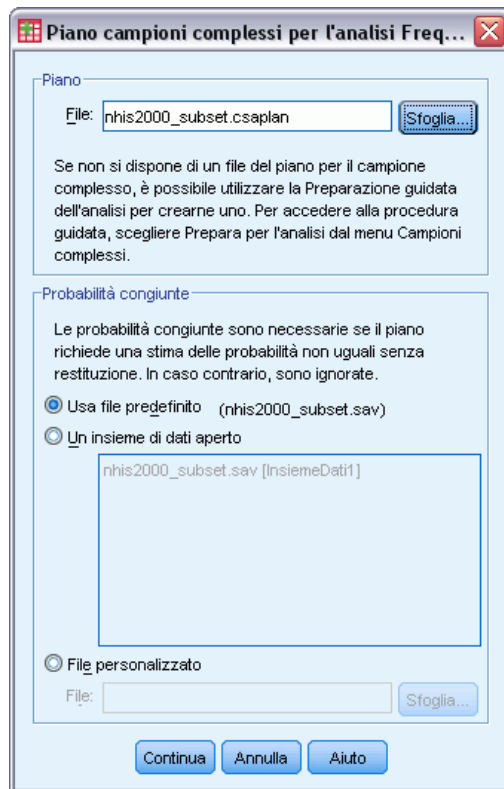
Si supponga che un ricercatore desideri studiare il consumo di supplementi nutrizionali tra la popolazione civile americana, e a tale scopo ricorre ai risultati del sondaggio National Health Interview Survey (NHIS) e al piano analisi precedentemente creato. [Per ulteriori informazioni, vedere l'argomento Utilizzo della Preparazione guidata dell'analisi Campioni complessi per leggere i dati pubblici NHIS in il capitolo 14 a pag. 145.](#)

Un sottoinsieme del sondaggio realizzato nel 2000 è contenuto nel file *nhis2000_subset.sav*. Il piano analisi viene memorizzato in *nhis2000_subset.csaplan*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20.](#) Utilizzare Frequenze campioni complessi per produrre statistiche relative al consumo di supplementi nutrizionali.

Esecuzione dell'analisi

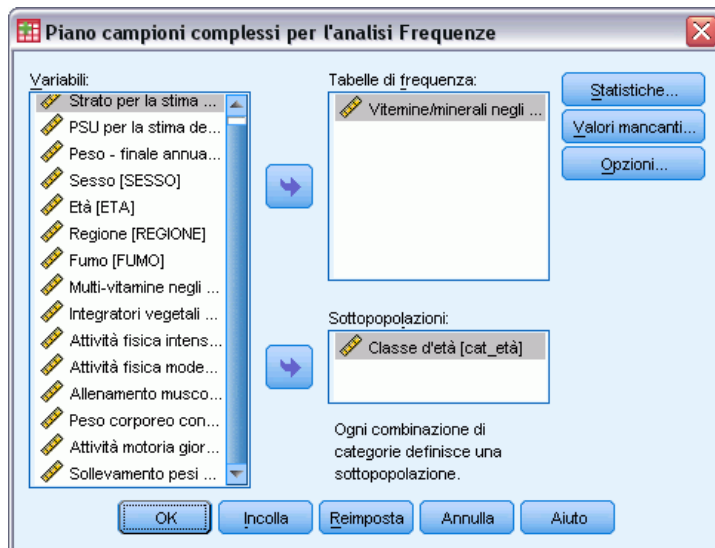
- Per eseguire un'analisi Frequenze campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Frequenze...

Figura 15-1
Finestra di dialogo Piano campioni complessi



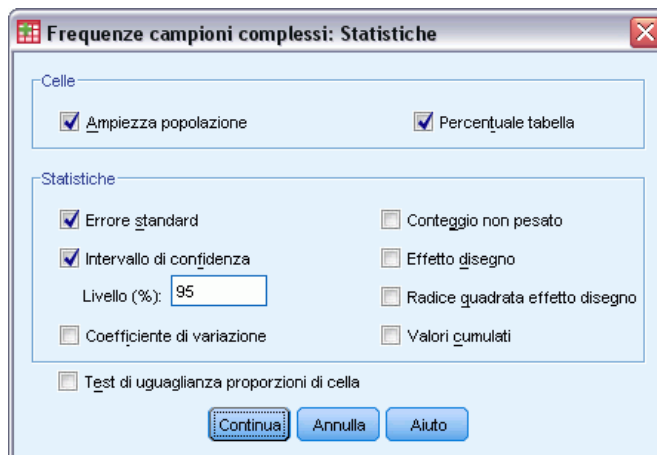
- ▶ Accedere alla finestra e selezionare *nhis2000_subset.csaplan*. Per ulteriori informazioni, vedere l'argomento *File di esempio* in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Fare clic su *Continua*.

Figura 15-2
Finestra di dialogo Frequenze



- ▶ Selezionare *Vitamin/mineral supplmnts-past 12 m* come variabile di frequenza.
- ▶ Selezionare *Categoria età* come variabile di sottopopolazione.
- ▶ Fare clic su Statistiche.

Figura 15-3
Finestra di dialogo Frequenze: Statistiche



- ▶ Selezionare Percentuale tabella nel gruppo Celle.
- ▶ Selezionare Intervallo di confidenza nel gruppo Statistiche.
- ▶ Fare clic su Continua.
- ▶ Fare clic su OK nella finestra di dialogo Frequenze.

Tabella di frequenza (Analisi delle corrispondenze)

Figura 15-4
Tabella di frequenza per variabile/situazione

		Stima	Errore standard	Intervallo di confidenza 95%	
				Inferiore	Superiore
Ampiezza popolazione	Tutti i giorni	102767095	1185127	100435967	105098223
	Alcuni giorni	90794234	1094402	88641560	92946908
	Totale	193561329	1789099	190042196	197080462
% del totale	Tutti i giorni	53.1%	0.4%	52%	54%
	Alcuni giorni	46.9%	0.4%	46.2%	47.6%
	Totale	100.0%	0.0%	100.0%	100.0%

Per ogni misura di cella selezionata viene calcolata ogni statistica selezionata. La prima colonna contiene le stime del numero e della percentuale della popolazione che consuma o meno supplementi di minerali o vitamine. Gli intervalli di confidenza non si sovrappongono, quindi ne deriva che, complessivamente, la maggior parte degli americani consuma supplementi di minerali e vitamine.

Frequenza per sottopopolazione

Figura 15-5
tabella di frequenza per sottopopolazione

Classe d'età			Stima	Errore standard	Intervallo di confidenza 95%	
					Inferiore	Superiore
18-24	Ampiezza popolazione	Tutti i giorni	10018312.000	350602.352	9328681.921	10707942.079
		Alcuni giorni	15472368.000	499182.391	14490482.996	16454253.004
		Totale	25490680.000	680732.812	24151687.776	26829672.224
	% del totale	Tutti i giorni	39.3%	1.0%	37.4%	41.2%
		Alcuni giorni	60.7%	1.0%	58.8%	62.6%
Totale		100.0%	.0%	100.0%	100.0%	
25-44	Ampiezza popolazione	Tutti i giorni	39163840.000	660855.719	37863945.748	40463734.252
		Alcuni giorni	39503150.000	645934.187	38232606.200	40773693.800
		Totale	78666990.000	961114.325	76776491.135	80557488.865
	% del totale	Tutti i giorni	49.8%	.6%	48.7%	50.9%
		Alcuni giorni	50.2%	.6%	49.1%	51.3%
Totale		100.0%	.0%	100.0%	100.0%	
45-64	Ampiezza popolazione	Tutti i giorni	34154952.000	598603.728	32977506.572	35332397.428
		Alcuni giorni	24005512.000	497723.833	23026495.959	24984528.041
		Totale	58160464.000	814680.415	56557998.654	59762929.346
	% del totale	Tutti i giorni	58.7%	.6%	57.5%	60.0%
		Alcuni giorni	41.3%	.6%	40.0%	42.5%
Totale		100.0%	.0%	100.0%	100.0%	
65+	Ampiezza popolazione	Tutti i giorni	19429991.000	439459.793	18565579.536	20294402.464
		Alcuni giorni	11813204.000	314238.078	11195101.955	12431306.045
		Totale	31243195.000	587623.439	30087347.652	32399042.348
	% del totale	Tutti i giorni	62.2%	.7%	60.7%	63.6%
		Alcuni giorni	37.8%	.7%	36.4%	39.3%
Totale		100.0%	.0%	100.0%	100.0%	

Nel calcolo delle statistiche per sottopopolazione, ogni statistica selezionata viene calcolata per ogni misura di cella selezionata in base al valore *Categoria età*. La prima colonna contiene le stime del numero e della percentuale della popolazione che consuma o meno supplementi

di minerali o vitamine. Gli intervalli di confidenza relativi alle percentuali della tabella sono tutti non sovrapponibili. Si può quindi concludere che l'impiego di supplementi di vitamine o minerali aumenta con l'età.

Riepilogo

L'utilizzo della procedura Frequenze campioni complessi consente di ottenere statistiche relative al consumo di supplementi nutrizionali tra la popolazione civile americana.

- Nel complesso, il numero di americani che consuma supplementi di vitamine e minerali è maggiore del numero di americani che non ne fa uso.
- Se suddivisi per categoria età, in termini proporzionali aumenta il numero di americani con età crescenti che consumano supplementi di vitamine e minerali.

Procedure correlate

La procedura Frequenze campioni complessi è uno strumento utile per ottenere statistiche descrittive univariate delle variabili categoriali relative alle osservazioni ottenute dal disegno di campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.
- **Preparazione guidata dell'analisi Campioni complessi** viene utilizzata per impostare le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura **Tavole di contingenza campioni complessi** fornisce statistiche descrittive per variabili di scala.
- La procedura **Descrittive campioni complessi** fornisce statistiche descrittive per variabili di scala.

Descrittive campioni complessi

La procedura Descrittive campioni complessi consente di visualizzare statistiche riassuntive univariate per numerose variabili. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Utilizzo di Descrittive campioni complessi per l'analisi dei livelli di attività

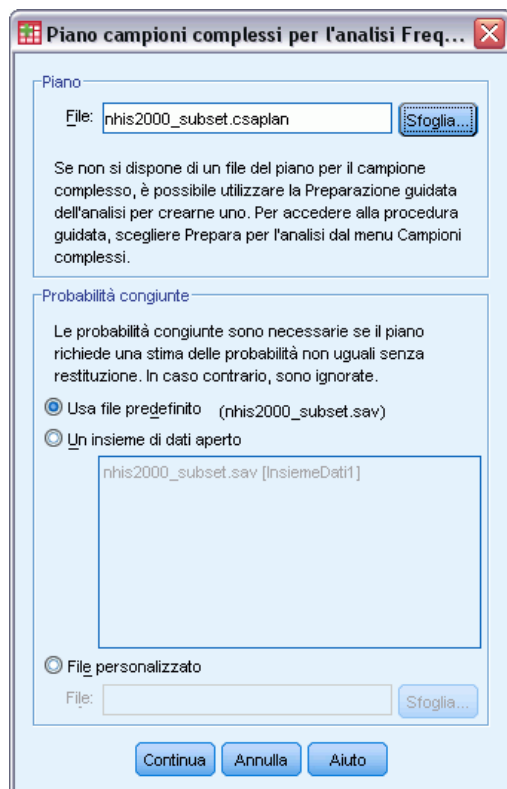
Si supponga che un ricercatore desideri studiare i livelli di attività nella popolazione civile americana, e a tale scopo ricorre ai risultati del sondaggio National Health Interview Survey (NHIS) e al piano analisi precedentemente creato. [Per ulteriori informazioni, vedere l'argomento Utilizzo della Preparazione guidata dell'analisi Campioni complessi per leggere i dati pubblici NHIS in il capitolo 14 a pag. 145.](#)

Un sottoinsieme del sondaggio realizzato nel 2000 è contenuto nel file *nhis2000_subset.sav*. Il piano analisi viene memorizzato in *nhis2000_subset.csaplan*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20.](#) Utilizzare Descrittive campioni complessi per produrre statistiche descrittive univariate relative ai livelli di attività.

Esecuzione dell'analisi

- ▶ Per eseguire un'analisi Descrittive campioni complessi, dai menu scegliere.
Analizza > Campioni complessi > Descrittive...

Figura 16-1
Finestra di dialogo Piano campioni complessi



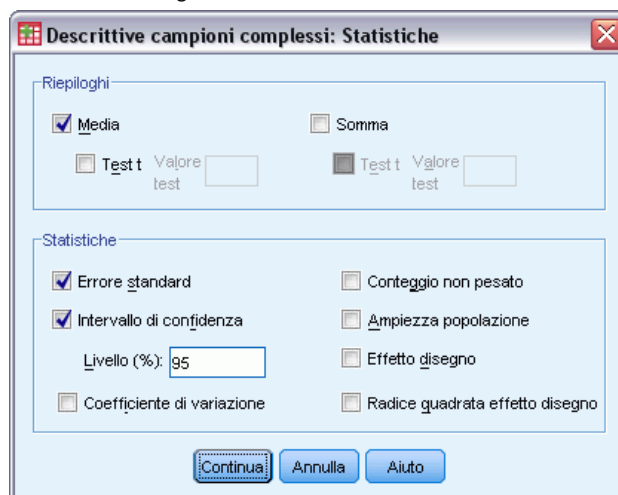
- ▶ Accedere alla finestra e selezionare *nhis2000_subset.csaplan*. Per ulteriori informazioni, vedere l'argomento *File di esempio* in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Fare clic su *Continua*.

Figura 16-2
Descrittive



- ▶ Selezionare *Frequenza attività vigorosa (numero di volte alla settimana)* fino a *Frequenza attività di forza (numero di volte alla settimana)* come variabili di misura.
- ▶ Selezionare *Categoria età* come variabile di sottopopolazione.
- ▶ Fare clic su *Statistiche*.

Figura 16-3
Finestra di dialogo Descrittive: Statistiche



- ▶ Selezionare *Intervallo di confidenza* nel gruppo *Statistiche*.
- ▶ Fare clic su *Continua*.
- ▶ Fare clic su *OK* nella finestra di dialogo *Descrittive campioni complessi*.

Statistiche univariate

Figura 16-4
Statistiche univariate

		Stima	Errore standard	Intervallo di confidenza 95%	
				Inferiore	Superiore
Media	Attività fisica intensa (settimanale)	3.73	.033	3.66	3.79
	Attività fisica moderata (settimanale)	4.90	.041	4.82	4.98
	Allenamento muscolare (settimanale)	3.52	.042	3.43	3.60

Per ogni variabile di misura selezionata viene calcolata ogni statistica selezionata. La prima colonna contiene le stime del numero medio di volte alla settimana in cui una persona realizza una determinata attività fisica. Gli intervalli di confidenza per le medie non si sovrappongono. Pertanto è possibile concludere che, complessivamente, gli americani prediligono le attività fisiche intensive rispetto alle attività fisiche di resistenza e le attività fisiche moderate rispetto a quelle intensive.

Statistiche univariate per sottopopolazione

Figura 16-5
Statistiche univariate per sottopopolazione

Classe d'età			Stima	Errore standard	Intervallo di confidenza 95%	
					Inferiore	Superiore
18-24	Media	Attività fisica intensa (settimanale)	3.92	.087	3.75	4.09
		Attività fisica moderata (settimanale)	5.18	.137	4.91	5.45
		Allenamento muscolare (settimanale)	3.45	.085	3.28	3.62
25-44	Media	Attività fisica intensa (settimanale)	3.55	.048	3.46	3.65
		Attività fisica moderata (settimanale)	4.73	.056	4.62	4.84
		Allenamento muscolare (settimanale)	3.28	.052	3.18	3.38
45-64	Media	Attività fisica intensa (settimanale)	3.79	.063	3.66	3.91
		Attività fisica moderata (settimanale)	4.88	.070	4.74	5.02
		Allenamento muscolare (settimanale)	3.65	.092	3.47	3.84
65+	Media	Attività fisica intensa (settimanale)	4.18	.111	3.96	4.39
		Attività fisica moderata (settimanale)	5.22	.084	5.06	5.39
		Allenamento muscolare (settimanale)	4.66	.155	4.36	4.97

Per ogni variabile di misura viene calcolata ogni statistica selezionata in base a valori di *Categoria età*. La prima colonna contiene le stime del numero medio di volte alla settimana in cui le persone di ogni categoria realizzano una determinata attività fisica. Gli intervalli di confidenza per le medie consentono di formulare alcune interessanti conclusioni.

- In termini di attività fisiche intense o moderate, le persone con 25–44 anni sono meno attive di quelle con 18–24 anni e 45–64, e le persone con età compresa tra 45–64 anni sono meno attive delle persone con 65 anni o più anziane.
- In termini di attività fisiche di resistenza, le persone con 25–44 anni sono meno attive di quelle con 45–64, e le persone con età compresa tra 18–24 e 45–64 sono meno attive delle persone con 65 anni o più anziane.

Riepilogo

L'utilizzo della procedura Descrittive campioni complessi consente di ottenere statistiche relative ai livelli di attività fisica della popolazione civile americana.

- Complessivamente, gli americani trascorrono quantità di tempo variabili in attività fisiche di vario tipo.
- Se suddivisi per età, in linea generale, risulta che gli americani dopo il college inizialmente sono meno attivi di quanto lo fossero durante gli studi ma diventano più coscienti sull'importanza di fare un'attività fisica invecchiando.

Procedure correlate

La procedura Descrittive campioni complessi è uno strumento utile per ottenere statistiche descrittive univariate delle misure di scala relative alle osservazioni ottenute dal disegno di campionamento complesso.

- [Campionamento complesso guidato](#) viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.
- [Preparazione guidata dell'analisi Campioni complessi](#) viene utilizzata per impostare le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura [Descrittive campioni complessi](#) fornisce statistiche descrittive per rapporti di misure di scala.
- La procedura [Descrittive campioni complessi](#) fornisce statistiche descrittive per variabili di scala.

Tavole di contingenza campioni complessi

La procedura Tavole di contingenza campioni complessi consente di creare tavole di contingenza per le coppie di variabili selezionate e di visualizzare statistiche a due vie. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Utilizzo delle tavole di contingenza campioni complessi per misurare il coefficiente di rischio di un evento

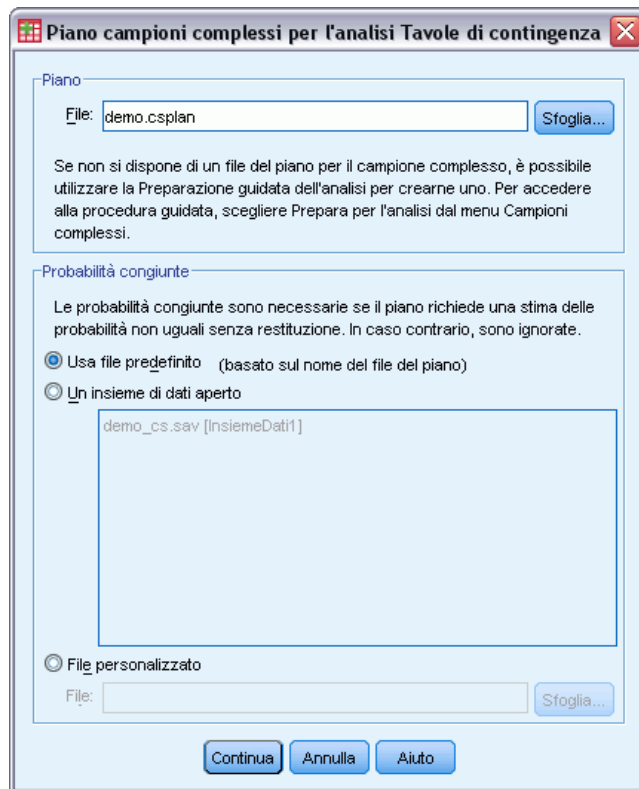
Per tradizione, una società dedicata alla vendita di abbonamenti a riviste invia mensilmente materiale promozionale a un database di nominativi acquistato. Il tasso di risposta in genere è basso, quindi è necessario cercare un sistema migliore per attirare potenziali clienti. Una proposta consiste nel concentrare l'invio di materiale promozionale ai sottoscrittori di quotidiani in quanto si presume che chi legge giornali sia più incline ad abbonarsi a una rivista.

Utilizzare la procedura Tavole di contingenza campioni complessi per verificare la validità di questa teoria. A tal fine, creare una tabella 2 x 2, *Abbonamento a quotidiani * Risposta* e calcolare il coefficiente di rischio in base al quale una persona con un abbonamento a un quotidiano risponderà al materiale promozionale inviato. Tali informazioni vengono raccolte nel file *demo_cs.sav* e devono essere analizzate tramite il file del piano di campionamento *demo.csplan*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20.](#)

Esecuzione dell'analisi

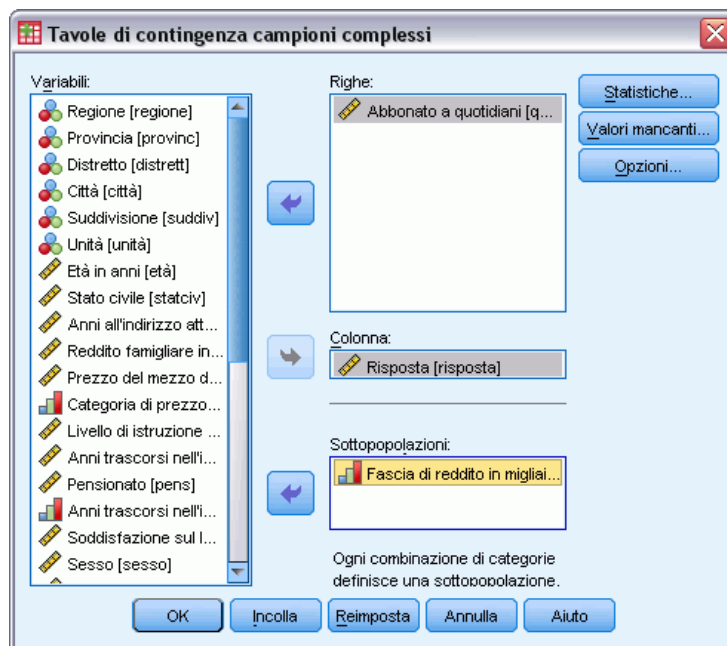
- Per eseguire un'analisi Tavole di contingenza campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Tavole di contingenza...

Figura 17-1
Finestra di dialogo Piano campioni complessi



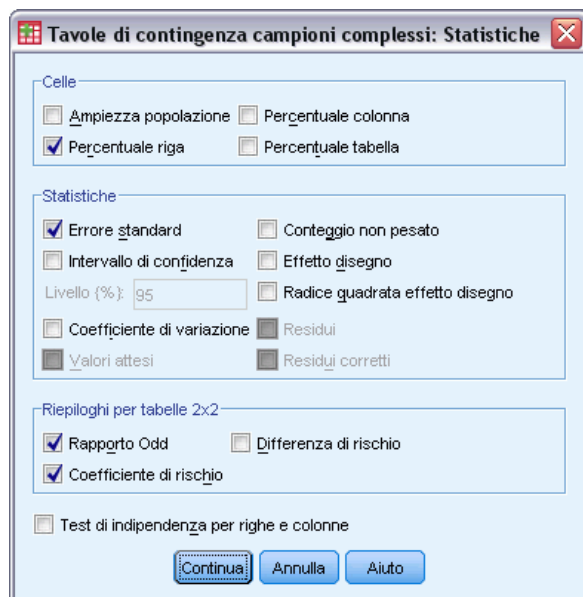
- ▶ Accedere alla finestra e selezionare *demo.csplan*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#).
- ▶ Fare clic su Continua.

Figura 17-2
Finestra di dialogo *Tavole di contingenza*



- ▶ Selezionare *Abbonamento a quotidiani* come variabile di riga.
- ▶ Selezionare *Risposta* come variabile di colonna.
- ▶ È inoltre interessante visualizzare i risultati suddivisi per categorie di reddito, quindi selezionare *Categoria di reddito in migliaia* come variabile di sottopopolazione.
- ▶ Fare clic su *Statistiche*.

Figura 17-3
Finestra di dialogo Tavole di contingenza: Statistiche



- ▶ Deselezionare Ampiezza popolazione e selezionare Percentuale di riga nel gruppo Celle.
- ▶ Selezionare Rapporto Odd e Coefficiente di rischio nel gruppo Riepiloghi per tabelle 2x2.
- ▶ Fare clic su Continua.
- ▶ Fare clic su OK nella finestra di dialogo Tavole di contingenza campioni complessi.

Tramite queste selezioni, verrà creata una tabella tavola di contingenza e una stima del rischio relativa a *Abbonamento a quotidiani *Risposta*. Inoltre vengono create tabelle separate con i risultati divisi in base a *Categoria di reddito in migliaia*.

Tavola di contingenza

Figura 17-4
Tavola di contingenza di abbonamento a quotidiani per risposta

Abbonato a quotidiani			Risposta		
			Sì	No	Totale
Sì	% entro	Stima	17.2%	82.8%	100.0%
	Abbonato a quotidiani	Errore standard	1.0%	1.0%	.0%
No	% entro	Stima	10.3%	89.7%	100.0%
	Abbonato a quotidiani	Errore standard	.7%	.7%	.0%
Totale	% entro	Stima	12.8%	87.2%	100.0%
	Abbonato a quotidiani	Errore standard	.7%	.7%	.0%

La tavola di contingenza mostra che complessivamente il numero di persone che ha risposto al materiale promozionale inviato è basso. Tuttavia, in proporzione ha risposto un numero maggiore di sottoscrittori a quotidiani.

Stima del rischio

Figura 17-5
Stima del rischio di abbonamento a quotidiani per risposta

		Stima
Abbonato a quotidiani	Rapporto Odd	1.812
* Risposta	Coefficiente di rischio	1.673
	Per cohort Response = Sì	
	Per cohort Response = No	.923

Le statistiche vengono calcolate solo per le tabelle 2x2 con tutte le celle osservate.

Il coefficiente di rischio è un rapporto di probabilità di eventi. Il coefficiente di rischio di una risposta a un invio di materiale promozionale è il rapporto tra la probabilità che un sottoscrittore risponda e la probabilità che un non sottoscrittore risponda. Quindi, la stima del coefficiente di rischio è semplicemente $17,2\%/10,3\% = 1,673$. Quindi, il coefficiente di rischio della non risposta è il rapporto tra la probabilità che un sottoscrittore non risponda e la probabilità che un non sottoscrittore non risponda. Nello specifico, la stima di questo coefficiente di rischio è 0,923. In base a questi risultati, è possibile stimare che un sottoscrittore a quotidiani sia 1,673 volte più incline a rispondere al materiale promozionale inviato rispetto a un non sottoscrittore, oppure 0,923 volte più incline di un non sottoscrittore a non rispondere.

Il Rapporto Odd è il rapporto tra gli odd degli eventi. Gli odd di un evento sono il rapporto tra la probabilità che un evento si verifichi e la probabilità che tale evento non si verifichi. Quindi, la stima degli odd relativa alla percentuale di risposta di un sottoscrittore a quotidiani è $17,2\%/82,8\% = 0,208$. Mentre la stima degli odd relativa alla percentuale di risposta di un non sottoscrittore è $10,3\%/89,7\% = 0,115$. La stima degli odd sarà quindi $0,208/0,115 = 1,812$ (tenendo conto del possibile errore di arrotondamento nei vari passaggi). Il Rapporto Odd è inoltre il rapporto tra il coefficiente di rischio di risposta e il coefficiente di rischio di non risposta, ovvero $1,673/0,923 = 1,812$.

Rapporto Odd vs. Coefficiente di rischio

Trattandosi di un rapporto tra proporzioni, il Rapporto Odd è un valore di difficile interpretazione. Il coefficiente di rischio è più facile da interpretare, quindi il Rapporto Odd preso singolarmente non è di grande utilità. Tuttavia, esistono situazioni abbastanza comuni in cui la stima del coefficiente di rischio non è molto valida ed è possibile utilizzare il Rapporto Odd per approssimarsi al coefficiente di rischio dell'evento di interesse. Il Rapporto Odd deve essere utilizzato come approssimazione del coefficiente di rischio dell'evento di interesse quando vengono soddisfatte le condizioni seguenti:

- La probabilità dell'evento di interesse è bassa ($<0,1$). Tale condizione assicura che il Rapporto Odd sia una valida approssimazione del coefficiente di rischio. In questo esempio, l'evento di interesse è la risposta al materiale promozionale inviato.
- Il disegno dello studio è di tipo "caso-controllo". Tale condizione segnala la probabilità che la stima abituale del coefficiente di rischio non sia valida. Uno studio caso-controllo è retrospettivo, spesso viene utilizzato quando l'evento di interesse è improbabile oppure quando il disegno di un esperimento potenziale è impraticabile o non etico.

In questo esempio, nessuna condizione viene soddisfatta in quanto la proporzione complessiva dei rispondenti era 12,8% e il disegno dello studio non era di tipo “caso-controllo”, quindi è più sicuro prendere in considerazione il valore di 1,673 come coefficiente di rischio, anziché il valore del Rapporto Odd.

Stima del rischio in base alla sottopopolazione

Figura 17-6

Stima del rischio di abbonamento quotidiani per risposta, con controllo sulla categoria di reddito

Anni trascorsi nell'impiego attuale				Stima
Under \$25	Abbonato a quotidiani * Risposta	Rapporto Odd		2.712
		Coefficiente di rischio	Per cohort Response = Sì	2.241
			Per cohort Response = No	.826
\$25 - \$49	Abbonato a quotidiani * Risposta	Rapporto Odd		1.794
		Coefficiente di rischio	Per cohort Response = Sì	1.645
			Per cohort Response = No	.917
\$50 - \$74	Abbonato a quotidiani * Risposta	Rapporto Odd		1.168
		Coefficiente di rischio	Per cohort Response = Sì	1.152
			Per cohort Response = No	.986
\$75+	Abbonato a quotidiani * Risposta	Rapporto Odd		1.242
		Coefficiente di rischio	Per cohort Response = Sì	1.227
			Per cohort Response = No	.988

Le statistiche vengono calcolate solo per le tabelle 2x2 con tutte le celle osservate.

Le stime del coefficiente di rischio vengono calcolate separatamente per ogni categoria di reddito. Si noti che il coefficiente di rischio di una risposta positiva da parte dei sottoscrittori a quotidiani sembra diminuire gradualmente all'aumentare del reddito, a indicare che è possibile definire con precisione il target dell'invio di materiale promozionale.

Riepilogo

Tramite le stime di rischio ottenute dalle tavole di contingenza campioni complessi, è possibile aumentare il tasso di risposta agli invii di materiale promozionale, definendo come target gli abbonati a quotidiani. Inoltre, viene messo in evidenza che le stime del rischio possono non essere costanti tra le *Categorie di reddito*, quindi non è possibile aumentare il tasso di risposta ulteriormente definendo come target i sottoscrittori di quotidiani con reddito inferiore.

Procedure correlate

La procedura Tavole di contingenza campioni complessi è uno strumento utile per realizzare statistiche descrittive delle tavole di contingenza delle variabili categoriali relative alle osservazioni ottenute dal disegno di un campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.
- **Preparazione guidata dell'analisi Campioni complessi** viene utilizzata per impostare le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura **Descrittive campioni complessi** fornisce statistiche descrittive per variabili di scala.

Rapporti campioni complessi

La procedura Rapporti campioni complessi consente di visualizzare statistiche riassuntive univariate per rapporti di variabili. Se necessario, è possibile richiedere statistiche basate su sottogruppi definiti in base a una o più variabili categoriali.

Utilizzo dei rapporti campioni complessi per agevolare la valutazione del valore della proprietà

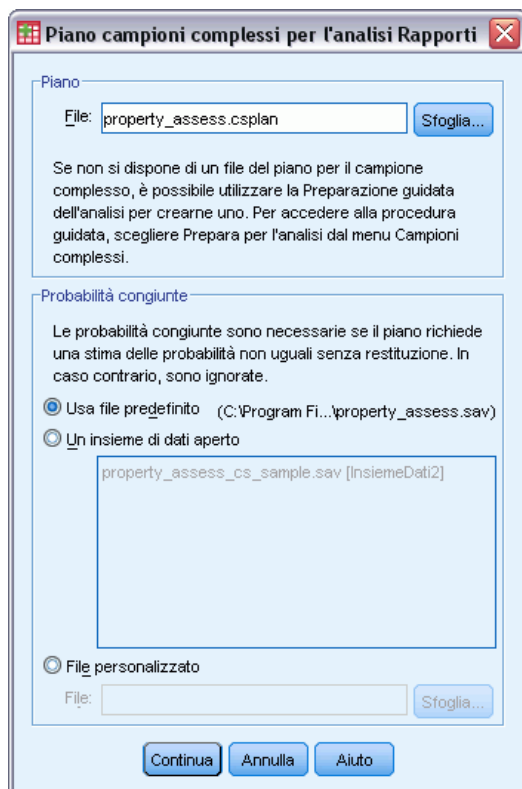
Un'agenzia statale deve garantire che le imposte sulla proprietà siano correttamente valutate nelle diverse zone. Le imposte si basano sul valore stimato della proprietà, quindi l'agenzia desidera tenere traccia dei valori della proprietà nelle diverse zone per essere certa che i record di ogni zona siano aggiornati allo stesso modo. Poiché le risorse per ottenere le attuali stime sono limitate, l'agenzia ha deciso di utilizzare una metodologia di campionamento complessa per selezionare le proprietà.

Il campione delle proprietà selezionate e le attuali informazioni sulla stima corrispondente sono contenute nel file *property_assess_cs_sample.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Utilizzare i Rapporti campioni complessi per valutare le modifiche nei valori della proprietà in cinque zone diverse dall'ultima stima realizzata.

Esecuzione dell'analisi

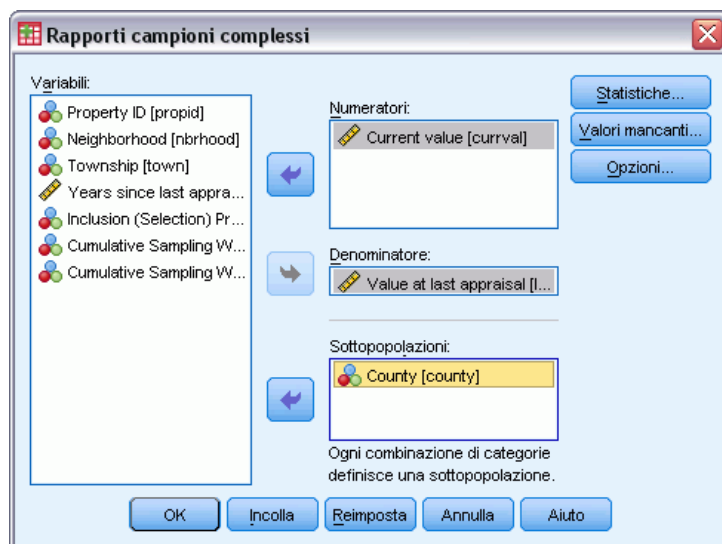
- ▶ Per eseguire un'analisi Rapporti campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Rapporti...

Figura 18-1
Finestra di dialogo Piano campioni complessi



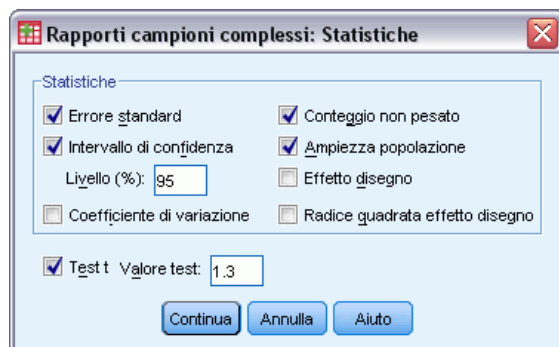
- ▶ Accedere alla finestra e selezionare *property_assess.csplan*. Per ulteriori informazioni, vedere l'argomento *File di esempio* in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Fare clic su *Continua*.

Figura 18-2
Finestra di dialogo Rapporti



- ▶ Selezionare *Valore corrente* come variabile numeratore.
- ▶ Selezionare *Valore all'ultima valutazione* come variabile denominatore.
- ▶ Selezionare *Paese* come variabile di sottopopolazione.
- ▶ Fare clic su *Statistiche*.

Figura 18-3
Finestra di dialogo Statistiche rapporti



- ▶ Selezionare Intervallo di confidenza, Conteggio non ponderato e Ampiezza popolazione nel gruppo Statistiche.
- ▶ Selezionare Test t e specificare 1,3 come valore del test.
- ▶ Fare clic su Continua.
- ▶ Fare clic su OK nella finestra di dialogo Rapporti campioni complessi.

Rapporti

Figura 18-4
tabella Rapporti

Zona	Numeratore	Denominatore	Stima del rapporto	Errore standard	Intervallo di confidenza 95%	
					Inferiore	Superiore
Orientale	Valore	Valore all'ultima valutazione	1,381	,068	1,236	1,525
Centrale	Valore	Valore all'ultima valutazione	1,364	,064	1,227	1,502
Occidentale	Valore	Valore all'ultima valutazione	1,524	,053	1,410	1,638
Settentrionale	Valore	Valore all'ultima valutazione	1,277	,032	1,208	1,346
Meridionale	Valore	Valore all'ultima valutazione	1,195	,029	1,134	1,256

La visualizzazione predefinita della tabella é molto ampia, quindi sarà necessario eseguire il pivot per migliorare la visualizzazione.

Eeguire il pivot della tabella Rapporti

- ▶ Fare doppio clic nella tabella pivot per attivarla.
- ▶ Dai menu del Viewer scegliere:
Pivot > Pivoting
- ▶ Trascinare *Numeratore* e quindi *Denominatore* dalla riga allo strato.
- ▶ Trascinare *Zona* dalla riga alla colonna.
- ▶ Trascinare *Statistiche* dalla colonna alla riga.
- ▶ Chiudere la finestra Pivoting.

Tabella pivot Rapporti

Figura 18-5
Tabella pivot Rapporti

		Zona				
		Orientale	Centrale	Occidentale	Settentrionale	Meridionale
Numeratore	Valore					
Denominatore	Valore all'ultima valutazione					
Stima del rapporto		1.381	1.384	1.524	1.277	1.195
Errore standard		.068	.068	.053	.032	.029
Intervallo di confidenza 95%	Inferiore	1.236	1.227	1.410	1.208	1.134
	Superiore	1.525	1.502	1.638	1.346	1.256
Test di ipotesi	Valore del test	1.3	1.3	1.3	1.3	1.3
	t	1.191	.997	4.201	-.702	-3.646
	df	15	15	15	15	15
	Sig.	.252	.334	.001	.493	.002
Ampiezza popolazione						
Conteggio non pesato		168	179	202	205	220

E' stato eseguito il pivot della tabella Rapporti per facilitare il confronto tra statistiche di zone diverse.

- Le stime dei rapporti variano da un valore basso di 1,195 nella zona meridionale a un valore alto di 1,524 nella zona occidentale
- Negli errori standard e' presente una leggera variabilità, che varia da un valore basso di 0,029 nella zona meridionale a 0,068 nella zona orientale.
- Alcuni degli intervalli di confidenza non si sovrappongono. Pertanto, è possibile concludere che i rapporti relativi alla zona occidentale sono maggiori dei rapporti esistenti nelle zone meridionali e settentrionali.
- Infine, come misura più oggettiva, si noti che i valori di significatività dei test *t* relativi alle zone occidentali e meridionali sono inferiori a 0,05. Quindi, è possibile concludere che il rapporto per la zona occidentale è maggiore di 1,3 e il rapporto per la zona meridionale è inferiore a 1,3.

Riepilogo

Tramite la procedura Rapporti campioni complessi, si sono ottenute varie statistiche per i rapporti di *Valore corrente* fino a *Valore all'ultima valutazione*. I risultati suggeriscono che potrebbero sussistere disuguaglianze nella valutazione delle imposte della proprietà tra le diverse zone.

Nello specifico:

- I rapporti relativi alla zona occidentale sono elevati. Ciò significa che i record corrispondenti non sono aggiornati come quelli di altre zone per quanto riguarda la valutazione dei valori della proprietà. Le imposte della proprietà sono probabilmente troppo basse in questa zona.
- I rapporti relativi alla zona meridionale sono bassi. Ciò significa che i record corrispondenti sono più aggiornati di quelli di altre zone per quanto riguarda la valutazione dei valori della proprietà. Le imposte della proprietà sono probabilmente troppo alte in questa zona.
- I rapporti relativi alla zona meridionale sono inferiori a quelli della zona occidentale, ma ancora compresi nell'obiettivo di 1,3.

Le risorse utilizzate per tenere traccia dei valori della proprietà nella zona meridionale verranno riassegnate alla zona occidentale per allineare i rapporti di queste zone con gli altri e con l'obiettivo di 1,3.

Procedure correlate

La procedura Rapporti campioni complessi e' uno strumento utile per ottenere statistiche descrittive univariate del rapporto di misure di scala relativo alle osservazioni ottenute dal disegno di campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.

- [Preparazione guidata dell'analisi Campioni complessi](#) viene utilizzata per impostare le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura [Descrittive campioni complessi](#) fornisce statistiche descrittive per variabili di scala.

Modello lineare generalizzato campioni complessi

La procedura Modello lineare generalizzato campioni complessi (CSGLM) esegue l'analisi di regressione lineare, nonché l'analisi della varianza e della covarianza su campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Utilizzo del Modello lineare generalizzato campioni complessi per l'adattamento a un'ANOVA a due fattori

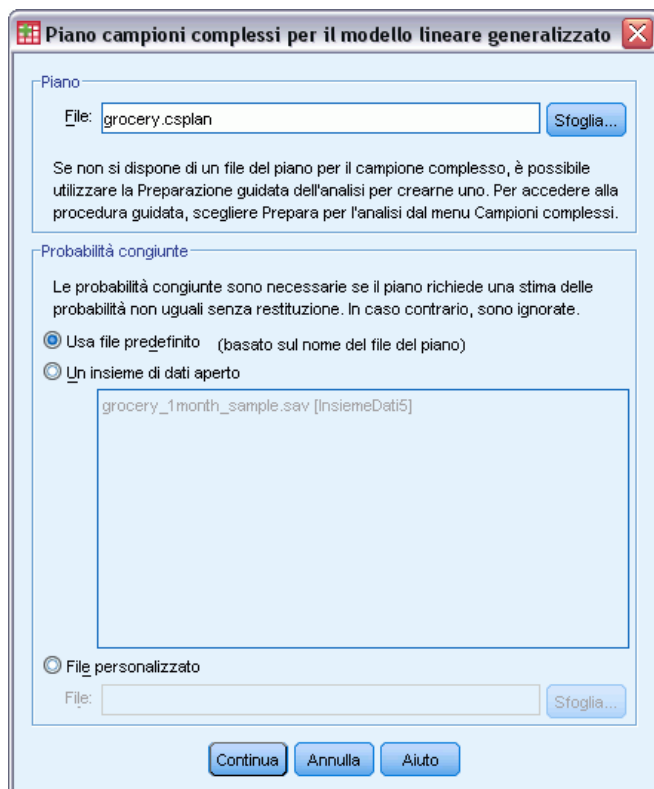
Una catena di drogherie ha effettuato un'indagine su un gruppo di clienti in relazione alle loro abitudini di acquisto, in base a un disegno complesso. Dati i risultati dell'indagine e la spesa di ogni cliente nel mese precedente, si vuole verificare se la frequenza di acquisto della clientela è correlata all'importo di spesa mensile, controllando il sesso del cliente e inserendo il disegno di campionamento.

Tali informazioni vengono raccolte nel file *grocery_1month_sample.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20.](#) Utilizzo del Modello lineare generalizzato campioni complessi per l'esecuzione di un'ANOVA a due fattori sugli importi di spesa.

Esecuzione dell'analisi

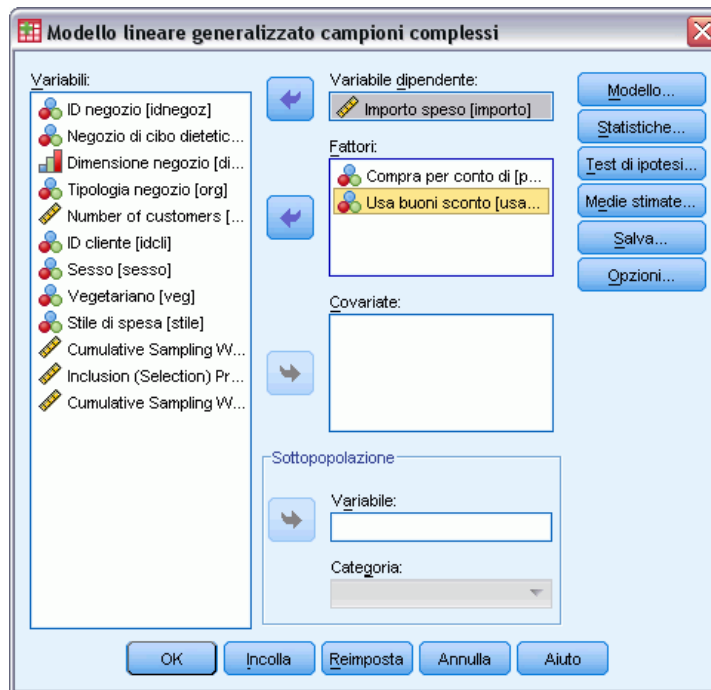
- Per eseguire un'analisi Modello lineare generalizzato campioni complessi, dai menu scegliere:
Analizza > Campioni complessi > Modello lineare generalizzato...

Figura 19-1
Finestra di dialogo *Piano campioni complessi*



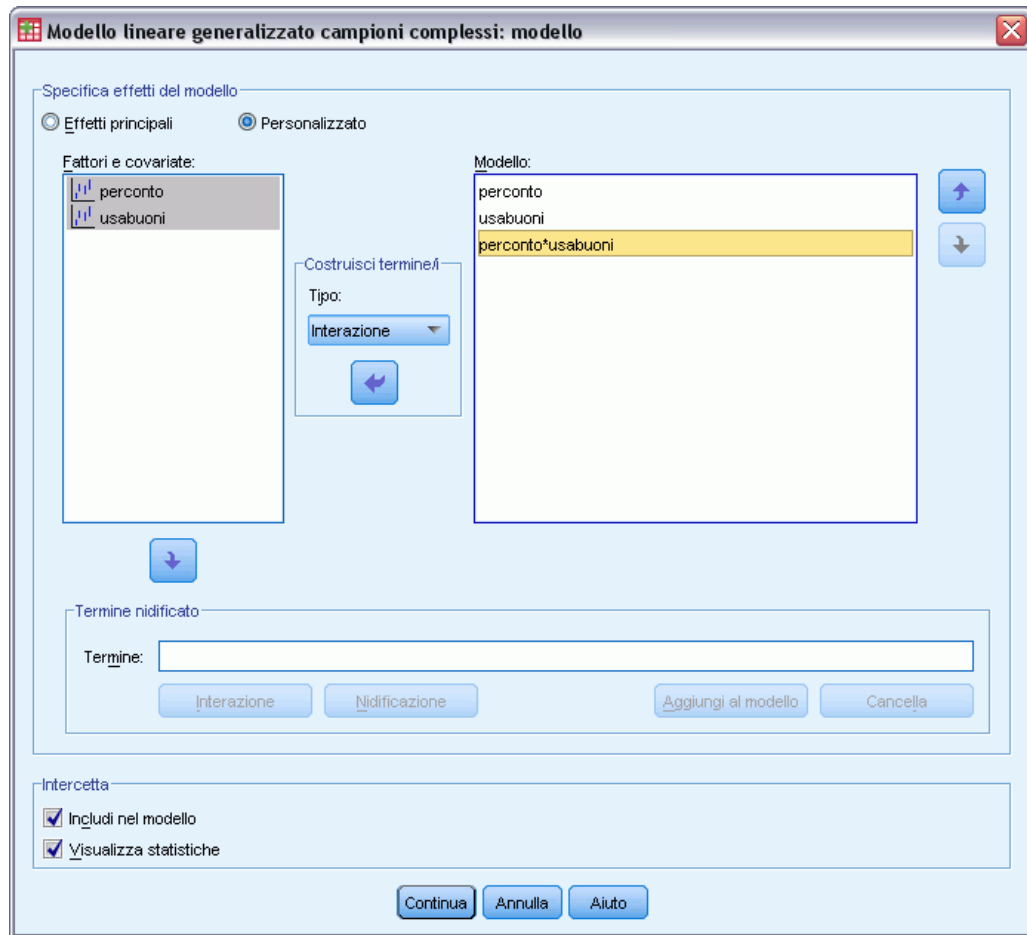
- ▶ Accedere alla finestra e selezionare il file *grocery.csplan*. Per ulteriori informazioni, vedere l'argomento *File di esempio* in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Fare clic su *Continua*.

Figura 19-2
Finestra di dialogo Modello lineare generalizzato



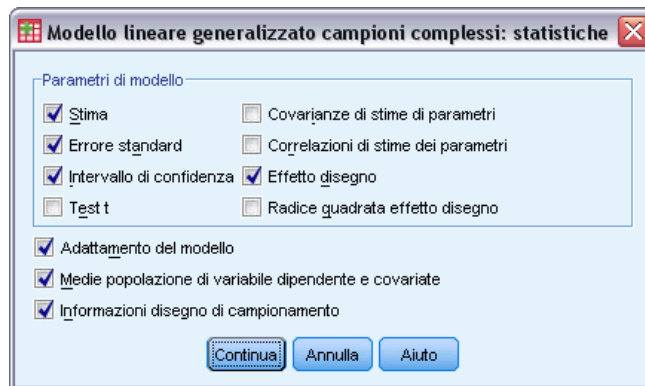
- ▶ Selezionare *Importo di spesa* come variabile dipendente.
- ▶ Selezionare *Destinatari spesa* e *Utilizzo buoni sconto* come fattori.
- ▶ Fare clic su *Modello*.

Figura 19-3
Finestra di dialogo Modello



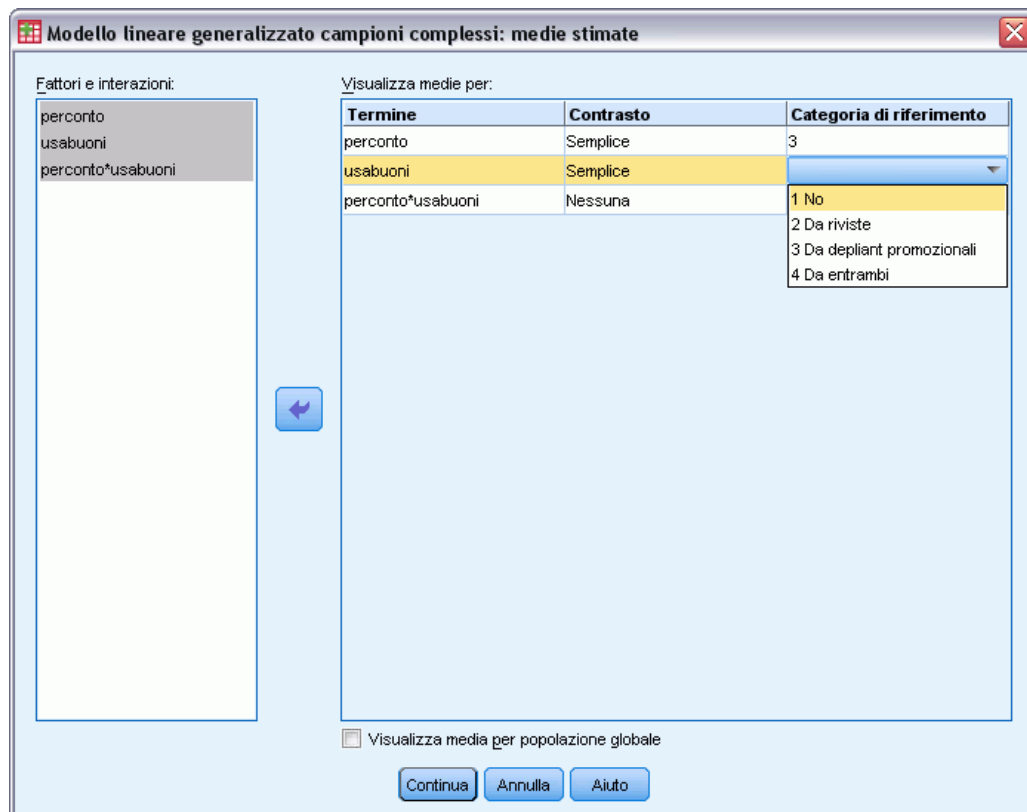
- ▶ Scegliere di generare un modello Personalizzato.
- ▶ Selezionare Effetti principali come tipo di termine da costruire e *destinatario* e *buoni* come termini del modello.
- ▶ Selezionare Interazione come tipo di termine da costruire e aggiungere l'interazione *destinatario*buoni* come termine del modello.
- ▶ Fare clic su Continua.
- ▶ Fare clic su Statistiche nella finestra di dialogo Modello lineare generalizzato.

Figura 19-4
Finestra di dialogo Statistiche modello lineare generalizzato



- ▶ Selezionare Stima, Errore standard, Intervallo di confidenza e Effetto disegno nel gruppo Parametri modello.
- ▶ Fare clic su Continua.
- ▶ Fare clic su Medie stimate nella finestra di dialogo Modello lineare generalizzato.

Figura 19-5
Finestra di dialogo Medie stimate modello lineare generalizzato



- ▶ Scegliere di visualizzare le medie per *destinatario*, *buoni* e per l'interazione *destinatario*buoni*.

- ▶ Selezionare un contrasto Semplice e 3 Se stesso e famiglia come categoria di riferimento per *destinatario*. Si noti che, una volta selezionata, la categoria viene visualizzata come “3” nella finestra di dialogo.
- ▶ Selezionare un contrasto Semplice e 1 No come categoria di riferimento per *buoni*.
- ▶ Fare clic su Continua.
- ▶ Fare clic su OK nella finestra di dialogo Modello lineare generalizzato.

Riepilogo del modello (Regressione output)

Figura 19-6
Statistica R-quadrato

R-quadrato	,601
------------	------

a. Modello: Importo speso = (Intercetta) + perconto + usabuoni + perconto * usabuoni

R-quadrato, il coefficiente di determinazione, è una misura della bontà di adattamento del modello lineare. Mostra che circa il 60% della variazione in *Importo di spesa* è spiegato dal modello, il che indica una buona capacità esplicativa. Potrebbe essere utile aggiungere altri predittori al modello per migliorare ulteriormente l'adattamento.

Test degli effetti del modello

Figura 19-7
Test degli effetti tra soggetti

Sorgente	df1	df2	F di Wald	Correzione per confronti multipli
(Modello corretto)	11.000	3.000	127.231	.001
(Intercetta)	1.000	13.000	6321.597	.000
perconto	2.000	12.000	643.593	.000
usabuoni	3.000	11.000	87.453	.000
perconto * usabuoni	6.000	8.000	10.688	.002

a. Modello: Importo speso = (Intercetta) + perconto + usabuoni + perconto * usabuoni

Ogni termine del modello, più il modello considerato globalmente, viene testato per verificare che il valore del suo effetto sia uguale a 0. I termini con valore di significatività inferiori a 0,05 hanno un effetto rilevabile. Di conseguenza, tutti i termini del modello contribuiscono al modello.

Stime di parametri

Figura 19-8
Stime dei parametri

Parametro	Stima	Errore standard	Intervallo di confidenza 95%		Effetto disegno
			Inferiore	Superiore	
(Intercetta)	518.249	11.731	492.905	543.592	1.387
[perconto=1]	-174.757	10.762	-198.0	-151.51	.950
[perconto=2]	-129.443	11.455	-154.2	-104.70	.925
[perconto=3]	.000 ^a
[usabuoni=1]	-140.838	10.180	-162.8	-118.85	.649
[usabuoni=2]	-63.026	13.195	-91.531	-34.520	.940
[usabuoni=3]	-31.375	9.726	-52.387	-10.363	.564
[usabuoni=4]	.000 ^a
[perconto=1] * [usabuoni=1]	41.693	11.170	17.562	65.824	.606
[perconto=1] * [usabuoni=2]	44.505	18.068	5.471	83.539	1.413
[perconto=1] * [usabuoni=3]	9.204	11.057	-14.684	33.092	.594
[perconto=1] * [usabuoni=4]	.000 ^a
[perconto=2] * [usabuoni=1]	89.211	10.967	65.518	112.903	.533
[perconto=2] * [usabuoni=2]	54.267	14.949	21.972	86.562	.836
[perconto=2] * [usabuoni=3]	17.884	13.753	-11.828	47.595	.797
[perconto=2] * [usabuoni=4]	.000 ^a
[perconto=3] * [usabuoni=1]	.000 ^a
[perconto=3] * [usabuoni=2]	.000 ^a
[perconto=3] * [usabuoni=3]	.000 ^a
[perconto=3] * [usabuoni=4]	.000 ^a

a. Impostare su zero perché questo parametro è dupli

b. Modello: Importo speso = (Intercetta) + perconto + usabuoni + perconto * usabuoni

Le stime dei parametri mostrano l'effetto di ogni predittore su *Importo di spesa*. Il valore di 518,249 per il termine di intercetta indica che la catena può aspettarsi che un cliente con famiglia che utilizza buoni sconto ricevuti per posta o trovati sui giornali spenda in media \$518,25. È possibile affermare che l'intercetta è associata a questi livelli di fattore perché i relativi parametri sono ridondanti.

- I coefficienti di *destinatario* suggeriscono che, tra i clienti che utilizzano buoni inviati per posta e trovati sui giornali, quelli senza famiglia tendono a spendere meno di quelli sposati, che a loro volta spendono meno di quelli con persone a carico. Poiché i test degli effetti del modello hanno mostrato che questo termine contribuisce al modello, tali differenze non sono dovute al caso.
- I coefficienti di *buoni* suggeriscono che la spesa tra i clienti con persone a carico si riduce in corrispondenza di un minore uso dei buoni sconto. Esiste una ridotta quantità di incertezza nelle stime, ma l'intervallo di confidenza non include 0.
- I coefficienti di interazione suggeriscono che i clienti che non utilizzano i buoni sconto oppure li ritagliano solo dai giornali e non hanno persone a carico tendono a spendere più di quanto ci si aspetterebbe. Il parametro di interazione è ridondante se una sua parte lo è.
- La deviazione nei valori degli effetti del disegno da 1 indica che alcuni degli errori standard calcolati per queste stime dei parametri sono maggiori di quelli che si sarebbero ottenuti presumendo che queste osservazioni derivassero da un semplice campione casuale, mentre gli altri sono minori. È fondamentale integrare le informazioni sul disegno di campionamento nell'analisi, in quanto diversamente sarebbe possibile, ad esempio, dedurre che il coefficiente *buoni=3* non sia diverso da 0.

Le stime dei parametri sono utili per quantificare l'effetto di ogni termine del modello, ma le tabelle delle medie marginali stimate possono semplificare l'interpretazione dei risultati del modello.

Medie marginali stimate

Figura 19-9
Medie marginali stimate per livelli di destinatario spesa

Compra per conto di	Media	Errore standard	Intervallo di confidenza 95%	
			Inferiore	Superiore
Sé stesso	308.5326	3.94286	300.0145	317.0506
Sé stesso e coniuge	370.3361	4.87908	359.7955	380.8767
Sé stesso e famiglia	459.4392	7.19769	443.8895	474.9888

La tabella mostra le medie marginali stimate per modello e gli errori standard di *Importo di spesa* ai livelli di fattore di *Destinatari spesa*. Questa tabella risulta utile per esaminare le differenze tra i livelli di questo fattore. Nell'esempio, si prevede che un cliente che fa spesa per se stesso spenda circa \$308,53, mentre per un cliente sposato la previsione è di \$370,34 e per un cliente con persone a carico di \$459,44. Per verificare se questa differenza è reale o dovuta a una variazione casuale, esaminare i risultati del test.

Figura 19-10
Risultati del singolo test per medie marginali stimate di genere

Contrasto semplice Compra per conto di ^a	Stima del contrasto	Valore ipotizzato	Differenza (stima - ipotesi)	Errore standard	df1	df2	F di Wald	Correzione per confronti multipli
Livello Sé stesso rispetto a livello Sé stesso e famiglia	-150.907	.000	-150.907	4.903	1.000	13.00	947.41	.000
Livello Sé stesso e coniuge rispetto a livello Sé stesso e famiglia	-89.103	.000	-89.103	5.903	1.000	13.00	227.84	.000

a. Categoria di riferimento = Sé stesso e famiglia

La tabella dei singoli test mostra due contrasti semplici nella spesa.

- La stima del contrasto è la differenza nella spesa per i livelli in elenco di *Destinatari spesa*.
- Il valore ipotizzato di 0,00 rappresenta la convinzione che non vi sia differenza di spesa.
- La statistica di Wald F , con i gradi visualizzati di libertà, viene utilizzata per verificare se la differenza tra una stima di contrasto e un valore ipotizzato sia dovuta a una variazione casuale.
- Poiché il valore di significatività è minore di 0,05 è possibile concludere che sono presenti differenze di spesa.

I valori delle stime di contrasto sono diversi dalle stime dei parametri. Ciò dipende dal fatto che è presente un termine di interazione contenente l'effetto *Destinatari spesa*. Di conseguenza, la stima del parametro per *destinatario=1* è un contrasto semplice tra i livelli *Se stesso* e *Se stesso e famiglia* al livello *Da entrambi* della variabile *Utilizzo buoni sconto*. La stima del contrasto nella tabella è mediata sui livelli di *Utilizzo buoni sconto*.

Figura 19-11

Risultati dei test globali per medie marginali stimate di genere

df1	df2	F di Wald	Correzione per confronti multipli
2.000	12.000	643.593	.000

La tabella dei test globali include i risultati di un test su tutti i contrasti nella tabella dei test singoli. Il valore di significatività inferiore a 0,05 conferma che esiste una differenza nella spesa tra i livelli di *Destinatari spesa*.

Figura 19-12

Medie marginali stimate per livelli di stile di spesa

	Media	Errore standard	Intervallo di confidenza 95%	
			Inferiore	Superiore
Usa buoni sconto				
No	319.6455	6.51429	305.5722	333.7188
Da riviste	386.7469	4.32295	377.4077	396.0861
Da depliant promozionali	394.5028	5.54218	382.5297	406.4760
Da entrambi	416.8486	6.51260	402.7790	430.9162

La tabella mostra le medie marginali stimate per modello e gli errori standard di *Importo di spesa* ai livelli di fattore di *Utilizzo buoni sconto*. Questa tabella risulta utile per esaminare le differenze tra i livelli di questo fattore. In questo esempio, si prevede che un cliente che non utilizza buoni sconto spenda circa \$319,65 e che chi utilizza i buoni spenda notevolmente di più.

Figura 19-13

Risultati del singolo test per medie marginali stimate di stile di spesa

Contrasto semplice Usa buoni sconto ^a	Stima del contrasto	Valore ipotizzato	Differenza (stima - ipotesi)	Errore standard	df1	df2	F di Wald	Correzione per confronti multipli
Livello Da riviste rispetto a livello No	67.101	.000	67.101	6.537	1.000	13.000	105.35	.000
Livello Da depliant promozionali rispetto a livello No	74.857	.000	74.857	5.875	1.000	13.000	162.33	.000
Livello Da entrambi rispetto a livello No	97.203	.000	97.203	5.603	1.000	13.000	300.92	.000

a. Categoria di riferimento = No

La tabella dei test singoli mostra tre contrasti semplici, che paragonano la spesa di clienti che non utilizzano i buoni sconto con quella dei clienti che ne fanno uso.

Poiché i valori di significatività dei test sono inferiori a 0,05, è possibile concludere che i clienti che utilizzano i buoni sconto tendono a spendere di più di quelli che non ne fanno uso.

Figura 19-14

Risultati dei test globali per medie marginali stimate di stile di spesa

df1	df2	F di Wald	Correzione per confronti multipli
3.000	11.000	87.453	.000

La tabella dei test globali mostra i risultati di un test su tutti i contrasti nella tabella dei test singoli. Il valore di significatività inferiore a 0,05 conferma che esiste una differenza nella spesa tra i livelli di *Utilizzo buoni sconto*. Si noti che i test globali per *Utilizzo buoni sconto* e *Destinatario*

spesa sono equivalenti ai test degli effetti del modello in quanto i valori del contrasto ipotizzato sono pari a 0.

Figura 19-15

Medie marginali stimate per livelli di genere per stile di spesa

Compra per conto di	Usa buoni sconto	Media	Errore standard	Intervallo di confidenza 95%	
				Inferiore	Superiore
Sé stesso	No	244.3471	6.00949	231.3644	257.3298
	Da riviste	324.9708	5.94134	312.1353	337.8063
	Da depliant promozionali	321.3207	4.11028	312.4410	330.2005
	Da entrambi	343.4916	6.57845	329.2797	357.7034
Sé stesso e coniuge	No	337.1783	7.12181	321.7925	352.5640
	Da riviste	380.0468	7.91038	362.9574	397.1361
	Da depliant promozionali	375.3141	6.22468	361.8665	388.7617
	Da entrambi	388.8054	7.12101	373.4214	404.1894
Sé stesso e famiglia	No	377.4111	11.58215	352.3894	402.4328
	Da riviste	455.2232	6.14420	441.9494	468.4969
	Da depliant promozionali	486.8736	10.76529	463.6166	510.1306
	Da entrambi	518.2488	11.73120	492.9050	543.5925

La tabella mostra le medie marginali stimate per modello, gli errori standard e gli intervalli di confidenza di *Importo di spesa* alle combinazioni di fattori di *Destinatari spesa* e *Utilizzo buoni sconto*. Questa tabella risulta utile per esaminare l'effetto di interazione tra questi due fattori, rilevato nei test degli effetti del modello.

Riepilogo

Nell'esempio, le medie marginali stimate hanno rivelato differenze nella spesa tra i clienti per livelli diversi di *Destinatari spesa* e *Utilizzo buoni sconto*. I test degli effetti del modello hanno confermato questa conclusione e il fatto che sembra essere presente un effetto di interazione *Destinatari spesa*Utilizzo buoni sconto*. La tabella di riepilogo del modello ha rivelato che il modello corrente spiega più della metà della variazione dei dati e può essere probabilmente migliorato aggiungendo ulteriori predittori.

Procedure correlate

La procedura Modello lineare generalizzato campioni complessi è un utile strumento per la creazione di modelli relativi a una variabile di scala per casi estratti in base a uno schema di campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.

- **Preparazione guidata dell'analisi Campioni complessi** viene utilizzata per definire le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura **Regressione logistica campioni complessi** consente di creare il modello di una risposta categoriale.
- La procedura **Regressione ordinale campioni complessi** consente di creare il modello di una risposta ordinale.

Regressione logistica campioni complessi

La procedura Regressione logistica campioni complessi esegue l'analisi di regressione logistica su una variabile dipendente binaria o multinomiale per campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Utilizzo della Regressione logistica campioni complessi per la valutazione del rischio di credito.

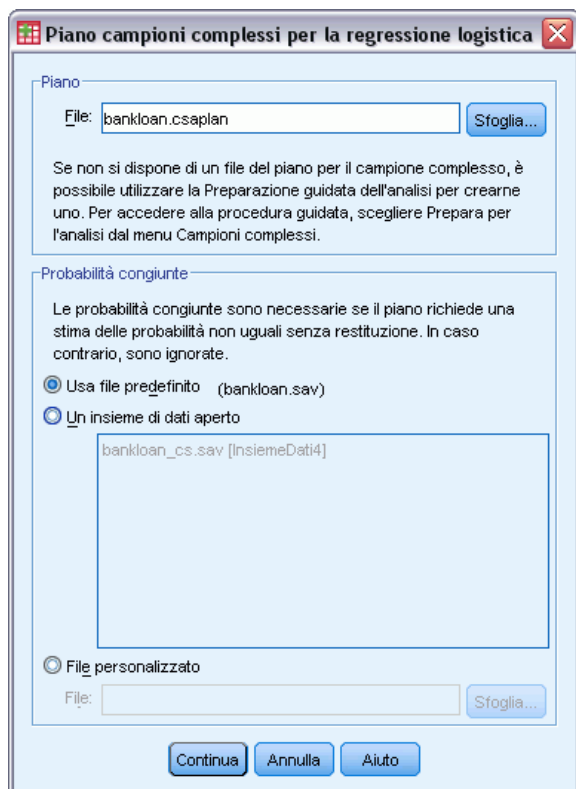
Un funzionario mutui presso una banca sarà interessato a identificare le caratteristiche indicative delle persone che tendenzialmente saranno inadempienti per quanto riguarda il rimborso dei prestiti e a utilizzare tali caratteristiche per identificare i rischi di credito positivi e negativi.

Si supponga che un funzionario mutui abbia raccolto i dati precedenti relativi a clienti cui siano stati accordati prestiti da parte di varie filiali, in base a un disegno complesso. Tali informazioni sono contenute nel file *bankloan_cs.sav*. [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Si desidera verificare se la probabilità di mancato rimborso da parte di un cliente è connessa all'età, alla storia lavorativa e all'importo dell'esposizione debitoria, inserendo il disegno di campionamento.

Esecuzione dell'analisi

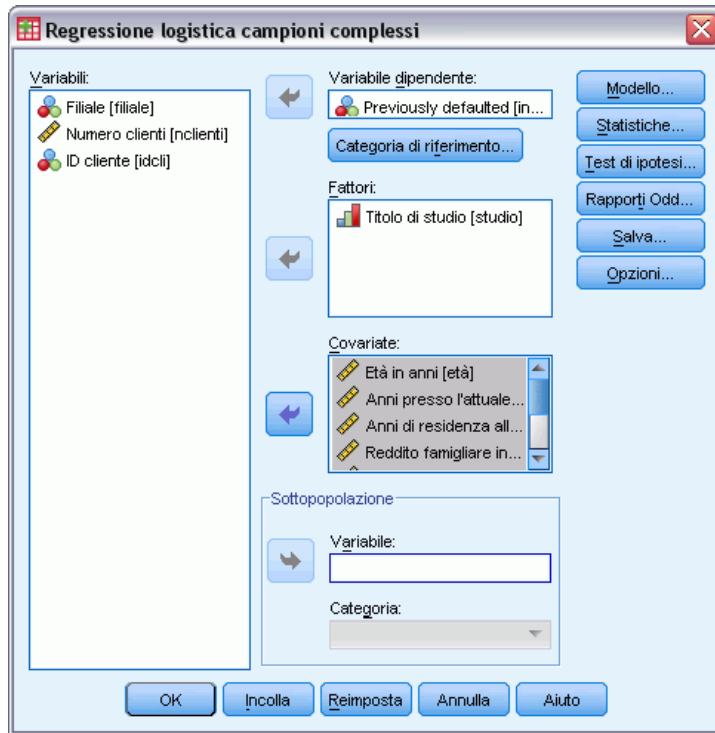
- ▶ Per creare il modello di regressione logistica, dai menu scegliere:
Analizza > Campioni complessi > Regressione logistica...

Figura 20-1
Finestra di dialogo Piano campioni complessi



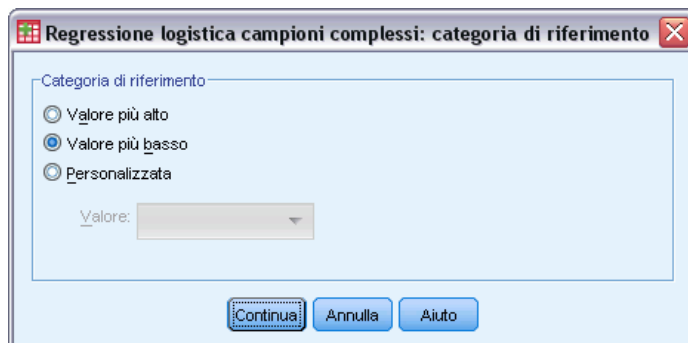
- ▶ Accedere alla finestra e selezionare il file *bankloan.csaplan*. Per ulteriori informazioni, vedere l'argomento *File di esempio* in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Fare clic su Continua.

Figura 20-2
Finestra di dialogo *Regressione logistica*



- ▶ Selezionare *Già inadempiente* come variabile dipendente.
- ▶ Selezionare *Livello di istruzione* come fattore.
- ▶ Selezionare da *Età in anni* a *Altri indebitamenti in migliaia* come covariate.
- ▶ Selezionare *Già inadempiente* e fare clic su *Categoria di riferimento*.

Figura 20-3
Finestra di dialogo Regressione logistica: Categoria riferimento

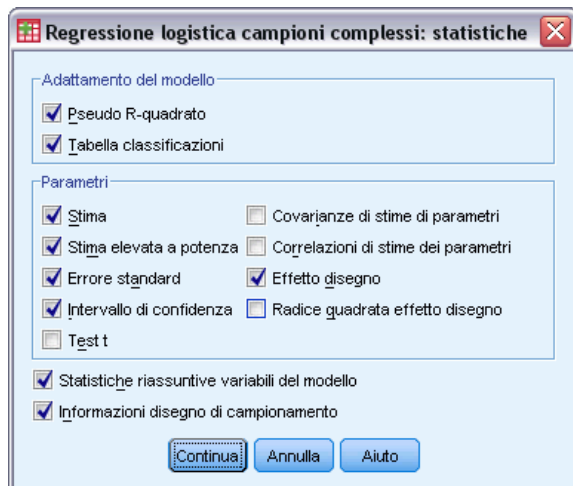


- Selezionare Valore minimo come categoria di riferimento.

Questo imposta la categoria “Nessun inadempimento precedente” come categoria di riferimento; di conseguenza, per i rapporti odd presenti nell’output, l’aumento dei rapporti odd corrisponderà all’aumento della probabilità di inadempimento.

- Fare clic su Continua.
- Nella finestra di dialogo Regressione logistica fare clic su Statistiche.

Figura 20-4
Finestra di dialogo Regressione logistica: Statistiche



- Selezionare Tabella di classificazione nel gruppo Adattamento del modello.
- Selezionare Stima, Stima elevata a potenza, Errore standard, Intervallo di confidenza e Effetto disegno nel gruppo Parametri.
- Fare clic su Continua.
- Nella finestra di dialogo Regressione logistica fare clic su Rapporti odd.

Figura 20-5
Finestra di dialogo *Regressione logistica: Rapporti odd*



- ▶ Scegliere di creare i rapporti odd per il fattore *istr* e le covariate *impiego* e *debred*.
- ▶ Fare clic su Continua.
- ▶ Nella finestra di dialogo *Regressione logistica* scegliere OK.

Pseudo R-quadrati

Figura 20-6
Statistiche pseudo R-quadrato

Cox e Snell	,330
Nagelkerke	,451
McFadden	,304

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello
: (Intercept), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

Nel modello di regressione lineare, il coefficiente di determinazione, l' R^2 , riepiloga la proporzione della varianza nella variabile dipendente associata alle variabili predittore (indipendente), con i valori più grandi dell' R^2 che indicano che il più della variazione viene spiegato dal modello, fino a un massimo di 1. Per i modelli di regressione con una variabile dipendente categoriale, non è possibile calcolare una singola statistica dell' R^2 che ha tutte le caratteristiche dell' R^2 nel modello di regressione lineare, pertanto, queste approssimazioni vengono calcolate al loro posto. I metodi seguenti consentono di stimare il coefficiente di determinazione.

- L' R^2 (Cox e Snell, 1989) di Cox e Snell è basato sul confronto fra la log likelihood del modello e la log likelihood di un modello di riferimento. Con risultati categoriali, tuttavia, presenta un valore massimo teorico inferiore a 1, anche per un modello “perfetto”.
- L' R^2 (Nagelkerke, 1991) di Nagelkerke è una versione corretta dell' R^2 quadro di Cox e Snell che rettifica la scala della statistica allo scopo di coprire la gamma completa da 0 a 1.
- L' R^2 (McFadden, 1974) di McFadden è un'altra versione, basata sui kernel di verosimiglianza del modello di solo intercetta e del modello a stima completa.

Ciò che costituisce un “buon” valore R^2 dipende dalle varie aree di applicazione. Se queste statistiche sono interessanti di per sé, risultano utili soprattutto durante il confronto di modelli concorrenti per gli stessi dati. Il modello con la statistica R^2 maggiore è il “migliore” in base a questa misura.

Classificazione

Figura 20-7
Tabella classificazioni

Osservati	Previsti		Percentuale correzione
	No	Sì	
No	188289,667	31871,267	85,5%
Sì	49970,600	77675,133	60,9%
Percentuale complessiva	68,5%	31,5%	76,5%

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello : (intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

La tabella di classificazione mostra i risultati pratici dell'utilizzo del modello di regressione logistica. Per ciascun caso, la risposta attesa è *Sì* se il logit previsto in base al modello del caso è maggiore di 0. I casi sono ponderati in base a *pesofinale*, di conseguenza la tabella di classificazione riporta i risultati del modello atteso nella popolazione.

- Le celle sulla diagonale rappresentano previsioni corrette.
- Le celle fuori dalla diagonale rappresentano previsioni non corrette.

In base ai casi utilizzati per creare il modello, è prevedibile una corretta classificazione dell'85,5% dei soggetti non inadempienti nella popolazione tramite il modello corrente. Analogamente, è prevedibile una corretta classificazione del 60,9% dei soggetti inadempienti. Nel complesso, è prevedibile una corretta classificazione del 76,5% dei casi; tuttavia, poiché la tabella è stata creata con i casi utilizzati per la generazione del modello, queste stime potrebbero essere eccessivamente ottimistiche.

Test degli effetti del modello

Figura 20-8
Test degli effetti tra soggetti

Sorgente	df1	df2	F di Wald	Correzione per confronti multipli
(Modello corretto)	11,000	4,000	14,669	,010
(Intercetta)	1,000	14,000	5,777	,031
studio	4,000	11,000	1,683	,224
età	1,000	14,000	5,352	,036
impiego	1,000	14,000	88,244	,000
residenz	1,000	14,000	1,123	,307
debredd	1,000	14,000	27,632	,000
debccred	1,000	14,000	33,402	,000
altdeb	1,000	14,000	,709	,414
reddito	1,000	14,000	,007	,932

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello : (Intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

Ogni termine del modello, più il modello considerato globalmente, viene testato per verificare che il valore del suo effetto sia uguale a 0. I termini con valore di significatività inferiori a 0,05 hanno un effetto rilevabile. Di conseguenza *età*, *impiego*, *debredd* e *creddeb* contribuiscono al modello, mentre gli altri effetti principali no. In un'ulteriore analisi dei dati si potrebbe forse eliminare *istr*, *indirizzo*, *reddito* e *altrideb* dalla considerazione del modello.

Stime di parametri

Figura 20-9
Stime dei parametri

Previamente default	Parametro	B	Errore standard	Intervallo di confidenza 95%		Effetto disegno	Exp(B)	Intervallo di confidenza 95% per Exp(B)	
				Inferiore	Superiore			Inferiore	Superiore
Si	(Intercetta)	-1,140	,399	-1,995	-,284	,665	,320	,136	,753
	[studio=1]	,720	,340	-,010	1,449	,862	2,054	,990	4,259
	[studio=2]	,684	,371	-,112	1,481	1,247	1,983	,894	4,397
	[studio=3]	,518	,307	-,140	1,177	,813	1,679	,869	3,244
	[studio=4]	,789	,302	,142	1,437	,817	2,202	1,152	4,208
	[studio=5]	,000 ^a					1,000		
	età	-,023	,010	-,043	-,002	,418	,978	,958	,998
	impiego	-,225	,024	-,277	-,174	1,200	,798	,758	,840
	residenz	-,028	,026	-,085	,029	,651	,972	,919	1,029
	debredd	,095	,018	,056	,134	1,222	1,100	1,058	1,143
	debccred	,493	,085	,310	,676	1,373	1,637	1,363	1,966
	altdeb	,026	,031	-,041	,094	1,219	1,027	,960	1,098
	reddito	,000	,003	-,007	,006	1,410	1,000	,993	1,006

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello : (Intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

a. Impostare su zero perché questo parametro è duplicato.

La tabella delle stime dei parametri riassume l'effetto di ogni predittore. Si noti che i valori dei parametri influenzano la probabilità della categoria "già inadempiente" rispetto alla categoria "nessun inadempimento precedente". Di conseguenza, i parametri con coefficienti positivi

aumentano la probabilità di inadempimento, mentre i parametri con coefficienti negativi riducono la probabilità di inadempimento.

Il significato di un coefficiente di regressione logistica non è così immediato come quello di un coefficiente di regressione lineare. Mentre B è adattato per verificare l'effetto dei modelli, $Exp(B)$ è più facile da interpretare. $Exp(B)$ rappresenta la modifica del rapporto negli odd dell'evento di interesse per una variazione di un'unità nel predittore per i predittori che non siano parte dei termini di interazione. Ad esempio, $Exp(B)$ per *impiego* è pari a 0,798, il che significa che gli odd dell'inadempimento per soggetti che hanno il proprio posto di lavoro da due anni sono 0,798 volte gli odd di inadempimento per coloro il cui rapporto di lavoro dura da un anno, a parità delle altre condizioni.

Gli effetti del disegno indicano che alcuni degli errori standard calcolati per queste stime dei parametri sono maggiori di quelli che si sarebbero ottenuti presumendo che queste osservazioni derivassero da un semplice campione casuale, mentre gli altri sono minori. È fondamentale integrare le informazioni sul disegno di campionamento nell'analisi, in quanto diversamente sarebbe possibile, ad esempio, dedurre che il coefficiente legato all'età non sia diverso da 0.

Rapporti Odd

Figura 20-10
Rapporti odd per livello di istruzione

		Previ ously def...	Rapporto Odd	Intervallo di confidenza 95%	
				Inferiore	Superiore
Titolo di studio	Non ha completato le superiori vs. Titolo post-laurea	Si	2,054	,990	4,259
	Licenza di scuola superiore vs. Titolo post-laurea	Si	1,983	,894	4,397
	Alcuni esami all'università vs. Titolo post-laurea	Si	1,679	,869	3,244
	Laurea vs. Titolo post-laurea	Si	2,202	1,152	4,208

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello
: (Intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

a. Fattori e covariate utilizzati nel calcolo sono stati fissati ai seguenti valori: Titolo di studio=Titolo post-laurea; Età in anni=34,19; Anni presso l'attuale datore di lavoro=6,99; Anni di residenza all'indirizzo attuale=6,32; Rapporto debiti / reddito (x100)=9,9341; Debito su carta di credito in migliaia=1,9764; Altri debiti in migliaia=3,9164; Reddito familiare in migliaia=60,1581

La tabella visualizza i rapporti odd di *Già inadempiente* ai livelli dei fattori di *Livello di istruzione*. I valori riportati sono i rapporti degli odd di inadempimento per le categorie da *Assenza di diploma di scuola superiore* a *Laurea*, a confronto con gli odd di inadempimento per *Master*. Di conseguenza, il rapporto odd pari a 2,054 nella prima riga della tabella indica che gli odd di inadempimento per un soggetto che non ha completato la scuola superiore sono 2,054 volte gli odd di inadempimento per un soggetto in possesso di master.

Figura 20-11
Rapporti odd per anni durata impiego attuale

Unità di variazione	Previously defaulted	Rapporto Odd	Intervallo di confidenza 95%	
			Inferiore	Superiore
Anni presso l'attuale datore di lavoro	Sì	,798	,758	,840

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello : (Intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

a. Fattori e covariate utilizzati nel calcolo sono stati fissati ai seguenti valori: Titolo di studio= Titolo post-laurea; Età in anni=34,19; Anni presso l'attuale datore di lavoro=6,99; 1; Anni di residenza all'indirizzo attuale=6,32; Rapporto debiti / reddito (x100)=9,9341; Debito su carta di credito in migliaia=1,9764; Altri debiti in migliaia=3,9164; Reddito familiare in migliaia=60,1581

La tabella visualizza i rapporti odd di *Già inadempiente* per una variazione unitaria nella covariata *Anni durata impiego attuale*. Il valore riportato è il rapporto degli odd di inadempimento per un soggetto il cui impiego corrente dura da 7,99 anni, rispetto agli odd di inadempimento per un soggetto il cui impiego dura da 6,99 (la media).

Figura 20-12
Rapporti odd per rapporto indebitamento/reddito

Unità di variazione	Previously defaulted	Rapporto Odd	Intervallo di confidenza 95%	
			Inferiore	Superiore
Rapporto debiti / reddito (x100)	Sì	1,100	1,058	1,143

Variabile dipendente: Previously defaulted (categoria di riferimento = No) Modello : (Intercetta), studio, età, impiego, residenz, debredd, debccred, altdeb, reddito

a. Fattori e covariate utilizzati nel calcolo sono stati fissati ai seguenti valori: Titolo di studio= Titolo post-laurea; Età in anni=34,19; Anni presso l'attuale datore di lavoro=6,99; 1; Anni di residenza all'indirizzo attuale=6,32; Rapporto debiti / reddito (x100)=9,9341; Debito su carta di credito in migliaia=1,9764; Altri debiti in migliaia=3,9164; Reddito familiare in migliaia=60,1581

La tabella visualizza i rapporti odd di *Già inadempiente* per una variazione unitaria nella covariata *Rapporto debito/reddito*. Il valore riportato è il rapporto degli odd di inadempimento per un soggetto con rapporto tra indebitamento e reddito pari a 10,9341, rispetto agli odd di inadempimento per un soggetto il cui rapporto è 9,9341 (la media).

Si noti che, poiché nessuno di questi predittori è parte dei termini di interazione, i valori dei rapporti odd riportati nelle tabelle sono pari ai valori delle stime dei parametri elevate a potenza. Quando un predittore è parte di un termine di interazione, il suo rapporto odd come riportato nelle tabelle dipenderà anche dai valori degli altri predittori che compongono l'interazione.

Riepilogo

Utilizzando la procedura di Regressione logistica campioni complessi è stato creato un modello per la previsione della probabilità che un determinato cliente sia inadempiente rispetto a un mutuo.

Un elemento fondamentale per i funzionari mutui è il costo degli errori di Tipo I e di Tipo II. In altre parole, qual è il costo della classificazione di un soggetto inadempiente come adempiente (Tipo I)? Qual è il costo della classificazione di un soggetto adempiente come inadempiente (Tipo II)? Se l'interesse principale è l'inadempienza, è utile ridurre l'errore di Tipo I e massimizzare la **sensibilità**. Se la priorità è ampliare la base clienti, è utile ridurre l'errore di Tipo II e massimizzare la **specificità**. Normalmente, entrambi questi elementi hanno importanza, quindi diventa necessario scegliere una regola decisionale per la classificazione della clientela che offra una combinazione ottimale di sensibilità e specificità.

Procedure correlate

La procedura Regressione logistica campioni complessi è un utile strumento per la creazione di modelli relativi a una variabile categoriale per casi estratti in base a uno schema di campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.
- **Preparazione guidata dell'analisi Campioni complessi** viene utilizzata per definire le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura **Modello lineare generalizzato campioni complessi** consente di creare il modello di una risposta scalare.
- La procedura **Regressione ordinale campioni complessi** consente di creare il modello di una risposta ordinale.

Regressione ordinale campioni complessi

La procedura Regressione ordinale campioni complessi consente di creare un modello predittivo per una variabile dipendente ordinale per campioni estratti da metodi di campionamento complessi. In alternativa, è possibile richiedere le analisi per una sottopopolazione.

Utilizzo di Regressione ordinale campioni complessi per l'analisi dei risultati di indagini

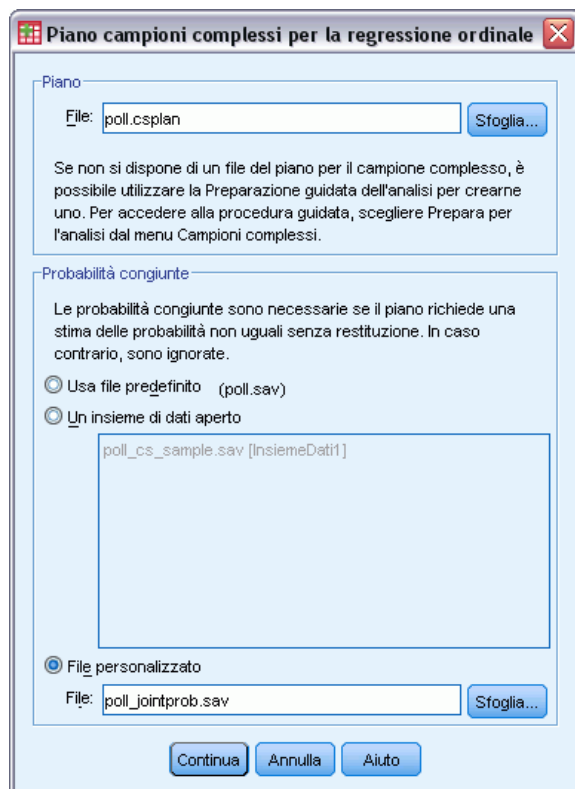
I deputati che discutono una legge in parlamento hanno l'interesse di sapere se la legge gode del sostegno pubblico e il modo in cui tale sostegno è correlato ai dati demografici degli elettori. Gli esperti in sondaggi di opinione sviluppano ed eseguono interviste in base a un disegno di campionamento complesso.

I risultati del sondaggio vengono raccolti nel file *poll_cs_sample.sav*. Il piano di campionamento utilizzato è contenuto in *poll.csplan*. Poiché fa uso del metodo PPS (probability-proportional-to-size, probabilità proporzionale alla dimensione), esiste anche un file contenente le probabilità di selezione congiunte (*poll_jointprob.sav*). [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). La procedura Regressione ordinale campioni complessi viene utilizzata per elaborare un modello per il livello di sostegno alla legge in base a dati demografici degli elettori.

Esecuzione dell'analisi

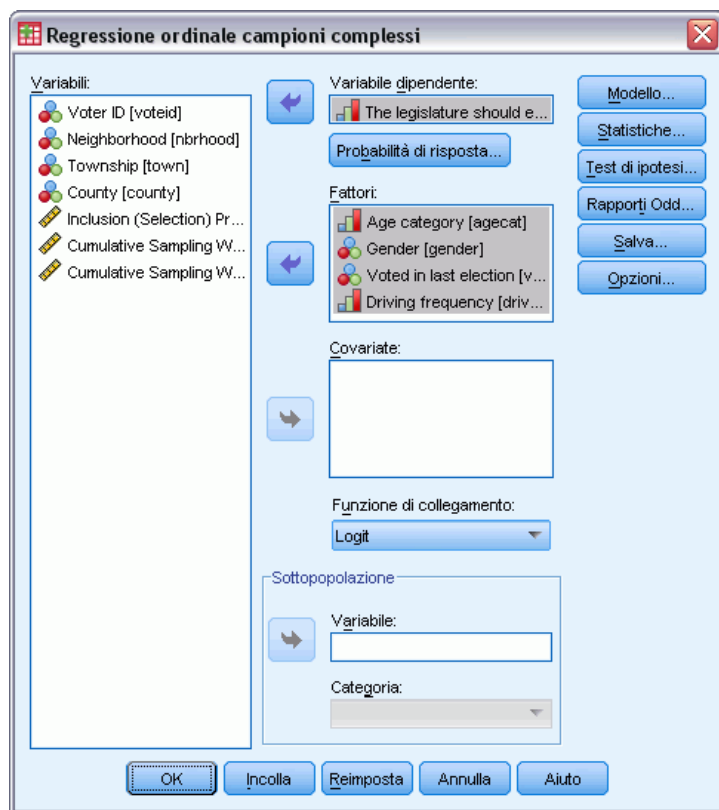
- ▶ Per eseguire un'analisi Regressione ordinale campioni complessi, dai menu scegliere:
Analizza > Complex Samples > Regressione ordinale...

Figura 21-1
Finestra di dialogo Piano campioni complessi



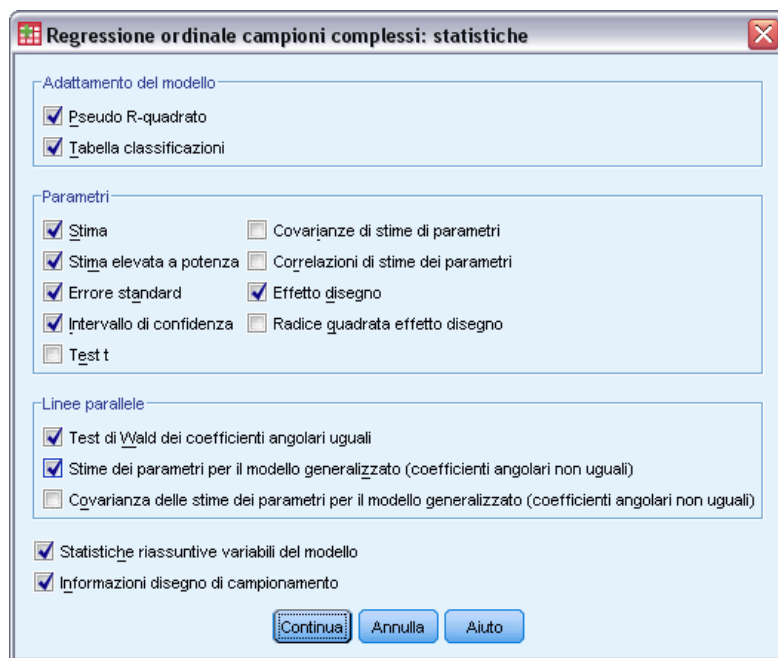
- ▶ Accedere alla finestra e selezionare il file *poll.csplan* come file del piano. Per ulteriori informazioni, vedere l'argomento [File di esempio](#) in l'appendice A in *IBM SPSS Complex Samples 20*.
- ▶ Selezionare il file delle probabilità congiunte *poll_jointprob.sav*.
- ▶ Fare clic su Continua.

Figura 21-2
Finestra di dialogo *Regressione ordinale*



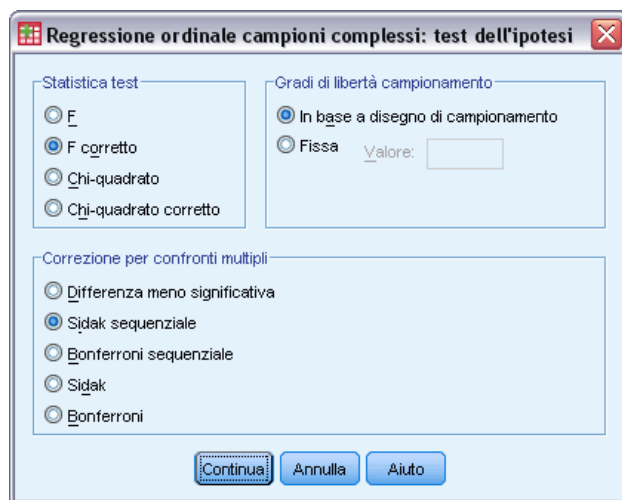
- ▶ Selezionare la variabile dipendente *\$\$\$The legislature should enact a gas tax.*
- ▶ Selezionare i fattori da *Categoria età* a *\$\$\$Driving frequency.*
- ▶ Fare clic su *Statistiche.*

Figura 21-3
Finestra di dialogo Regressione ordinale: Statistiche



- ▶ Selezionare Tabella di classificazione nel gruppo Adattamento del modello.
- ▶ Selezionare Stima, Stima elevata a potenza, Errore standard, Intervallo di confidenza e Effetto disegno nel gruppo Parametri.
- ▶ Selezionare Test di Wald dei coefficienti angolari uguali e Stime dei parametri per il modello generalizzato (coefficienti angolari non uguali).
- ▶ Fare clic su Continua.
- ▶ Fare clic su Test di ipotesi nella finestra di dialogo Regressione ordinale campioni complessi.

Figura 21-4
Finestra di dialogo *Test di ipotesi*



Persino per un numero moderato di predittori e categorie di risposta il test Wald F può essere inestimabile per il test delle linee parallele.

- ▶ Selezionare F corretto nel gruppo Test.
- ▶ Selezionare Sidak sequenziale come metodo di correzione per confronti multipli.
- ▶ Fare clic su Continua.
- ▶ Fare clic su Rapporti Odd nella finestra di dialogo Regressione ordinale campioni complessi.

Figura 21-5
Finestra di dialogo *Regressione lordinale: Rapporti odd*



- ▶ Scegliere di produrre rapporti odd cumulativi per *Categoria età* e *Driving frequency*.
- ▶ Selezionare *10-14.999 miles/year*, un chilometraggio annuale più comune rispetto al massimo, come categoria di riferimento per *Driving frequency*.
- ▶ Fare clic su Continua.

- Fare clic su OK nella finestra di dialogo Regressione ordinale campioni complessi.

Pseudo R-quadrati

Figura 21-6

Pseudo R-quadrati

Cox e Snell	,179
Nagelkerke	,191
McFadden	,071

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

Nel modello di regressione lineare, il coefficiente di determinazione, l' R^2 , riassume la proporzione della varianza nella variabile dipendente associata alle variabili predittore (indipendente), con i valori più grandi dell' R^2 che indicano che il più della variazione viene spiegato dal modello, fino a un massimo di 1. Per i modelli di regressione con una variabile dipendente categoriale, non è possibile calcolare una singola statistica dell' R^2 che ha tutte le caratteristiche dell' R^2 nel modello di regressione lineare, pertanto, queste approssimazioni vengono calcolate al loro posto. I metodi seguenti consentono di stimare il coefficiente di determinazione.

- L' R^2 (Cox e Snell, 1989) di Cox e Snell è basato sul confronto fra la log likelihood del modello e la log likelihood di un modello di riferimento. Con risultati categoriali, tuttavia, presenta un valore massimo teorico inferiore a 1, anche per un modello “perfetto”.
- L' R^2 (Nagelkerke, 1991) di Nagelkerke è una versione corretta dell' R^2 quadro di Cox e Snell che rettifica la scala della statistica allo scopo di coprire la gamma completa da 0 a 1.
- L' R^2 (McFadden, 1974) di McFadden è un'altra versione, basata sui kernel di verosimiglianza del modello di solo intercetta e del modello a stima completa.

Ciò che costituisce un “buon” valore R^2 dipende dalle varie aree di applicazione. Se queste statistiche sono interessanti di per sé, risultano utili soprattutto durante il confronto di modelli concorrenti per gli stessi dati. Il modello con la statistica R^2 maggiore è il “migliore” in base a questa misura.

Test degli effetti del modello

Figura 21-7
Test degli effetti del modello

Sorgente	df1	df2	F di Wald corretto	Correzione per confronti multipli	Sig. Sidak sequenziale
agecat	2,283	31,966	6,215	,004	,003
gender	1,000	14,000	,046	,834	,834
votelast	1,000	14,000	,076	,787	,787
drivefreq	3,785	52,987	228,015	,000	,000

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

Ogni termine del modello viene testato per verificare che il valore del suo effetto sia uguale a 0. I termini con valore di significatività inferiori a 0,05 hanno un effetto rilevabile. Pertanto *agecat* e *drivefreq* danno un contributo al modello diversamente dagli altri effetti principali. In un'ulteriore analisi dei dati è possibile prendere in considerazione la rimozione di *gender* (sesso) e *votelast* (ultimovoto) dal modello.

Stime di parametri

La tabella delle stime dei parametri riepiloga l'effetto di ogni predittore. Se l'interpretazione dei coefficienti in questo modello è difficile a causa della natura della funzione di collegamento, i segni dei coefficienti per le covariate e i valori relativi dei coefficienti per livelli di fattore rivelano molto bene gli effetti dei predittori nel modello.

- Relativamente alle covariate, i coefficienti positivi (negativi) indicano relazioni positive (inverse) tra predittori e risultato. Un valore crescente di una covariata con un coefficiente positivo corrisponde a una crescente probabilità di collocarsi tra le categorie di risultato cumulativo "superiori".
- Relativamente ai fattori, un livello di fattore con un coefficiente maggiore indica una maggiore probabilità di collocarsi tra le categorie di risultato cumulativo "superiori". Il segno di un coefficiente per un livello di fattore dipende dall'effetto del livello di fattore relativo alla categoria di riferimento.

Figura 21-8
Stime dei parametri

The legislature should enact a gas tax	Parametro	B	Errore standard	Intervallo di confidenza 95%		
				Inferiore	Superiore	
Strongly agree	(Soglia)	-3,681	,221	-4,155	-3,207	
	[agecat=1]	-,320	,096	-,525	-,115	
	[agecat=2]	-,075	,071	-,227	,077	
	[agecat=3]	-,022	,073	-,180	,135	
	[agecat=4]	,000 ^a	.	.	.	
	[gender=0]	-,082	,054	-,197	,033	
	[gender=1]	,000 ^a	.	.	.	
	[votelast=0]	,008	,052	-,104	,120	
	[votelast=1]	,000 ^a	.	.	.	
	[drivefreq=1]	-4,096	,267	-4,669	-3,523	
	[drivefreq=2]	-3,367	,237	-3,876	-2,857	
	[drivefreq=3]	-2,678	,224	-3,158	-2,199	
	[drivefreq=4]	-1,928	,213	-2,384	-1,471	
[drivefreq=5]	-1,015	,252	-1,555	-,476		
[drivefreq=6]	,000 ^a	.	.	.		
Agree	(Soglia)	-1,963	,153	-2,291	-1,635	
	[agecat=1]	-,385	,095	-,587	-,182	
	[agecat=2]	-,130	,069	-,279	,018	
	[agecat=3]	-,139	,101	-,356	,077	
	[agecat=4]	,000 ^a	.	.	.	
	[drivefreq=1]	-3,867	,318	-4,549	-3,185	
	[drivefreq=2]	-3,005	,175	-3,380	-2,630	
	[drivefreq=3]	-2,290	,187	-2,691	-1,888	
	[drivefreq=4]	-1,633	,166	-1,988	-1,278	
	[drivefreq=5]	-,909	,137	-1,204	-,615	
	[drivefreq=6]	,000 ^a	.	.	.	
	Disagree	(Soglia)	-,648	,087	-,836	-,460
		[agecat=1]	-,231	,081	-,405	-,058
[agecat=2]		-,199	,065	-,338	-,059	
[agecat=3]		-,101	,070	-,250	,049	
[agecat=4]		,000 ^a	.	.	.	
[gender=0]		,054	,041	-,034	,142	
[gender=1]		,000 ^a	.	.	.	
[votelast=0]		-,047	,048	-,151	,056	
[votelast=1]		,000 ^a	.	.	.	
[drivefreq=1]		-3,618	,406	-4,490	-2,747	
[drivefreq=2]		-3,134	,208	-3,580	-2,689	
[drivefreq=3]		-2,327	,112	-2,568	-2,087	
[drivefreq=4]		-1,522	,089	-1,713	-1,331	
[drivefreq=5]	-,779	,093	-,980	-,579		
[drivefreq=6]	,000 ^a	.	.	.		

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

a. Impostare su zero perché questo parametro è duplicato.

È possibile elaborare le interpretazioni seguenti in base alle stime dei parametri:

- Gli individui nelle fasce di età inferiori si dichiarano favorevoli alla legge più di coloro nelle fasce di età superiori.

- Gli individui che guidano di meno si dichiarano favorevoli alla legge più di coloro che guidano spesso.
- I coefficienti delle variabili *sex* e *lastvote*, oltre a non essere statisticamente significativi, sembrano bassi rispetto ad altri coefficienti.

Gli effetti del disegno indicano che alcuni degli errori standard calcolati per queste stime dei parametri sono maggiori di quelli che si sarebbero ottenuti utilizzando un semplice campione casuale, mentre gli altri sono minori. È fondamentale integrare le informazioni sul disegno di campionamento nell'analisi, in quanto diversamente sarebbe possibile, ad esempio, dedurre che il coefficiente relativo al terzo livello di *Age Category* (Categoria d'età), [*etacat=3*], sia notevolmente diverso da 0.

Classificazione

Figura 21-9
Informazioni variabili categoriali

Informazioni su variabili categoriali		Conteggio ponderato	Percentuale ponderata
The legislature should enact a gas tax ^a	Strongly agree	25132,955	21,3%
	Agree	32261,425	27,3%
	Disagree	29477,417	24,9%
	Strongly disagree	31314,203	26,5%
Age category	18-30	20509,504	17,4%
	31-45	35380,506	29,9%
	46-60	34865,792	29,5%
	>60	27430,198	23,2%
Gender	Male	61424,547	52,0%
	Female	56761,453	48,0%
Voted in last election	No	70607,216	59,7%
	Yes	47578,784	40,3%
Driving frequency	Do not own car	3437,137	2,9%
	<10,000 miles/year	10816,349	9,2%
	10-14,999 miles/year	32539,364	27,5%
	15-19,999 miles/year	39179,814	33,2%
	20-29,999 miles/year	25617,804	21,7%
	>=30,000 miles/year	6595,532	5,6%
Ampiezza popolazione		118186,000	100,0%

a. I valori delle variabili dipendenti sono ordinati in ordine crescente.

Considerati i dati osservati, il modello “nullo” (ovvero quello privo di predittori) classificherebbe tutti i clienti nel gruppo modale *Agree*. Quindi, il modello nullo sarebbe corretto il 27,3% delle volte.

Figura 21-10
Tabella classificazioni

Osservati	Previsti				
	Strongly agree	Agree	Disagree	Strongly disagree	Percentuale correzione
Strongly agree	7067,567	12130,814	3875,825	2058,750	28,1%
Agree	4271,234	14464,286	7320,767	6205,137	44,8%
Disagree	2024,816	11703,368	7108,487	8640,746	24,1%
Strongly disagree	889,869	8169,109	6946,522	15308,703	48,9%
Percentuale complessiva	12,1%	39,3%	21,4%	27,3%	37,2%

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

La tabella di classificazione mostra i risultati pratici dell'utilizzo del modello. Per ogni caso la risposta prevista è la categoria di risposta con la maggiore probabilità prevista dal modello. I casi vengono ponderati in base a *Peso di campionamento finale*, di conseguenza la tabella di classificazione riporta i risultati del modello previsti nella popolazione.

- Le celle sulla diagonale rappresentano previsioni corrette.
- Le celle fuori dalla diagonale rappresentano previsioni non corrette.

Il modello classifica correttamente un ulteriore 9,9%, ovvero il 37,2% dei casi. In particolare il modello classifica sicuramente meglio coloro il cui giudizio è *Agree* (D'accordo) o *Strongly disagree* (Completamente in disaccordo) e leggermente peggio coloro il cui giudizio è *Disagree* (In disaccordo).

Rapporti Odd

I **rapporti odd cumulativi** vengono definiti come il rapporto tra la probabilità che la variabile dipendente assuma un valore inferiore o uguale a una data categoria di risposta e la probabilità che assuma un valore superiore alla stessa categoria di risposta. Il **rapporto odd cumulativi** è il rapporto degli odd cumulativi per i diversi valori dei predittori ed è strettamente correlato alle stime dei parametri elevate a potenza. È interessante notare come il rapporto odd cumulato stesso non dipenda dalla categoria di risposta.

Figura 21-11
Rapporti odd cumulativi per Categoria età

Rapporti Odd cumulativi 1^a

	Rapporto Odd cumulativo	Intervallo di confidenza 95%		Effetto disegno	Radice quadrata dell'effetto disegno
		Inferiore	Superiore		
Age category 18-30 vs. >60	1,383	1,166	1,639	1,793	1,339
31-45 vs. >60	1,148	1,022	1,290	1,158	1,076
46-60 vs. >60	1,100	,935	1,294	2,206	1,485

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

- a. Fattori e covariate utilizzati nel calcolo sono stati fissati ai seguenti valori: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

La tabella visualizza i rapporti odd cumulativi per i livelli di fattore di *Age category* (Categoria età). I valori riportati sono i rapporti degli odd cumulativi per la fascia di età 18–30 fino a 46–60, rispetto agli odd cumulativi per la fascia >60. Quindi, il rapporto odd 1,383 nella prima riga della tabella significa che gli odd cumulativi di un individuo di età compresa tra 18 e 30 anni corrispondono a 1,383 volte gli odd cumulativi di un individuo con più di 60 anni. Si noti che poiché *Age category* (Categoria età) non rientra in alcun termine di interazione, i rapporti odd rappresentano semplicemente i rapporti delle stime dei parametri elevate a potenza. Ad esempio, il rapporto odd cumulativi tra 18–30 e >60 è pari a $1,00/0,723 = 1,383$

Figura 21-12
Rapporti odd per la frequenza di guida

Rapporti Odd cumulativi 2^a

	Rapporto Odd cumulativo	Intervallo di confidenza 95%		Effetto disegno	Radice quadrata dell'effetto disegno
		Inferiore	Superiore		
Driving frequency Do not own car vs. 10-14,999 miles/year	4,288	2,878	6,390	2,345	1,531
<10,000 miles/year vs. 10-14,999 miles/year	2,030	1,656	2,488	1,838	1,356
15-19,999 miles/year vs. 10-14,999 miles/year	,484	,430	,546	1,450	1,204
20-29,999 miles/year vs. 10-14,999 miles/year	,227	,193	,267	2,095	1,448
>=30,000 miles/year vs. 10-14,999 miles/year	,101	,079	,129	1,585	1,259

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

- a. Fattori e covariate utilizzati nel calcolo sono stati fissati ai seguenti valori: Age category=>60; Gender=Female; Voted in last election=Yes; Driving frequency=>=30,000 miles/year

La tabella visualizza i rapporti odd cumulativi per i livelli di fattore di *Driving frequency* (Frequenza di guida), in cui 14,999 miles/year (10-14.999 miglia/anno) è utilizzato come categoria di riferimento. Poiché *Driving frequency* non rientra in alcun termine di interazione, i rapporti odd rappresentano semplicemente i rapporti delle stime dei parametri elevate a potenza. Ad

esempio il rapporto odd cumulativi tra 20–29.999 miglia/anno e 10–14.999 miles/year (14.999 miglia/anno) è pari a $0,101/0,444 = 0,227$.

Modello cumulativo generalizzato

Figura 21-13
Test di linee parallele

df1	df2	F di Wald corretto	Correzione per confronti multipli	Sig. Sidak sequenziale
8,769	122,767	1,894	,061	,392

Variabile dipendente: The legislature should enact a gas tax (Crescente)

Modello: (Soglia), agecat, gender, votelast, drivefreq

Funzione di collegamento: Logit

Il test delle linee parallele consente di appurare se l'ipotesi che i parametri sono gli stessi per tutte le categorie di risposta è accettabile. Il test opera un confronto tra il modello stimato con un insieme di coefficienti per tutte le categorie e un modello generalizzato con un insieme separato di coefficienti per ogni categoria.

Il test Wald F è un test onnicomprensivo della matrice di contrasto per l'ipotesi delle linee parallele che fornisce valori p asintoticamente corretti; per i campioni medio-piccoli, il test Wald F rettificato produce buoni risultati. Il valore di significatività è prossimo a 0,05, suggerendo che il modello generalizzato potrebbe migliorare l'adattamento del modello. Il test Sidak sequenziale corretto, tuttavia, produce un valore di significatività sufficientemente elevato (0,392) il quale, nel complesso, suggerisce che non vi sia nessuna prova per negare l'ipotesi delle linee parallele. Il test Sidak sequenziale inizia con singoli test Wald di contrasto per fornire un valore p globale e questi risultati dovrebbero essere paragonabili al risultato del test Wald onnicomprensivo. Il fatto che siano così diversi in questo esempio è abbastanza sorprendente ma potrebbe venire spiegato con l'esistenza di molti contrasti nel test e con gradi di libertà del disegno relativamente ridotti.

Figura 21-14
Stime dei parametri per il modello cumulativo generalizzato (illustrato parzialmente)

The legislature should enact a gas tax	Parametro	B	Errore standard	Intervallo di confidenza 95%	
				Inferiore	Superiore
Strongly agree	(Soglia)	-3,681	,221	-4,155	-3,207
	[agecat=1]	,320	,096	,525	,116
	[agecat=2]	-,075	,071	-,227	,077
	[agecat=3]	-,022	,073	-,180	,135
	[agecat=4]	,000 ^a	.	.	.
	[gender=0]	-,082	,054	-,197	,033
	[gender=1]	,000 ^a	.	.	.
	[votelas=0]	,008	,052	-,104	,120
	[votelas=1]	,000 ^a	.	.	.
	[drivefreq=1]	-4,096	,267	-4,669	-3,523
	[drivefreq=2]	-3,367	,237	-3,876	-2,857
	[drivefreq=3]	-2,678	,224	-3,158	-2,199
	[drivefreq=4]	-1,928	,213	-2,384	-1,471
	[drivefreq=5]	-1,015	,252	-1,555	-,476
[drivefreq=6]	,000 ^a	.	.	.	
Agree	(Soglia)	-1,963	,153	-2,291	-1,635
	[agecat=1]	-,385	,095	-,587	-,182
	[agecat=2]	-,130	,069	-,279	,018
	[agecat=3]	-,139	,101	-,356	,077
	[agecat=4]	,000 ^a	.	.	.
	[gender=0]	-,004	,040	-,090	,082
	[gender=1]	,000 ^a	.	.	.
	[votelas=0]	,009	,059	-,117	,135
	[votelas=1]	,000 ^a	.	.	.
	[drivefreq=1]	-3,867	,318	-4,549	-3,185
	[drivefreq=2]	-3,005	,175	-3,380	-2,630
	[drivefreq=3]	-2,290	,187	-2,691	-1,888
	[drivefreq=4]	-1,633	,166	-1,988	-1,278

Inoltre i valori stimati dei coefficienti del modello generalizzato non sembrano discostarsi molto dalle stime ottenute con l'ipotesi delle linee parallele.

Rimozione dei predittori non significativi

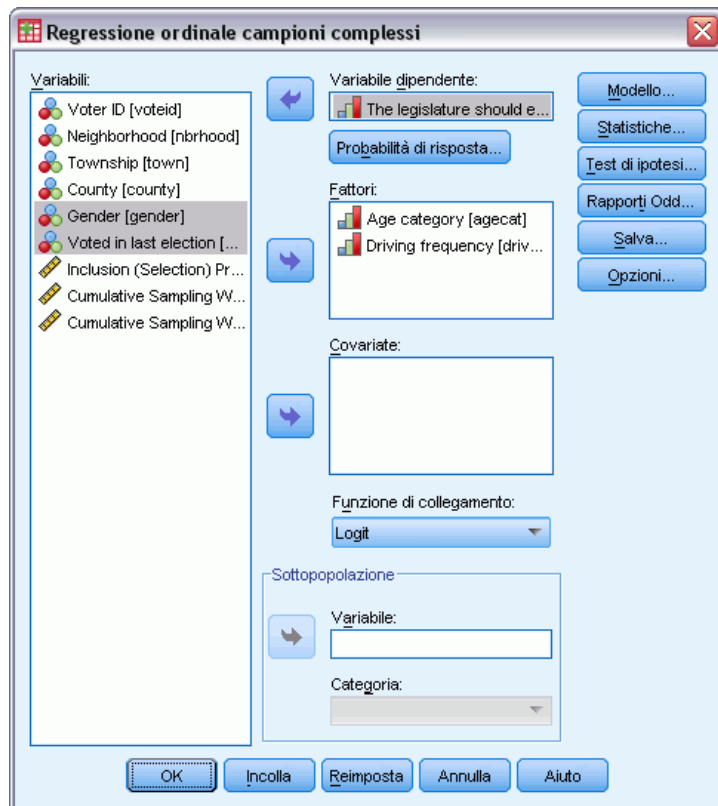
I test degli effetti del modello hanno dimostrato che i coefficienti del modello relativi a *Gender* (Sesso) e *Voted in last election* (Votato in ultima elezione) non sono statisticamente significativamente diversi da 0.

- Per produrre un modello ridotto, richiamare la finestra di dialogo *Regressione ordinale campioni complessi*.

- Fare clic su Continua nella finestra di dialogo Piano.

Figura 21-15

Finestra di dialogo Regressione ordinale



- Deselezionare i fattori *Sesso* e *Voted in last election*.
- Fare clic su Opzioni.

Figura 21-16
Finestra di dialogo *Regressione ordinale: Opzioni*

- Selezionare *Visualizza cronologia iterazioni*.

La cronologia delle iterazioni risulta utile per la risoluzione dei problemi riscontrati dall'algorithm di stima.

- Fare clic su *Continua*.
- Fare clic su *OK* nella finestra di dialogo *Regressione ordinale campioni complessi*.

Avvisi

Figura 21-17
Avvisi per il modello ridotto

The log-likelihood value cannot be increased after the maximum number of steps in the step-halving method.
La procedura CSORDINAL continua nonostante gli avvisi sopra riportati. I successivi risultati descritti sono basati sull'ultima iterazione. La validità dell'adattamento del modello non è sicura.
Il seguente messaggio si applica la modello cumulativo generalizzato
The log-likelihood value cannot be increased after the maximum number of steps in the step-halving method.

Gli avvisi sottolineano come la stima del modello ridotto sia terminata prima che le stime dei parametri abbiano raggiunto la convergenza perché non è stato possibile aumentare la verosimiglianza con nessuna modifica, o passo, nei valori correnti delle stime dei parametri.

Figura 21-18
Avvisi per il modello ridotto

Numero di iterazione ^a	N dimezzamenti	Pseudo -2 Log verosimiglianza	Soglia			Regressione								
			[opinion_gastax =1]	[opinion_gastax =2]	[opinion_gastax =3]	[ageca t=1]	[ageca t=2]	[ageca t=3]	[drivefr eq=1]	[drivefr eq=2]	[drivefr eq=3]	[drivefr eq=4]	[drivefr eq=5]	
0	0	326640.341	-1.309	-0.58	1.020	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	0	303567.549	-3.242	-1.881	-.704	-.323	-.137	-.094	-3.841	-2.970	-2.248	-1.563	-.835	
2	0	303336.336	-3.327	-1.897	-.664	-.325	-.139	-.095	-3.740	-2.998	-2.291	-1.568	-.811	
3	0	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812	
4	0	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812	
5 ^b	5	303335.933	-3.333	-1.900	-.664	-.326	-.139	-.096	-3.750	-3.003	-2.295	-1.570	-.812	

I parametri duplicati non vengono visualizzati. I rispettivi valori sono sempre pari a zero in tutte le iterazioni.
Variabile dipendente: The legislature should enact a gas tax (Crescente)
Modello: (Soglia), agecat, drivefr
Funzione di collegamento: Logit

a. È stato utilizzato il metodo Newton-Raphson per stimare i parametri.
b. Il valore di log-verosimiglianza non può essere aumentato oltre il numero massimo di fasi nel metodo di dimezzamento.

Osservando la cronologia delle iterazioni, le modifiche alle stime dei parametri relative alle ultime iterazioni sono di così scarsa rilevanza da non dover temere il messaggio di avviso.

Confronto di modelli

Figura 21-19
Pseudo R-quadrati per modello ridotto

Cox e Snell	,179
Nagelkerke	,191
McFadden	,071

Variabile dipendente: The legislature should enact a gas tax (Crescente)
Modello: (Soglia), agecat, gender, votelast, drivefr
Funzione di collegamento: Logit

I valori R^2 per il modello ridotto sono identici ai valori per il modello originale. Ciò è a favore del modello ridotto.

Figura 21-20
Tabella di classificazione del modello ridotto

Osservati	Previsti				Percentuale correzione
	Strongly agree	Agree	Disagree	Strongly disagree	
Strongly agree	7067.567	12130.814	3875.825	2058.750	28.1%
Agree	4271.234	14464.286	7320.767	6205.137	44.8%
Disagree	2024.816	11703.368	7108.487	8640.746	24.1%
Strongly disagree	889.869	8169.109	6946.522	15308.703	48.9%
Percentuale complessiva	12.1%	39.3%	21.4%	27.3%	37.2%

Variabile dipendente: The legislature should enact a gas tax (Crescente)
Modello: (Soglia), agecat, gender, votelast, drivefr
Funzione di collegamento: Logit

La tabella di classificazione complica leggermente la situazione. Il tasso globale di classificazione del 37,0% relativo al modello ridotto è paragonabile al modello originale, circostanza a favore del modello ridotto. Il modello ridotto, tuttavia, sposta la risposta prevista del 3,8% degli elettori da $Disagree$ a $Agree$, e si è osservato che più della metà di questi ha risposto $Disagree$ o $Strongly disagree$. Questa differenza è molto importante e merita un'attenta considerazione prima di scegliere il modello ridotto.

Riepilogo

La procedura Regressione ordinale campioni complessi ha consentito di costruire modelli concorrenti per il livello di sostegno alla proposta di legge in base ai dati demografici degli elettori. Il test delle linee parallele dimostra che un modello cumulativo generalizzato non è necessario. I test degli effetti del modello suggeriscono che *Sesso* e *Voted in last election* potrebbero essere omessi dal modello e il modello ridotto produce buoni risultati in termini di R^2 e tasso globale di classificazione rispetto al modello originale. Il modello ridotto tuttavia classifica in modo errato gli elettori nelle categorie *Agree/Disagree*, quindi per il momento i legislatori preferiscono mantenere il modello originale.

Procedure correlate

La procedura Regressione ordinale campioni complessi è un utile strumento per la creazione di modelli relativi a una variabile ordinale per casi estratti in base a uno schema di campionamento complesso.

- **Campionamento complesso guidato** viene utilizzato per determinare le specifiche di disegno del campionamento complesso ed ottenere un campione. Il file del piano di campionamento creato tramite Campionamento guidato contiene il piano analisi predefinito ed è possibile specificarlo nella finestra di dialogo Piano quando viene analizzato il campione ottenuto in base a tale piano.
- **Preparazione guidata dell'analisi Campioni complessi** viene utilizzata per definire le specifiche di analisi relative a un campione complesso esistente. È possibile specificare il file del piano analisi creato dal Campionamento guidato nella finestra di dialogo Piano quando viene analizzato il campione corrispondente a tale piano.
- La procedura **Modello lineare generalizzato campioni complessi** consente di creare il modello di una risposta scalare.
- La procedura **Regressione logistica campioni complessi** consente di creare il modello di una risposta categoriale.

Regressione Cox di campioni complessi

La procedura Regressione Cox di campioni complessi esegue l'analisi di sopravvivenza su campioni estratti da metodi di campionamento complessi.

Uso di Regressione Cox di campioni complessi: Definisci predictor dipendente dal tempo

L'impegno delle Forze dell'Ordine nel tentativo di valutare il tasso di recidività nella propria area di giurisdizione. Una delle misure di recidività è il tempo che trascorre fino al secondo arresto del trasgressore. L'agenzia vorrebbe creare un modello di tempo del nuovo arresto utilizzando la Regressione di Cox su un campione estratto tramite metodi di campionamento complessi ma il problema potrebbe essere l'invalidità dell'ipotesi della proporzionalità del rischio per le categorie di età.

Sono state selezionate in diversi dipartimenti campione le persone rilasciate al primo arresto nel mese di giugno 2003 e sono stati controllati fino alla fine di giugno 2006. Il campione è contenuto nel file *recidivism_cs_sample.sav*. Il piano di campionamento utilizzato è contenuto in *recidivism_cs.csplan*. Poiché fa uso del metodo PPS (probability-proportional-to-size, probabilità proporzionale alla dimensione), esiste anche un file contenente le probabilità di selezione congiunte (*recidivism_cs_jointprob.sav*). [Per ulteriori informazioni, vedere l'argomento File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Utilizzare la Regressione Cox di campioni complessi per valutare la validità dell'ipotesi della proporzionalità del rischio e generare un modello con predittori dipendenti dal tempo.

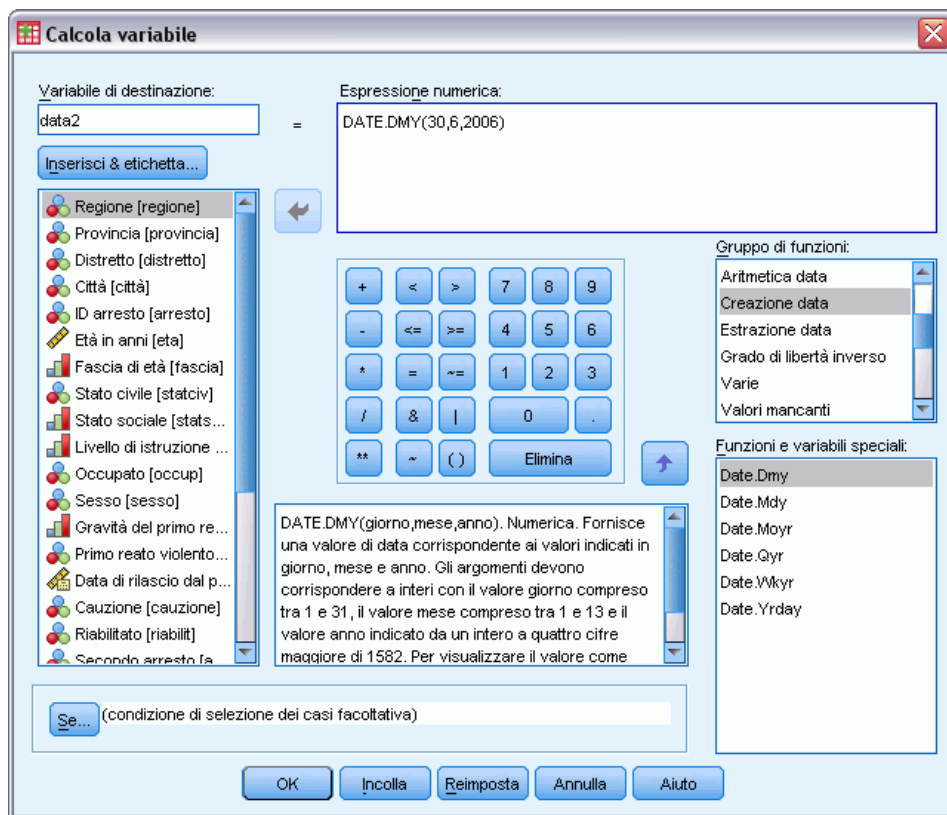
Preparazione dei dati

L'insieme di dati contiene le date di rilascio dopo il primo e dopo il secondo arresto; poiché la Regressione di Cox analizza il tempo di sopravvivenza, è necessario calcolare il tempo trascorso tra tali date.

Tuttavia, *Date of second arrest [date2]* (*Data del secondo arresto [data2]*) contiene i casi con il valore 03/10/1582, un valore mancante per le variabili della data. Queste sono persone che non hanno commesso un secondo reato e che vogliamo includere nel modello come casi troncati. La fine del periodo di controllo era il 30 giugno 2006, pertanto, è necessario ricodificare 03/10/1582 in 30/06/2006.

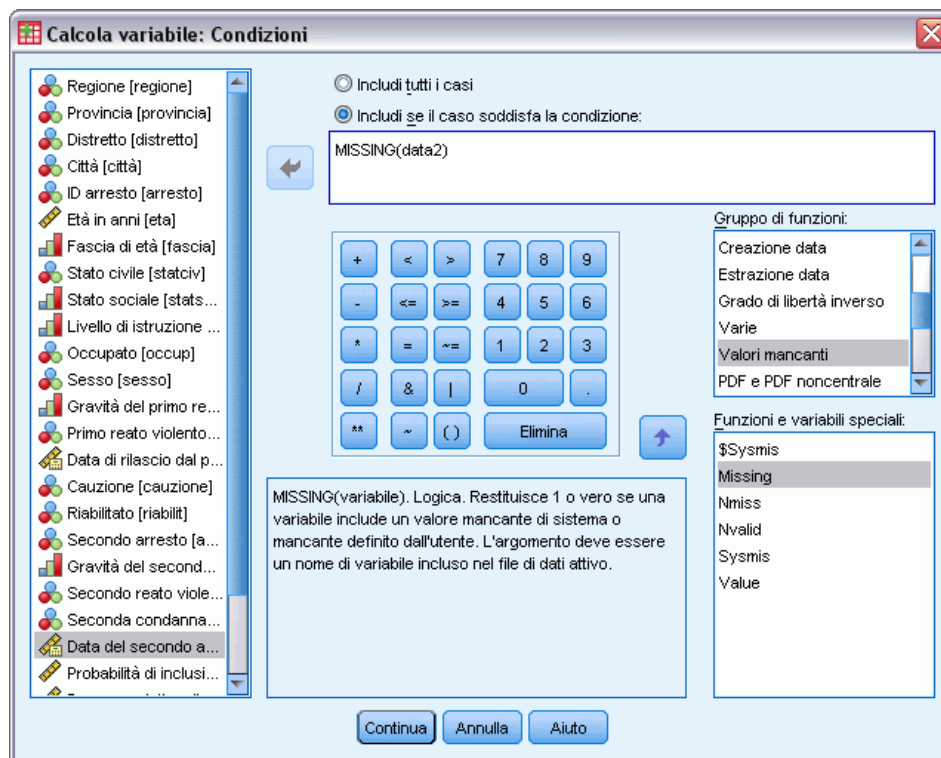
- Per ricodificare questi valori, dai menu scegliere:
Trasforma > Calcola variabile...

Figura 22-1
Finestra di dialogo *Calcola variabile*



- Digitare data2 (data2) come variabile di destinazione.
- Digitare DATE.DMY(6,30,2006) come espressione numerica.
- Fare clic su Se.

Figura 22-2
Finestra di dialogo Calcola variabile: Condizioni



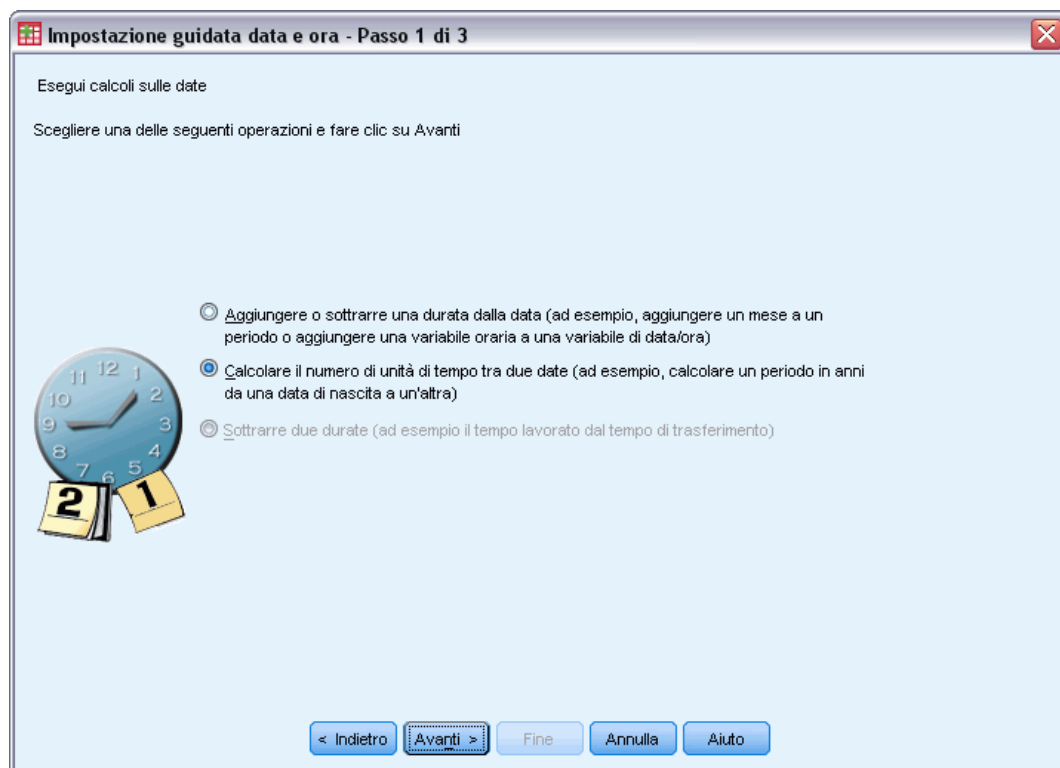
- ▶ Selezionare Includi se il caso soddisfa la condizione.
- ▶ Digitare MISSING(date2) (MANCANTE (data2)) come espressione numerica.
- ▶ Fare clic su Continua.
- ▶ Fare clic su OK nella finestra di dialogo Calcola variabile.
- ▶ Quindi, per calcolare il tempo trascorso tra il primo e il secondo arresto, dai menu scegliere: Trasforma > Procedura guidata Data e ora...

Figura 22-3
Impostazione guidata data e ora, finestra di benvenuto



- ▶ Selezionare Esegui calcoli con date e ore.
- ▶ Fare clic su Avanti.

Figura 22-4
Impostazione guidata data e ora, Eseguire calcoli sulle date



- ▶ Selezionare Calcolare il numero di unità di tempo tra due date.
- ▶ Fare clic su Avanti.

Figura 22-5
Impostazione guidata data e ora, Calcolare il numero di unità di tempo tra due date

Calcolare il numero di unità di tempo tra due variabili di data o di data/ora.

Il risultato sarà una variabile a numero intero. Tutte le parti frazionarie di un'unità verranno scartate. Il risultato sarà una variabile di durata. L'elenco di variabili sottostante visualizza solo le variabili di durata.

Variabili:

- Data e ora correnti [\$TI...]

Data1:

Data del secondo arresto [data2]

meno Data2:

Data di rilascio dal primo arresto [data1]

Unità:

Giorni

Trattamento risultato

Tronca a numero intero

Arrotonda a numero intero

Mantieni parte frazionaria

Per le unità di mese e anno, il risultato è basato sulla lunghezza media dell'unità, a meno che non venga utilizzato il troncamento.

\$TIME rappresenta la data e l'ora correnti.

< Indietro Avanti > Fine Annulla Aiuto

- ▶ Selezionare *Date of second arrest [date2]* (*Data del secondo arresto [data2]*) come prima data.
- ▶ Selezionare *Date of release from first arrest [date1]* (*Data di rilascio dal primo arresto [data1]*) come data da sottrarre dalla prima data.
- ▶ Selezionare Giorni come unità.
- ▶ Fare clic su Avanti.

Figura 22-6
Impostazione guidata data e ora, Calcolo

Calcolo: data2 - data1

Variabile risultante: tempo_per_caso Unità: Giorni

Etichetta variabile: Secondo il tempo di arresto

Esecuzione

Crea la variabile ora Incolla la sintassi nell'apposita finestra

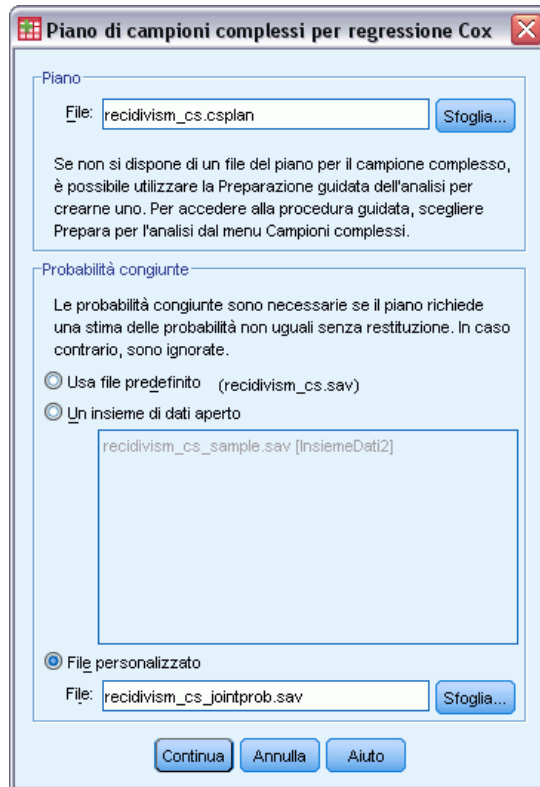
< Indietro Avanti > Fine Annulla Aiuto

- ▶ Digitare *time_to_event* (*tempo_fino_a_evento*) come nome della variabile che rappresenta il tempo tra le due date.
- ▶ Digitare *Time to second arrest* (*Tempo fino al secondo arresto*) come etichetta della variabile.
- ▶ Fare clic su Fine.

Esecuzione dell'analisi

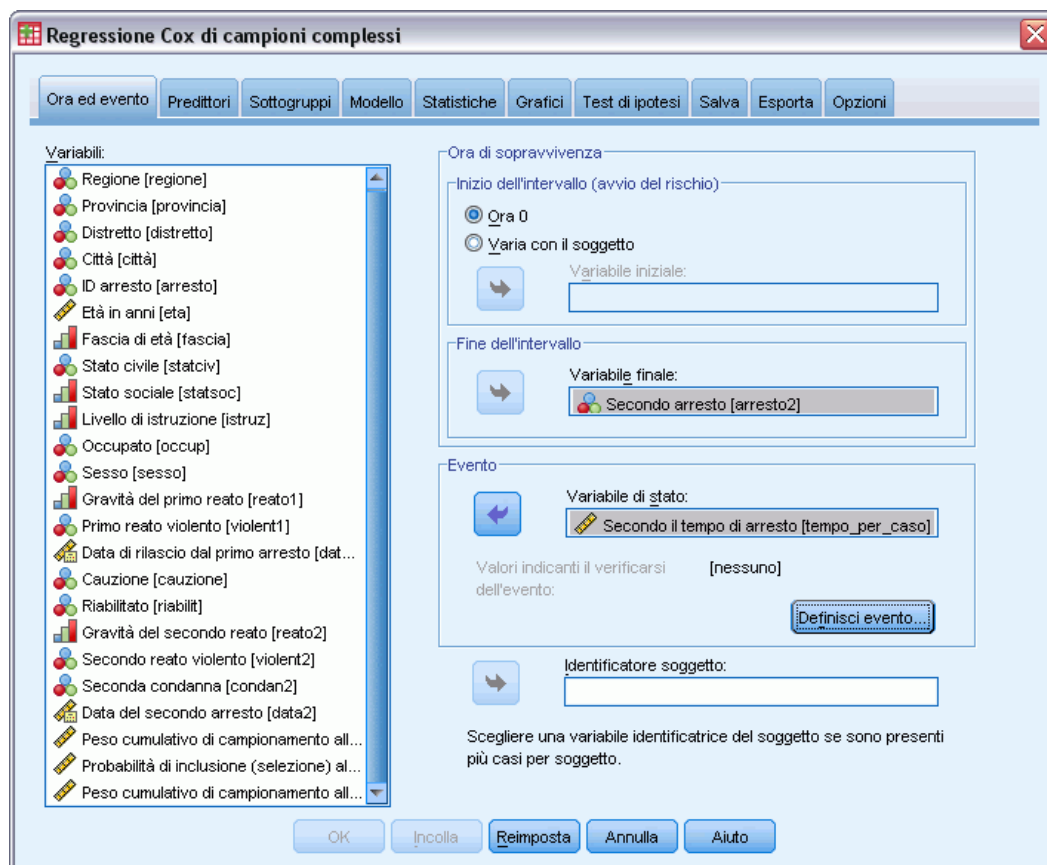
- ▶ Per eseguire un'analisi Regressione Cox di campioni complessi, dai menu scegliere:
Analizza > Complex Samples > Regressione di Cox...

Figura 22-7
Finestra di dialogo Piano di campioni complessi per regressione Cox



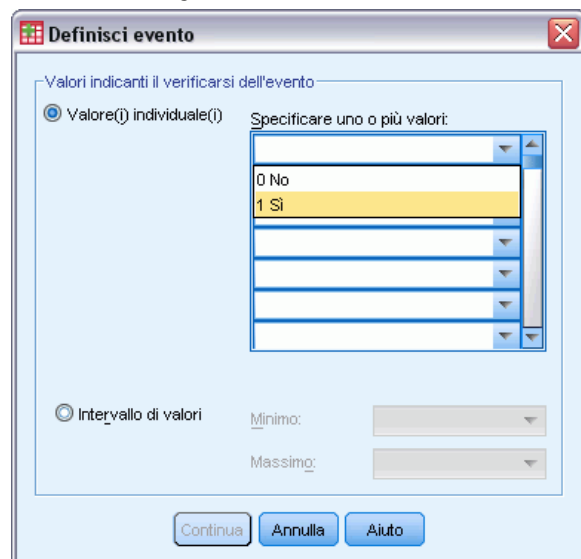
- ▶ Accedere alla directory dei file di esempio e selezionare *recidivism_cs.csplan* come file del piano.
- ▶ Selezionare File personalizzato nel gruppo Probabilità congiunte, accedere alla directory dei file di esempio e selezionare *recidivism_cs_jointprob.sav*.
- ▶ Fare clic su Continua.

Figura 22-8
Finestra di dialogo regressione di Cox, scheda Ora ed evento



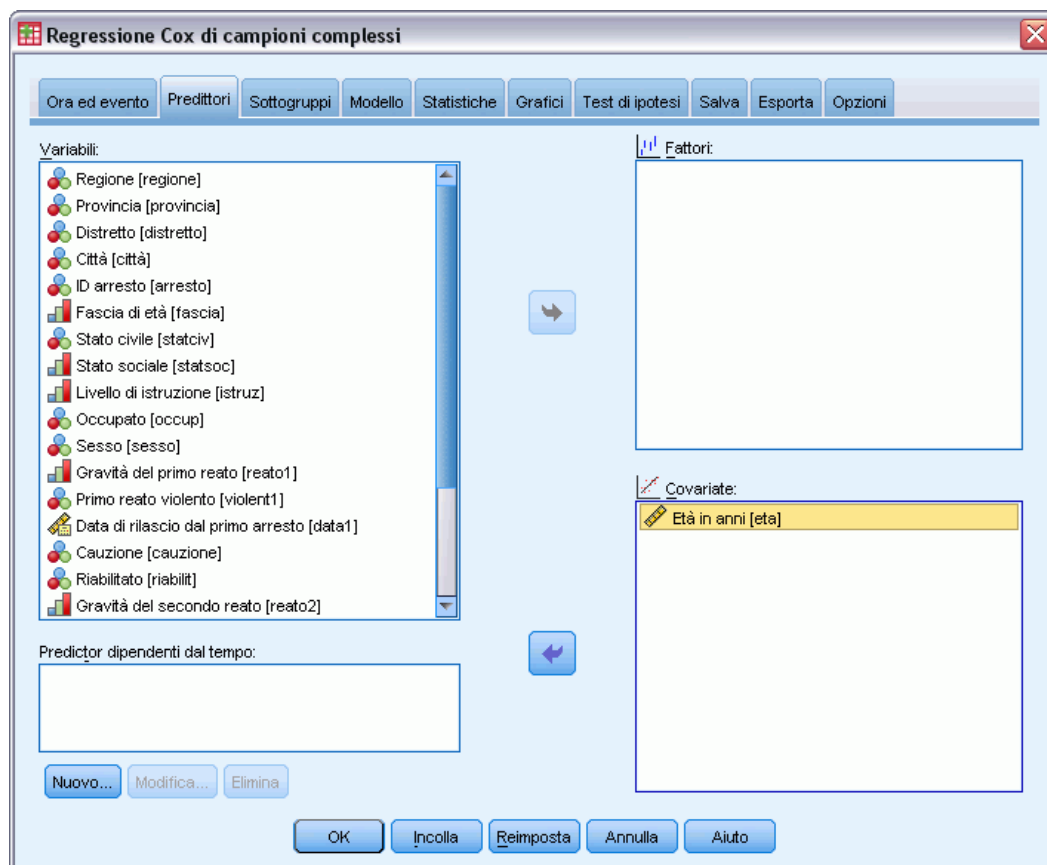
- ▶ Selezionare *Time to second arrest* (*Tempo fino al secondo arresto [tempo_fino_a_evento]*) come variabile che definisce la fine dell'intervallo.
- ▶ Selezionare *Second arrest* [*arresto2*] (*Secondo arresto [arresto2]*) come variabile che definisce se l'evento si è verificato o meno.
- ▶ Fare clic su Definisci evento.

Figura 22-9
Finestra di dialogo Definisci evento



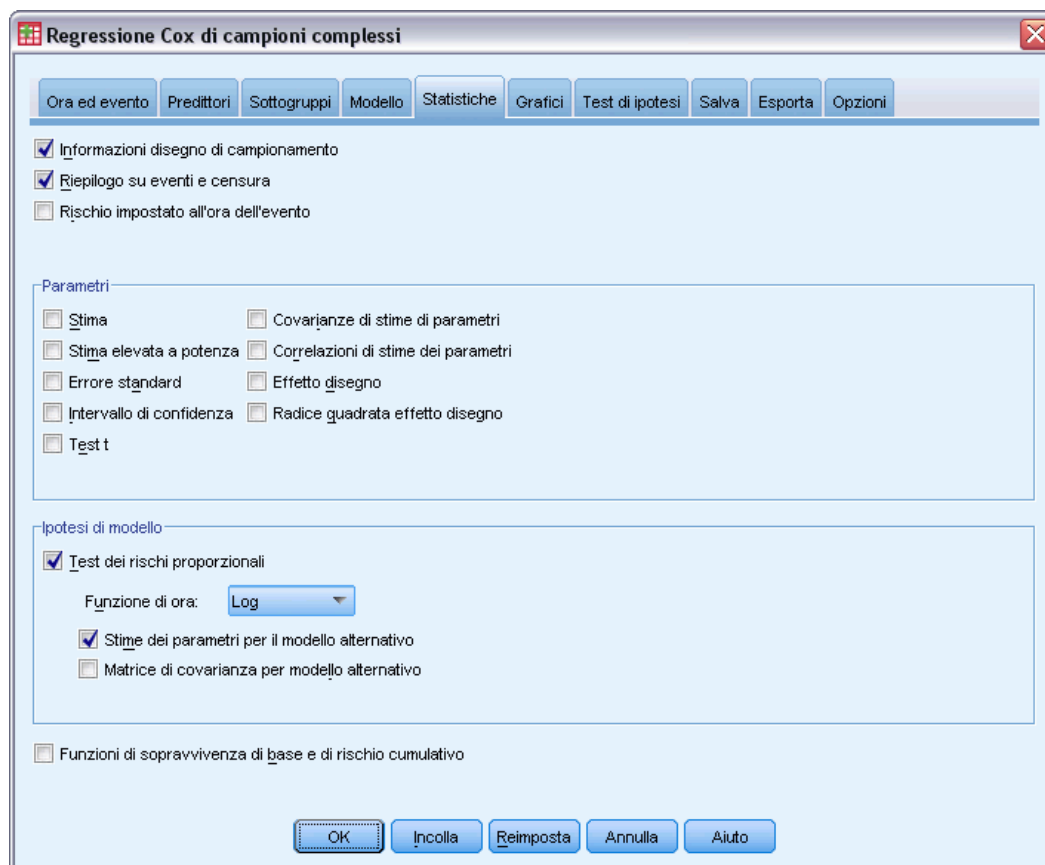
- ▶ Selezionare 1 Yes (1 Si) come valore che indica il verificarsi dell'evento di interesse (nuovo arresto).
- ▶ Fare clic su Continua.
- ▶ Fare clic sulla scheda Predittori.

Figura 22-10
Finestra di dialogo Regressione di Cox, scheda Predittori



- ▶ Selezionare *Age in years [age]* (*Età in anni [età]*) come covariata.
- ▶ Fare clic sulla scheda Statistiche.

Figura 22-11
Finestra di dialogo *Regressione di Cox*, scheda *Statistiche*



- ▶ Selezionare Test dei rischi proporzionali, quindi Log come funzione nel gruppo Ipotesi di modello.
- ▶ Selezionare Stime dei parametri per il modello alternativo.
- ▶ Fare clic su OK.

Informazioni disegno di campionamento

Figura 22-12
Informazioni disegno di campionamento

			N
Conteggi non ponderati	Validi	Soggetti	5687
		Casi	5687
		Casi non validi	0
		Casi totali	5687
Validi		Dimensione soggetti popolazione	307583.898
Stadio 1	Validi	Strati	4
		Unità	20
Validi		Gradi di libertà del disegno campionario	16

Questa tabella contiene le informazioni sul disegno di campionamento relative alla stima del modello.

- È presente un caso per soggetto e tutti i 5.687 casi vengono utilizzati nell'analisi.
- Il campione rappresenta meno del 2% dell'intera popolazione stimata.
- Il disegno ha richiesto 4 strati e 5 unità per strato per un totale di 20 unità nel primo stadio del disegno. I gradi di libertà del disegno di campionamento sono stimati in $20-4=16$.

Test degli effetti del modello

Figura 22-13
Test degli effetti del modello

Sorgente	df1	df2	F di Wald	Correzione per confronti multipli
age	1.000	16.000	504.787	1.580E-13

Variabile ora di sopravvivenza: Time to second arrest
Variabile di stato evento: Second arrest = 1
Modello: age

Nel modello di proporzionalità del rischio, il valore di significatività di 'età del predittore è inferiore a 0,05 e pertanto, sembra contribuire al modello.

Test dei rischi proporzionali

Figura 22-14
Test complessivo di rischi proporzionali

df1	df2	F di Wald	Correzione per confronti multipli
1.000	16.000	29.924	5.136E-5

Variabile ora di sopravvivenza: Time to second arrest
Variabile di stato evento: Second arrest = 1
Modello: age, age*_TF

Figura 22-15
Stime dei parametri per il modello alternativo

Parametro	B	Errore standard	Intervallo di confidenza 90%	
			Inferiore	Superiore
age	-.002	.014	-.025	.022
age*_TF ^a	-.012	.002	-.016	-.009

Variabile ora di sopravvivenza: Time to second arrest
Variabile di stato evento: Second arrest = 1
Modello: age, age*_TF

a. Funzione di ora: Log

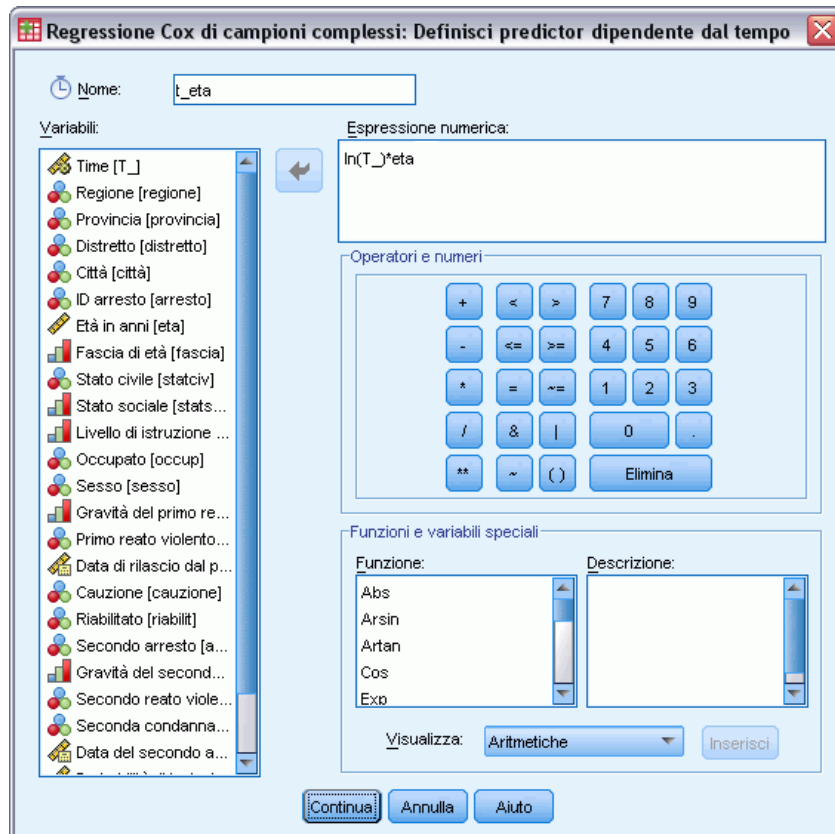
Il valore di significatività del test complessivo di rischi proporzionali è inferiore a 0,05; questo indica che l'ipotesi della proporzionalità del rischio è stata violata. La funzione log time viene utilizzata per il modello alternativo, pertanto risulterà facile replicare questo predittore dipendente dal tempo.

Aggiunta di un predittore dipendente dal tempo

- ▶ Aprire la finestra di dialogo Regressione Cox di campioni complessi e fare clic sulla scheda Predittori.
- ▶ Fare clic su Nuovo.

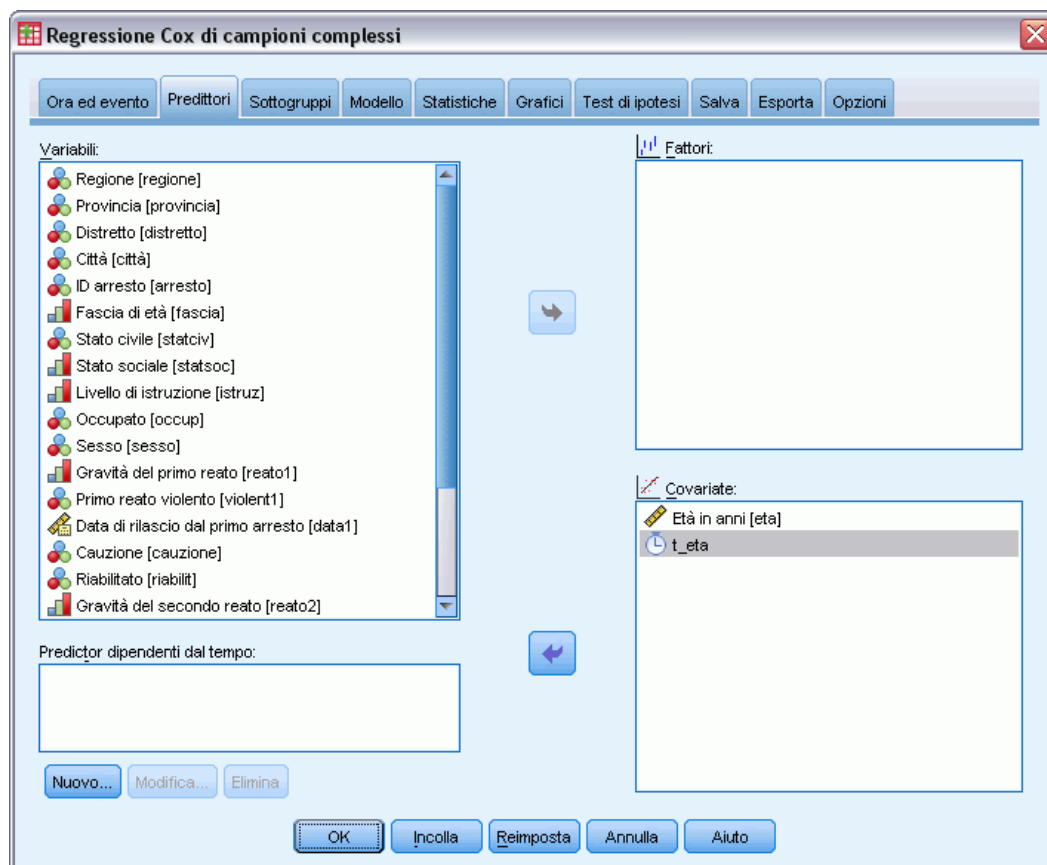
Figura 22-16

Finestra di dialogo Regressione Cox di campioni complessi: Definisci predictor dipendente dal tempo



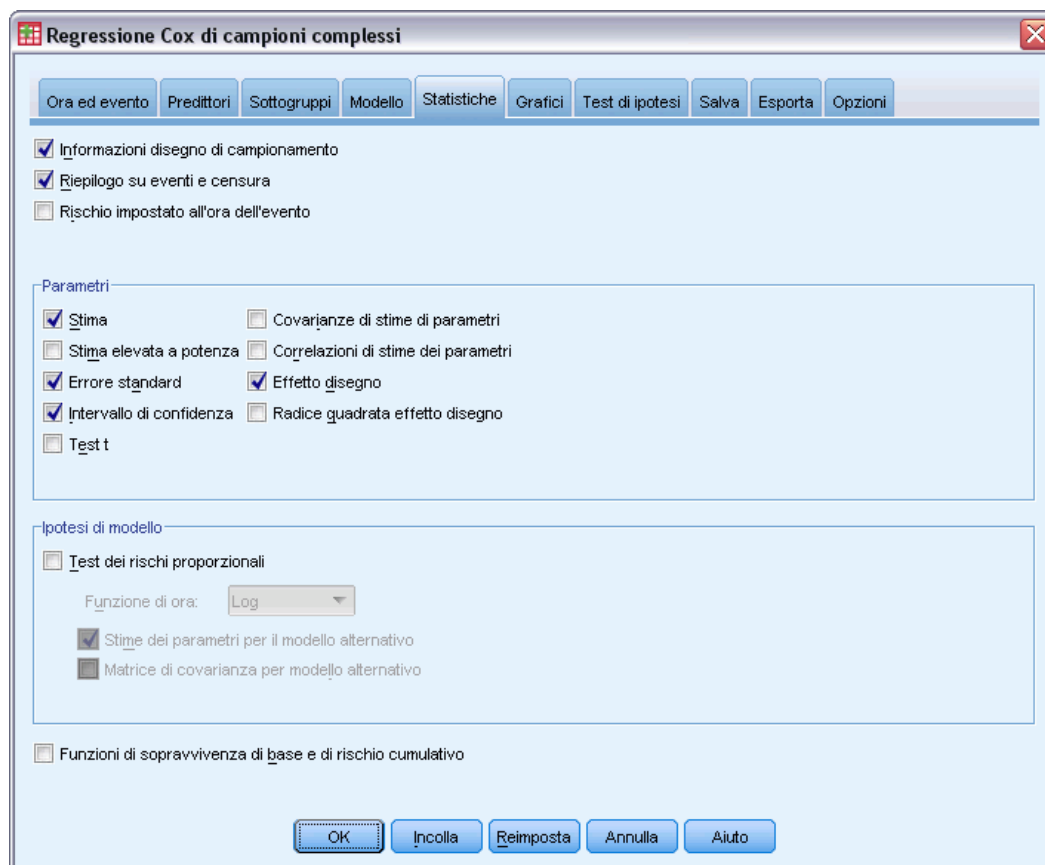
- ▶ Digitare t_age (t_età) come nome del predittore dipendente dal tempo da definire.
- ▶ Digitare ln(T_)*age (ln(T_)*età) come espressione numerica.
- ▶ Fare clic su Continua.

Figura 22-17
Finestra di dialogo Regressione di Cox, scheda Predittori



- Selezionare t_age ($t_età$) come covariata.
- Fare clic sulla scheda Statistiche.

Figura 22-18
Finestra di dialogo *Regressione di Cox*, scheda *Predittori*



- ▶ Selezionare *Stima*, *Errore standard*, *Intervallo di confidenza* e *Effetto disegno* nel gruppo *Parametri*.
- ▶ Deselezionare *Test dei rischi proporzionali*, quindi *Stime dei parametri per il modello alternativo* nel gruppo *Ipotesi di modello*.
- ▶ Fare clic su *OK*.

Test degli effetti del modello

Figura 22-19
Test degli effetti del modello

Sorgente	df1	df2	F di Wald	Correzione per confronti multipli
age	1.000	16.000	.015	0.91
age* _{TF} ^a	1.000	16.000	29.924	5.136E-5

Variabile ora di sopravvivenza: Time to second arrest
Variabile di stato evento: Second arrest= 1
Modello: age, age*_{TF}

Con l'aggiunta del predittore dipendente dal tempo, il valore di significatività di *age* (*età*) è 0,91; questo indica che il contributo per il modello è sostituito da quello di *t_age* (*t_età*).

Stime di parametri

Figura 22-20
Stime dei parametri

Parametro	B	Errore standard	Intervallo di confidenza 90%		Effetto disegno
			Inferiore	Superiore	
age	-.002	.014	-.030	.027	.702
age*_TF ^a	-.012	.002	-.017	-.008	.666

Variabile ora di sopravvivenza: Time to second arrest
 Variabile di stato evento: Second arrest = 1
 Modello: age, age*_TF

Dall'osservazione delle stime dei parametri e degli errori standard, è possibile notare che è stato replicato il modello alternativo dal test dei rischi proporzionali. Indicando esplicitamente il modello, è possibile richiedere statistiche dei parametri e grafici aggiuntivi. Qui è stato richiesto l'effetto disegno; il valore di t_{age} inferiore a 1 indica che l'errore standard di t_{age} ($t_{età}$) è inferiore a quello che si sarebbe ottenuto nel caso l'insieme di dati fosse un semplice campione casuale. In questo caso, l'effetto di t_{age} ($t_{età}$) sarebbe ancora significativo a livello statistico ma gli intervalli di confidenza sarebbero più ampi.

Più casi per soggetto in Regressione Cox di campioni complessi

I ricercatori che analizzano i tempi di sopravvivenza dei pazienti che escono da un programma di riabilitazione da attacchi ischemici devono affrontare una serie di difficoltà.

Più casi per soggetto. Le variabili che rappresentano l'anamnesi del paziente dovrebbero essere utili come predittori. Nel tempo, i pazienti potrebbero subire eventi medici più gravi che alterano la loro anamnesi. In questo insieme di dati, sono indicati i casi di infarto del miocardio, di attacco ischemico o di emorragia celebrale ed è indicato anche il periodo in cui si è verificato l'evento. È possibile creare nella procedura covariate calcolabili dipendenti dal tempo per includere tali informazioni nel modello ma sarebbe meglio utilizzare più casi per soggetto. Notare che le variabili in origine erano codificate è pertanto l'anamnesi del paziente è registrata nelle variabili; sarà necessario quindi ristrutturare l'insieme di dati.

Troncamento a sinistra. L'avvio del rischio inizia al momento dell'attacco ischemico. Tuttavia, il campione include solo i pazienti che hanno superato il programma di riabilitazione, quindi il campione è troncato a sinistra nel senso che i tempi di sopravvivenza osservati sono "gonfiati" dalla lunghezza del periodo di riabilitazione. È possibile rappresentare questo specificando il periodo in cui i pazienti hanno concluso la riabilitazione come il periodo di ingresso nel caso di studio.

Nessun piano di campionamento. L'insieme di dati non è stato raccolto tramite un piano di campionamento complesso ed è considerato un semplice campione casuale. Sarà necessario creare un piano di analisi per utilizzare la Regressione Cox di campioni complessi.

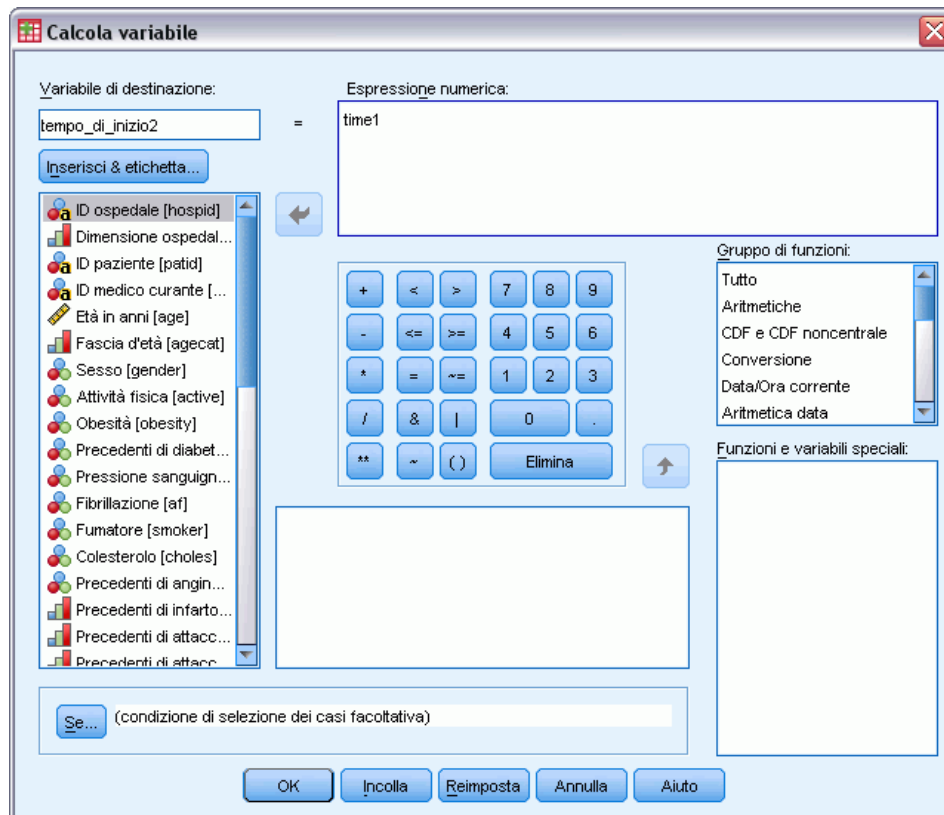
L'insieme di dati è raccolto nel file *stroke_survival.sav*. Per ulteriori informazioni, vedere l'argomento [File di esempio in l'appendice A in IBM SPSS Complex Samples 20](#). Utilizzare Ristrutturazione di dati guidata per preparare i dati all'analisi, quindi Preparazione guidata dell'analisi per creare un piano di campionamento casuale semplice e infine la Regressione Cox di campioni complessi per creare un modello per i tempi di sopravvivenza.

Preparazione dei dati per l'analisi

Prima di ristrutturare i dati, sarà necessario creare due variabili ausiliarie di aiuto nella ristrutturazione.

- Per calcolare una nuova variabile, dai menu scegliere:
Trasforma > Calcola variabile...

Figura 22-21
Finestra di dialogo Calcola variabile

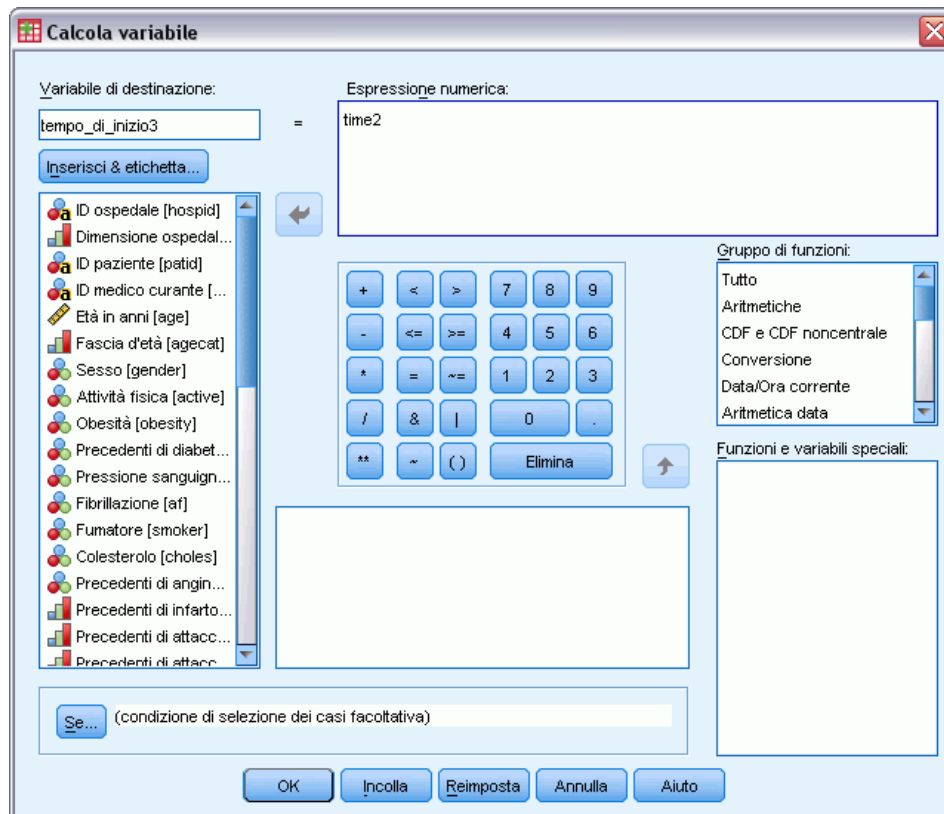


- Digitare *start_time2* (*periodo_iniziale*) come variabile di destinazione.
- Digitare *time1* (*periodo1*) come espressione numerica.
- Fare clic su OK.

- Richiamare la finestra di dialogo Calcola variabile.

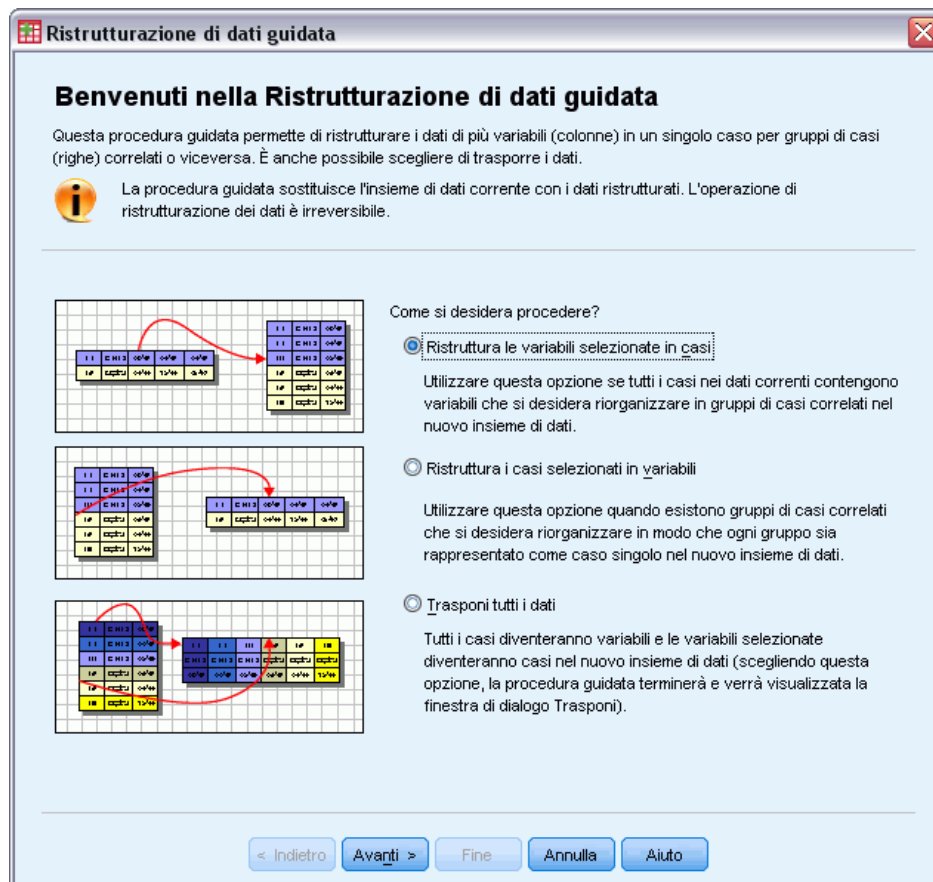
Figura 22-22

Finestra di dialogo Calcola variabile



- Digitare *start_time3* (*periodo_iniziale3*) come variabile di destinazione.
- Digitare *time2* (*periodo2*) come espressione numerica.
- Fare clic su OK.
- Per ristrutturare i dati dalle variabili ai casi, dai menu scegliere:
Dati > Ristruttura...

Figura 22-23
Ristrutturazione di dati guidata, finestra di benvenuto



- ▶ Verificare che l'opzione Ristruttura i casi selezionati in variabili sia selezionata.
- ▶ Fare clic su Avanti.

Figura 22-24

Ristrutturazione di dati guidata, Variabili in casi: Numero di gruppi di variabili

Ristrutturazione di dati guidata - Passo 2 di 7

Variabili in casi: Numero di gruppi di variabili

Si è scelto di ristrutturare le variabili selezionate in gruppi di casi collegati nel file nuovo.

Un gruppo di variabili collegate, denominato gruppo di variabili, rappresenta misurazioni in una variabile.

Ad esempio, la variabile potrebbe riferirsi alla larghezza. Se è registrata in tre diverse misurazioni, ognuna delle quali rappresenta un diverso punto temporale (w1, w2 e w3), i dati vengono organizzati in un gruppo di variabili.

Se esistono più variabili nel file, spesso vengono registrate anch'esse in un gruppo di variabili, ad esempio l'altezza, registrata in h1, h2 e h3.

Quanti gruppi di variabili si desidera ristrutturare?

Uno (ad esempio, a1, a2 e a3)

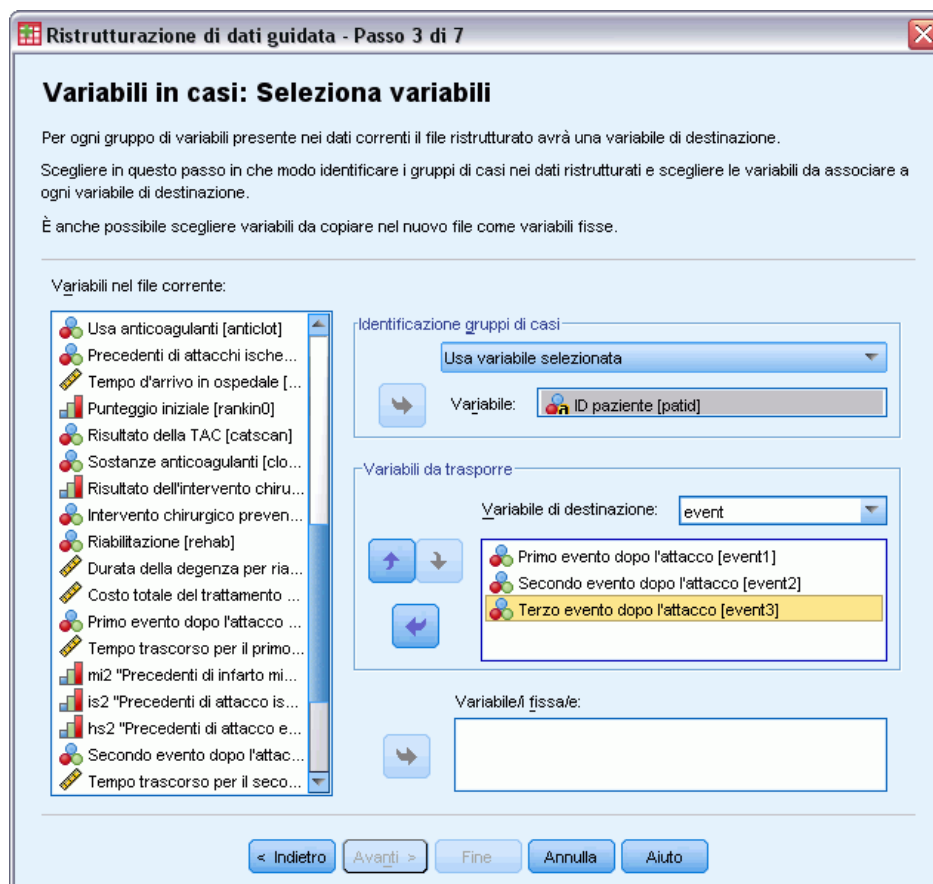
Più di uno (ad esempio, a1, a2, a3 e h1, h2, h3, ecc.)

Quanti?

< Indietro Avanti > Fine Annulla Aiuto

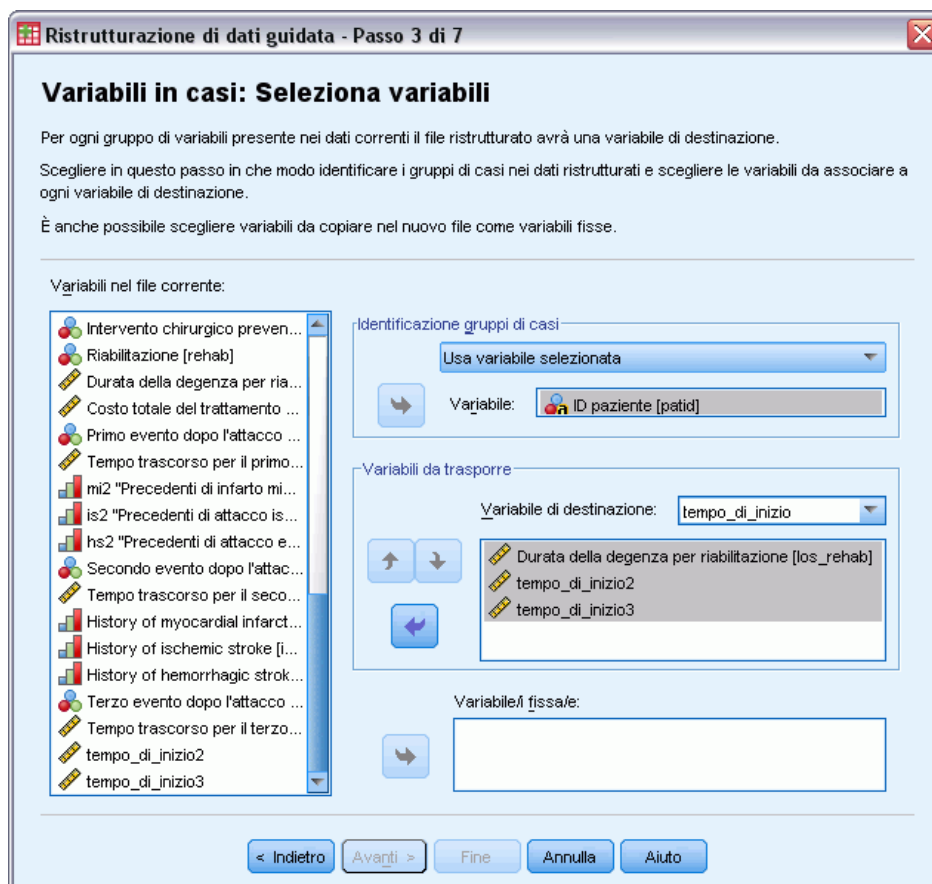
- ▶ Selezionare il gruppo di variabili Più di uno da ristrutturare.
- ▶ Digitare 6 come numero di gruppi.
- ▶ Fare clic su Avanti.

Figura 22-25
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



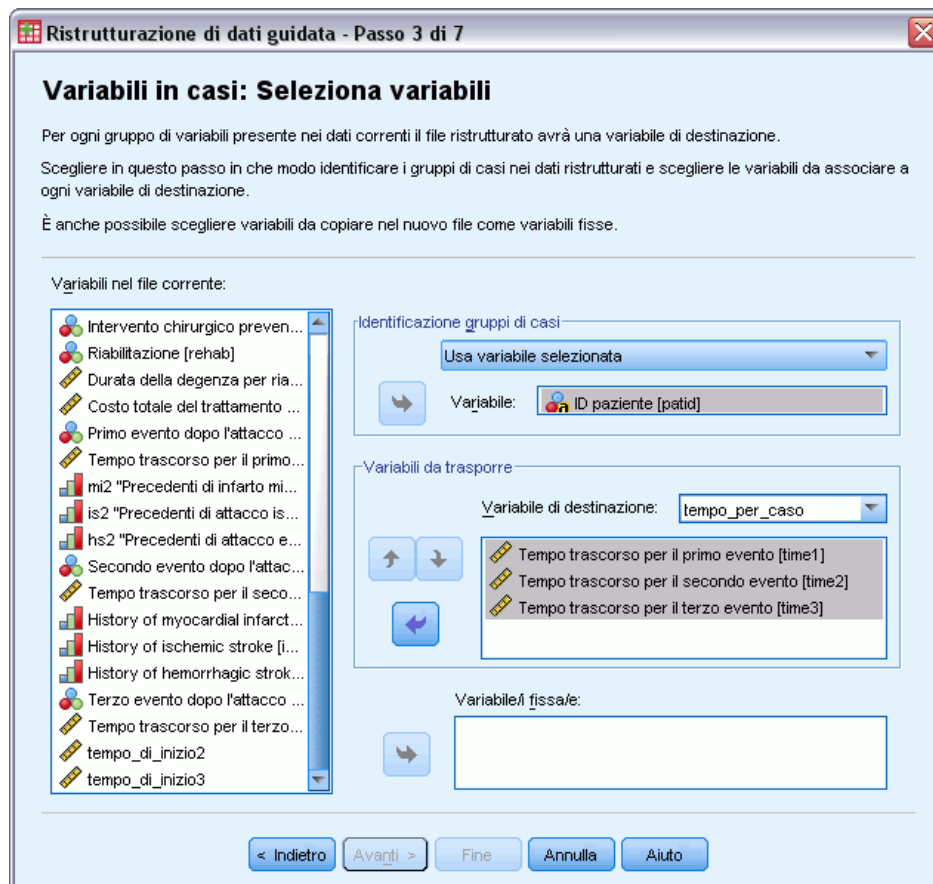
- ▶ Nel gruppo Identificazione gruppi di casi, selezionare Usa variabile selezionata e *Patient ID [patid]* (*ID paziente [idpaz]*) come identificatore dei soggetti.
- ▶ Digitare event (evento) come prima variabile di destinazione.
- ▶ Selezionare *First event post-attack [event1]* (*Primo evento post-attacco [evento1]*), *Second event post-attack [event2]* (*Secondo evento post-attacco [evento2]*) e *Third event post-attack [event3]* (*Terzo evento post-attacco [evento3]*) come variabili da trasportare.
- ▶ Selezionare *trans2 (trasp2)* dall'elenco delle variabili di destinazione.

Figura 22-26
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



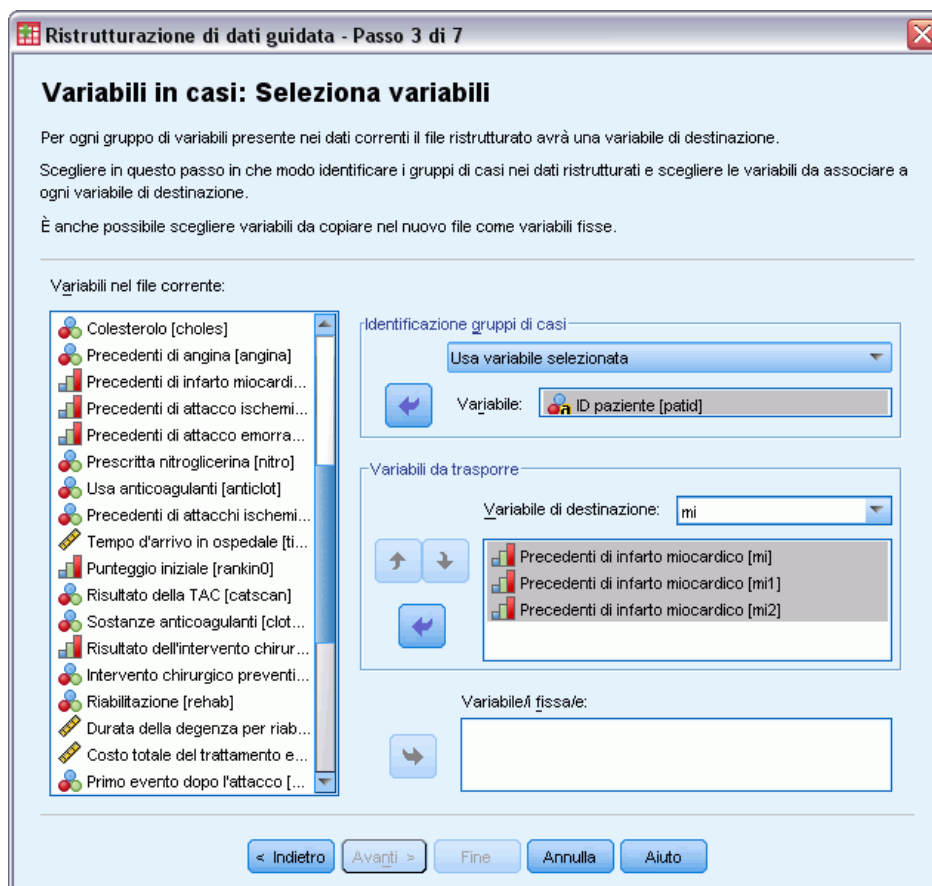
- ▶ Digitare `start_time` (`periodo_iniziale`) come variabile di destinazione.
- ▶ Selezionare *Length of stay for rehabilitation* [`los_rehab`] (*Durata della riabilitazione* [`dur_rehab`]), *start_time2* (`periodo_iniziale2`) e *start_time3* (`periodo_iniziale3`) come variabili da trasportare. *Time to first event post-attack* [`time1`] (*Periodo primo evento post-attacco* [`periodo1`]) e *Time to second event post-attack* [`time2`] (*Periodo secondo evento post-attacco* [`periodo2`]) verranno utilizzati per creare i tempi finali e ogni variabile può essere presente in un solo elenco di variabili da trasportare, quindi *start_time2* (`periodo_iniziale`) e *start_time3* (`periodo_iniziale3`) erano valori necessari.
- ▶ Selezionare *trans3* (`trasp3`) dall'elenco delle variabili di destinazione.

Figura 22-27
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



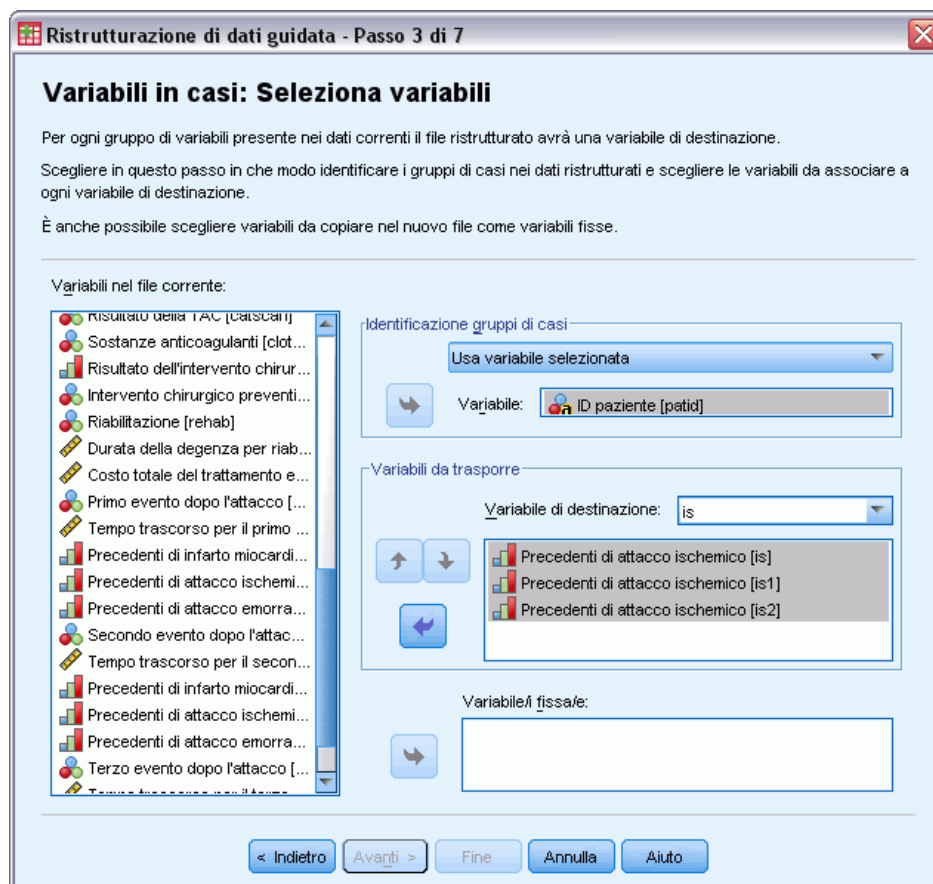
- ▶ Digitare `time_to_event` (`tempo_fino_a_evento`) come variabile di destinazione.
- ▶ Selezionare *First event post-attack [time1]* (*Primo evento post-attacco [periodo1]*), *Second event post-attack [time2]* (*Secondo evento post-attacco [periodo2]*) e *Third event post-attack [time3]* (*Terzo evento post-attacco [periodo3]*) come variabili da trasportare.
- ▶ Selezionare *trans4* (*trasp4*) dall'elenco delle variabili di destinazione.

Figura 22-28
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



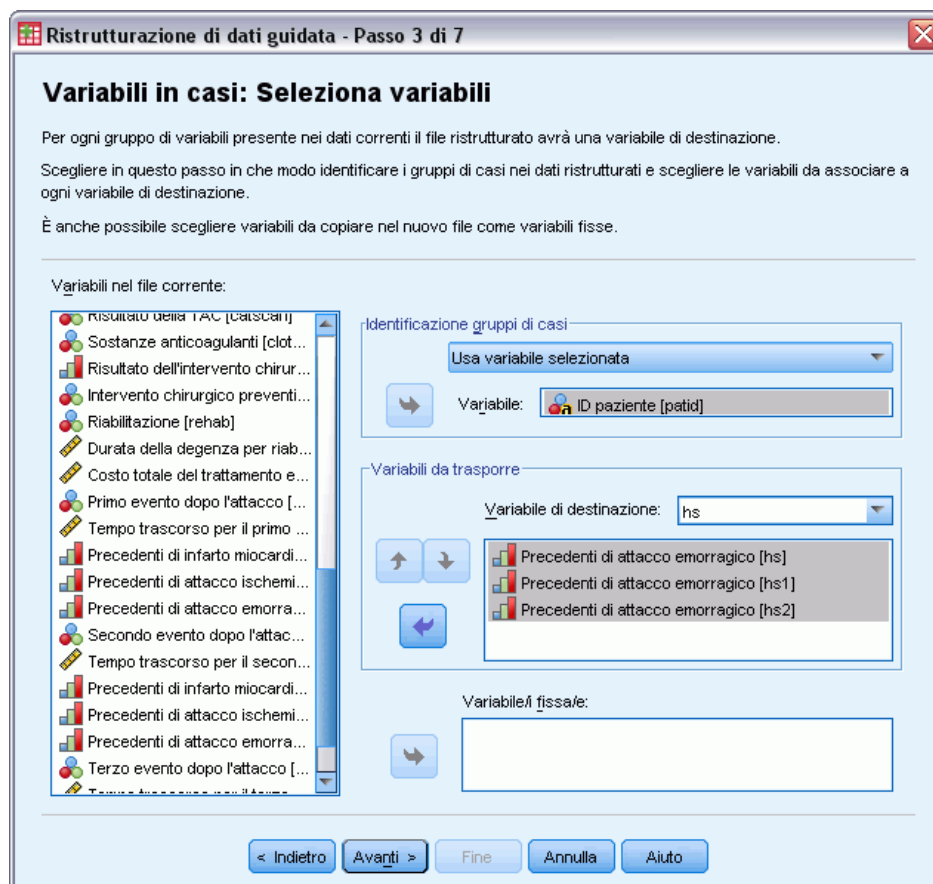
- ▶ Digitare mi (inf) come variabile di destinazione.
- ▶ Selezionare *History of myocardial infarction [mi]* (*Anamnesi infarto del miocardio [inf]*), *History of myocardial infarction [mi1]* (*Anamnesi infarto del miocardio [inf1]*) e *History of myocardial infarction [mi2]* (*Anamnesi infarto del miocardio [inf2]*) come variabili da trasporre.
- ▶ Selezionare *trans5 (trasp5)* dall'elenco delle variabili di destinazione.

Figura 22-29
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



- ▶ Digitare is come variabile di destinazione.
- ▶ Selezionare *History of ischemic stroke [is]* (*Anamnesi dell'attacco ischemico [is]*), *History of ischemic stroke [is1]* (*Anamnesi dell'attacco ischemico [is1]*) e *History of ischemic stroke [is2]* (*Anamnesi dell'attacco ischemico [is2]*) come variabili da trasporre.
- ▶ Selezionare *trans6 (trasp6)* dall'elenco delle variabili di destinazione.

Figura 22-30
Ristrutturazione di dati guidata, Variabili in casi: Seleziona variabili



- ▶ Digitare hs (em) come variabile di destinazione.
- ▶ Selezionare *History of hemorrhagic stroke [hs]* (*Anamnesi emorragia cerebrale [em]*), *History of hemorrhagic stroke [hs1]* (*Anamnesi emorragia cerebrale [em1]*) e *History of hemorrhagic stroke [hs2]* (*Anamnesi emorragia cerebrale [em2]*) come variabili da trasportare.
- ▶ Fare clic su Avanti, quindi ancora su Avanti nel passo Crea variabili indice.

Figura 22-31
Ristrutturazione di dati guidata, Variabili in casi: Crea una variabile indice

Variabili in casi: Crea una variabile indice

Si è scelto di creare una sola variabile indice. I valori della variabile possono essere numeri in sequenza o i nomi delle variabili in un gruppo.

Nella tabella è possibile specificare il nome e l'etichetta per la variabile indice.

Che tipo di valori indice?

Numeri sequenziali
Valori indice: 1, 2, 3

Nomi di variabile
Valori indice: event1, event2, event3

Modifica l'etichetta e il nome della variabile indice:

	Nome	Etichetta	Livelli	Valori indice
1	caso_indice	Caso Indice	3	1, 2, 3

< Indietro Avanti > Fine Annulla Aiuto

- ▶ Digitare event_index (indice_evento) come nome della variabile indice e digitare Event index (Indice evento) come etichetta della variabile.
- ▶ Fare clic su Avanti.

Figura 22-32
Ristrutturazione di dati guidata, Variabili in casi: Crea una variabile indice

Ristrutturazione di dati guidata - Passo 6 di 7

Variabili in casi: Opzioni

In questo passaggio è possibile impostare le opzioni che verranno applicate al file di dati ristrutturati.

Gestione di variabili non selezionate

Rilascia variabile/i dal nuovo file di dati

Mantieni e tratta come variabile/i fissate

Mancante di sistema o valori vuoti in tutte le variabili trasposte

Crea un caso nel nuovo file

Scarta dati

Variabile conteggio casi

Conta il numero di nuovi casi creato dal caso nei dati correnti

Nome:

Etichetta:

< Indietro Avanti > Fine Annulla Aiuto

- ▶ Verificare che l'opzione Mantieni e tratta come variabile/i sia selezionata.
- ▶ Fare clic su Fine.

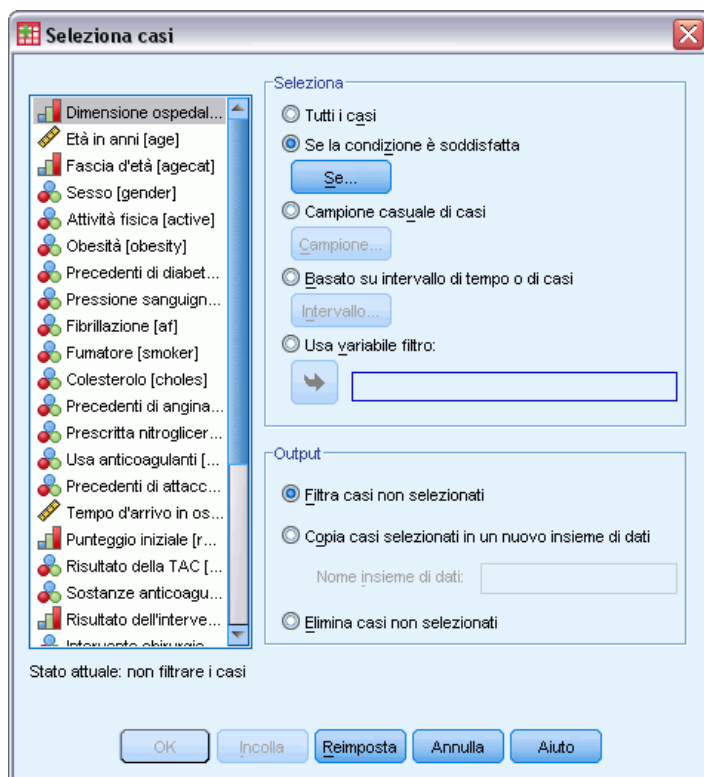
Figura 22-33
Dati ristrutturati

event_index	event	start_time	time_to_event	mi	is	hs
1	0	3	1500	0	1	0
2	-4	1500	-4	-4	-4	-4
3	-4	.	-4	-4	-4	-4
1	1	33	1311	0	1	0
2	4	1311	1325	1	1	0
3	-3	1325	-3	-3	-3	-3
1	4	12	1098	1	1	0
2	-3	1098	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3
1	4	4	1356	0	1	0
2	-3	1356	-3	-3	-3	-3
3	-3	.	-3	-3	-3	-3

I dati ristrutturati contengono tre casi per ogni paziente; tuttavia, molti pazienti sono andati incontro a meno di tre eventi, pertanto sono presenti molti casi con valori negativi (mancanti) per *evento*. È possibile semplicemente filtrare questi casi dall'insieme di dati

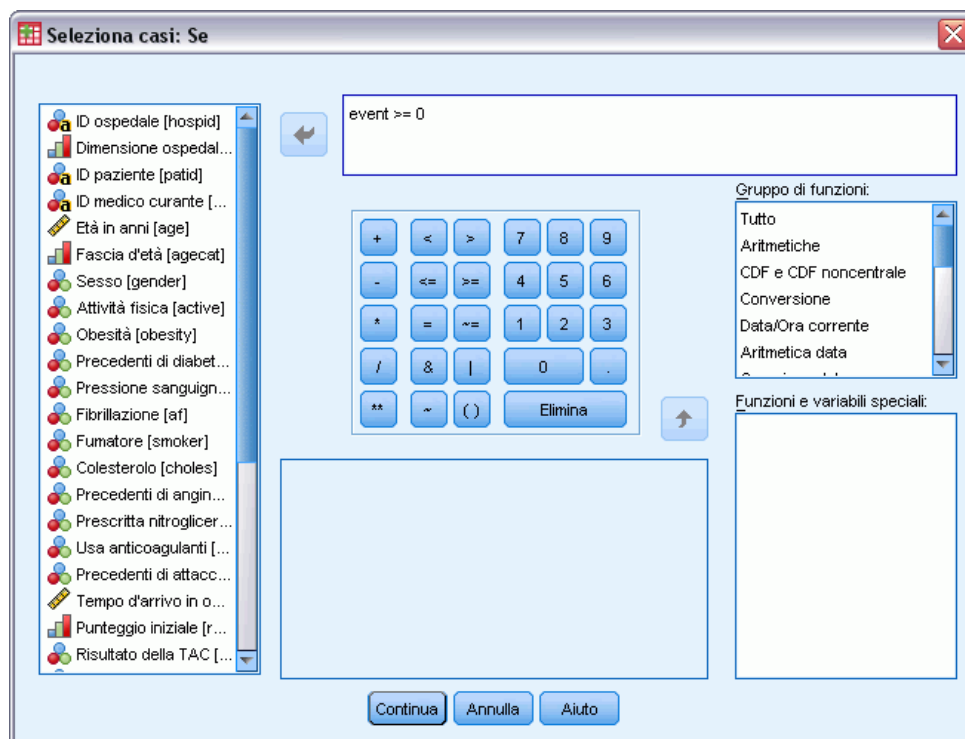
- Per filtrare tali casi, selezionare dai menu:
Dati > Seleziona casi...

Figura 22-34
Finestra di dialogo *Seleziona casi*



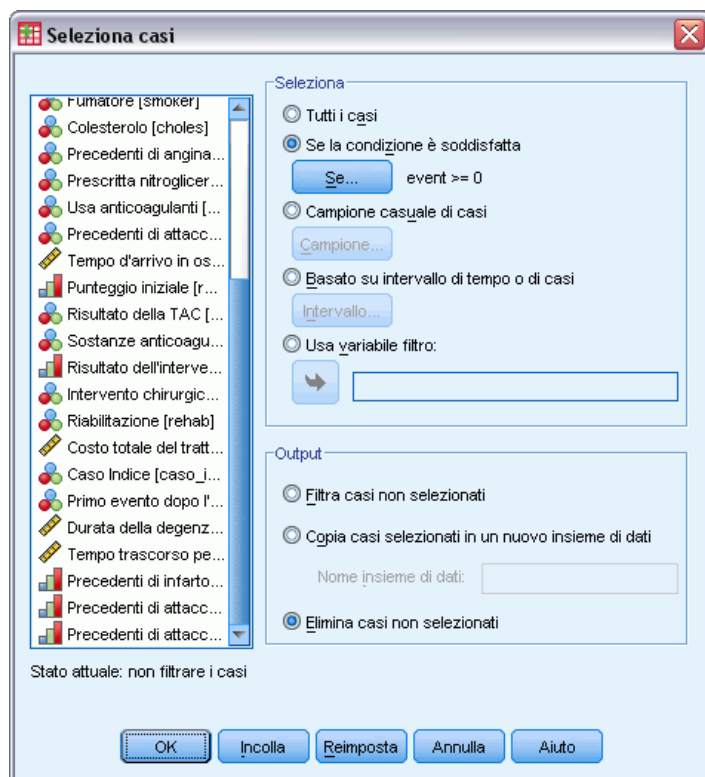
- ▶ Selezionare *Se la condizione è soddisfatta*.
- ▶ Fare clic su *Se*.

Figura 22-35
Finestra di dialogo *Seleziona casi: Se*



- Digitare event >= 0 (evento >= 0) come espressione condizionale.
- Fare clic su Continua.

Figura 22-36
Finestra di dialogo *Seleziona casi*



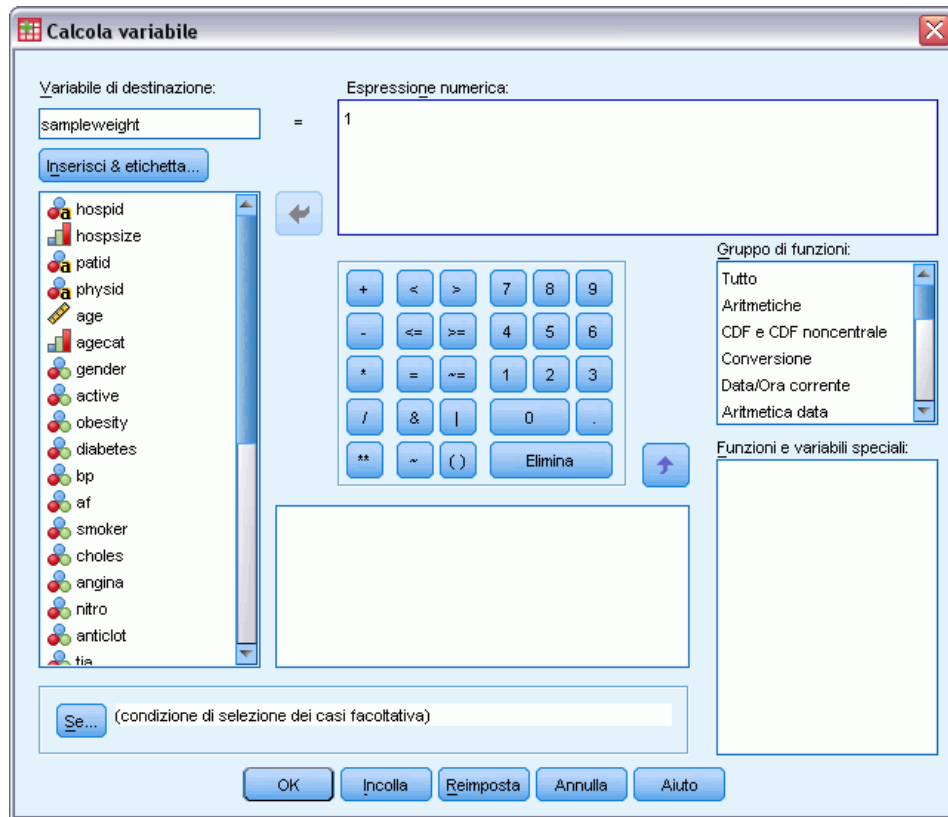
- ▶ Selezionare Elimina casi non selezionati.
- ▶ Fare clic su OK.

Creazione di un piano di analisi di campionamento casuale semplice

È ora possibile creare il piano di analisi di campionamento casuale semplice.

- ▶ Innanzitutto è necessario creare una variabile di ponderazione di campionamento. Dai menu, scegliere:
Trasforma > Calcola variabile...

Figura 22-37
Finestra di dialogo principale Regressione di Cox



- ▶ Digitare sampleweight (pesocampione) come variabile di destinazione.
- ▶ Digitare 1 come espressione numerica.
- ▶ Fare clic su OK.

È ora possibile creare il piano analisi.

Nota: È disponibile un file del piano, *srs.csaplan*, nella directory dei file di esempio, che è possibile utilizzare se si desidera ignorare le istruzioni seguenti e passare direttamente all'analisi dei dati.

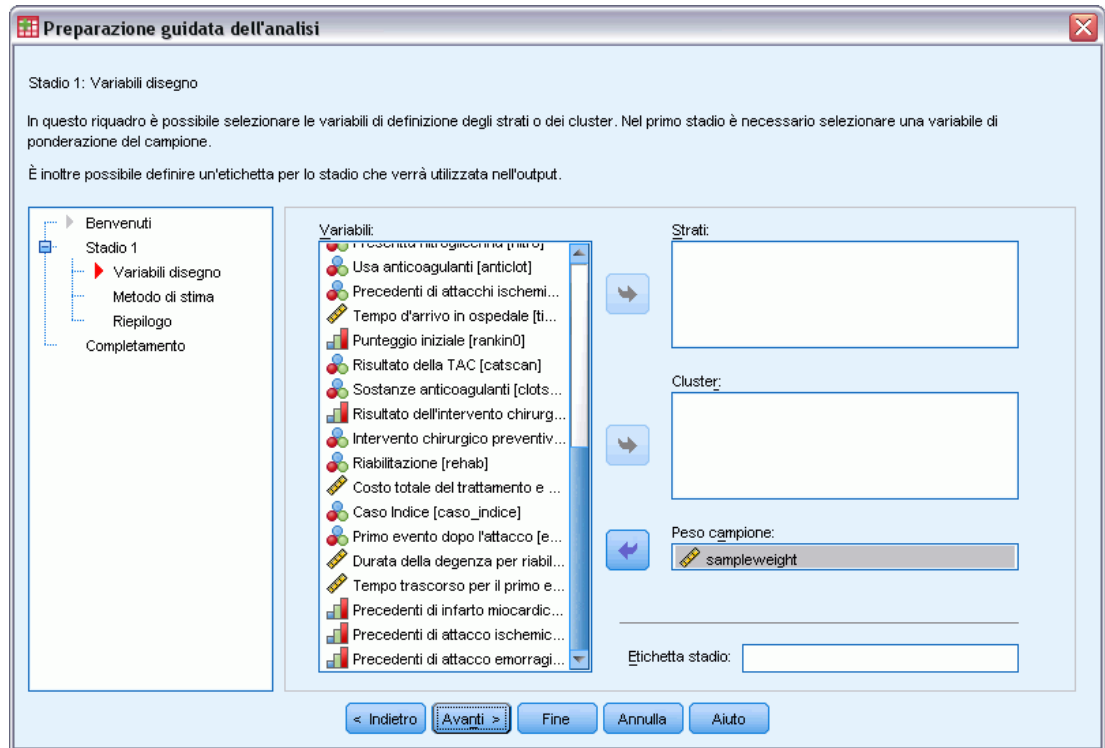
- ▶ Per creare il piano analisi, dai menu scegliere:
Analizza > Complex Samples > Prepara per l'analisi...

Figura 22-38
Preparazione guidata dell'analisi, Finestra di benvenuto



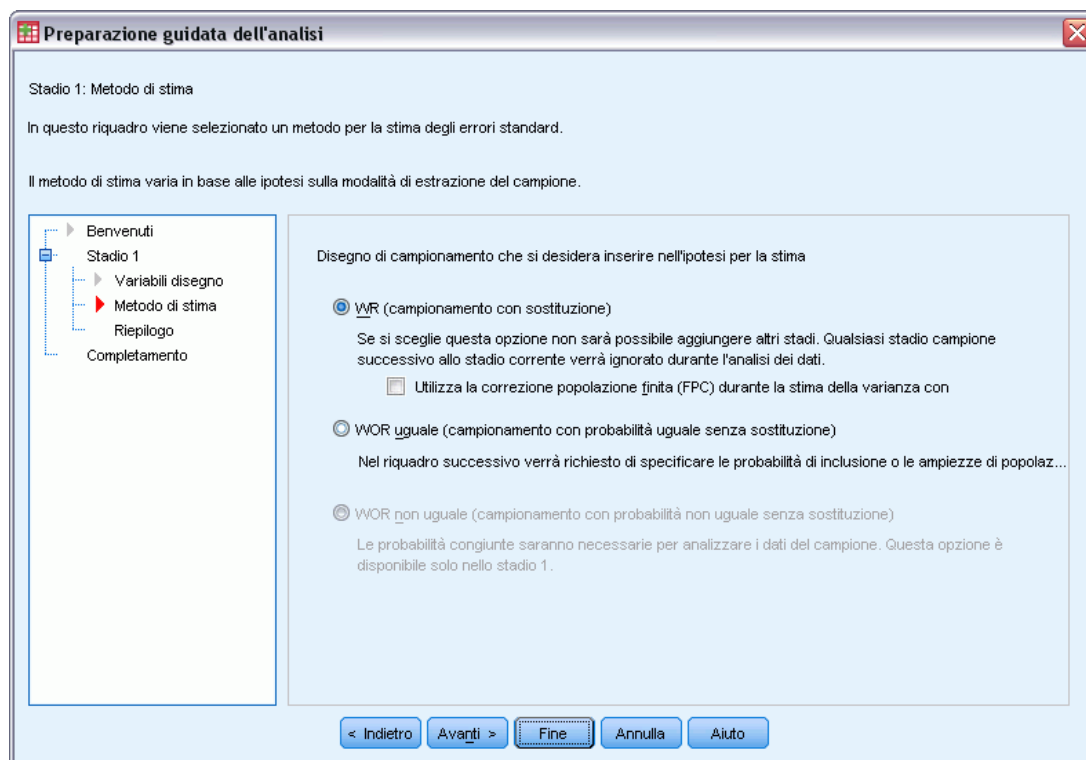
- ▶ Selezionare Crea un file del piano e digitare srs.csaplan come nome del file. In alternativa, accedere alla posizione in cui eseguire il salvataggio.
- ▶ Fare clic su Avanti.

Figura 22-39
Preparazione guidata dell'analisi, Variabili disegno



- ▶ Selezionare *sampleweight* (*pesocampione*) come variabile di ponderazione del campione.
- ▶ Fare clic su Avanti.

Figura 22-40
Preparazione guidata dell'analisi, Metodo di stima



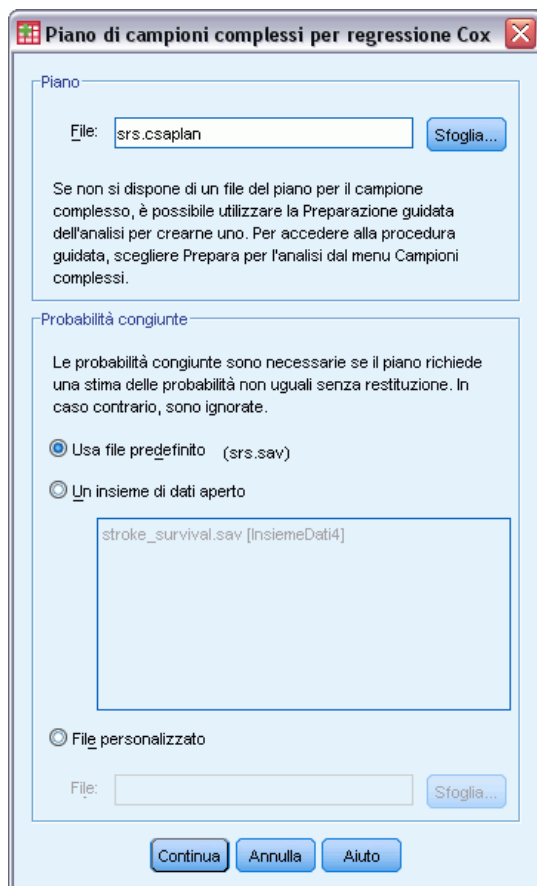
- ▶ Deselezionare Utilizza la correzione popolazione finita.
- ▶ Fare clic su Fine.

È ora possibile eseguire l'analisi.

Esecuzione dell'analisi

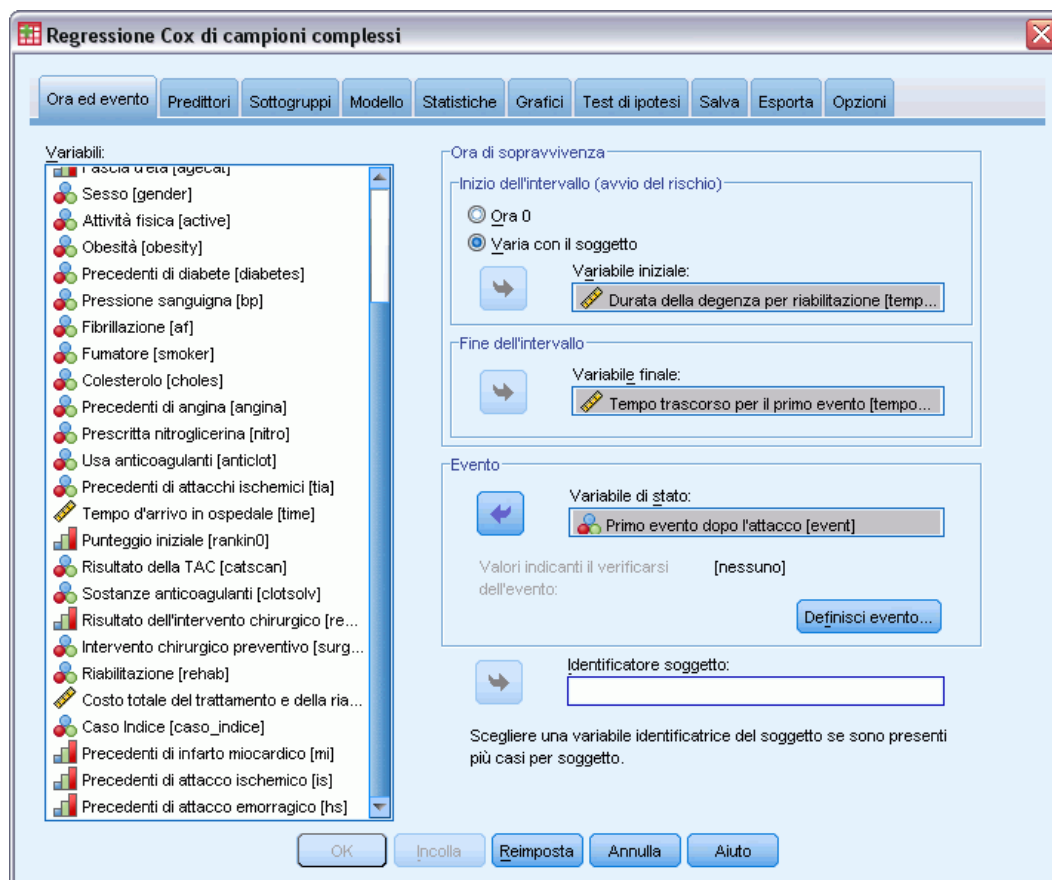
- ▶ Per eseguire un'analisi Regressione Cox di campioni complessi, dai menu scegliere:
Analizza > Complex Samples > Regressione di Cox...

Figura 22-41
Finestra di dialogo Piano di campioni complessi per regressione Cox



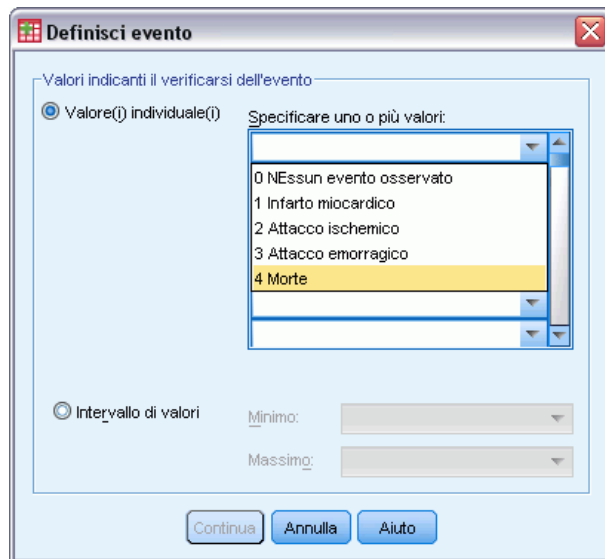
- ▶ Accedere alla posizione nella quale è stato salvato il piano di analisi di campionamento casuale semplice oppure alla directory dei file di esempio e selezionare *srs.csaplan*.
- ▶ Fare clic su Continua.

Figura 22-42
Finestra di dialogo regressione di Cox, scheda Ora ed evento



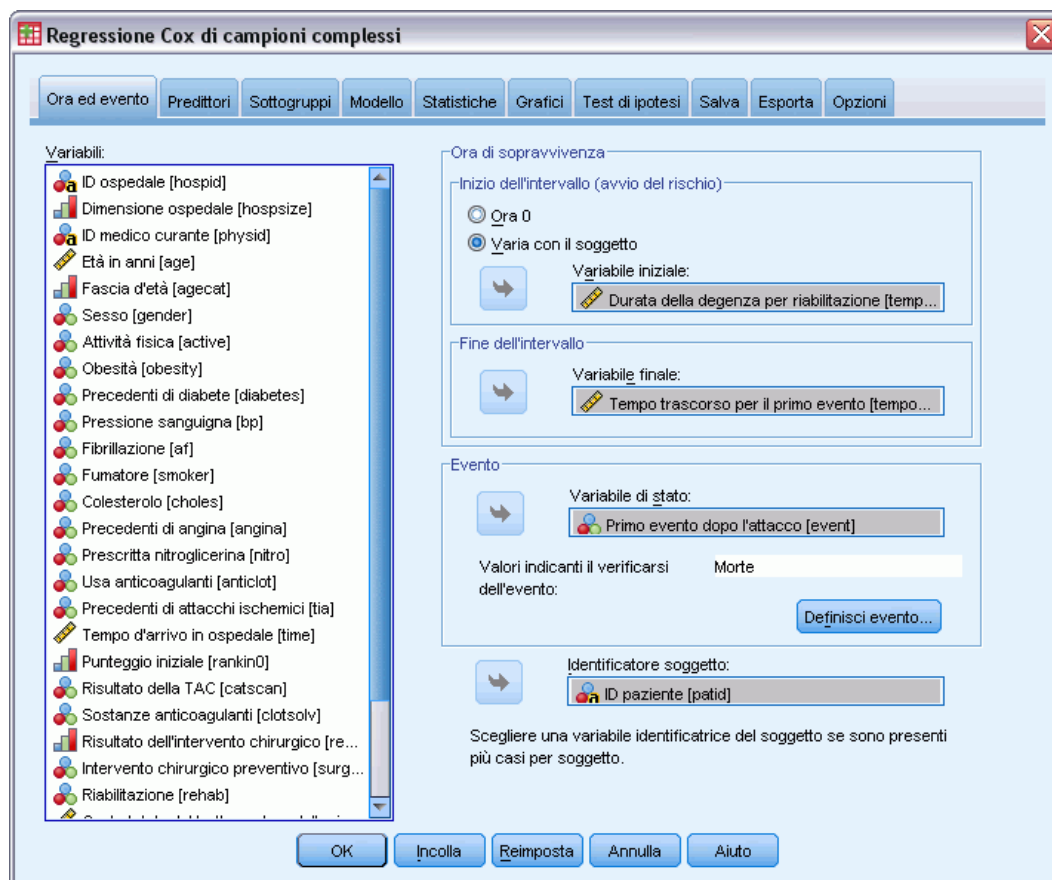
- ▶ Selezionare *Vary with the subject*, quindi *Length of stay for rehabilitation [los_rehab]* (*Durata della riabilitazione [dur_riab]*) come variabile iniziale. Tenere presente che la variabile ristrutturata ha ricavato l'etichetta dalla prima variabile utilizzata per la sua creazione, sebbene l'etichetta non sia necessariamente adatta alla variabile creata.
- ▶ Selezionare *Time to first event post-attack [time_to_event]* (*Primo evento post-attacco [tempo_fino_a_evento]*) come variabile finale.
- ▶ Selezionare *First event post-attack [event]* (*Primo evento post-attacco [evento]*) come variabile di stato.
- ▶ Fare clic su *Definisci evento*.

Figura 22-43
Finestra di dialogo *Definisci evento*



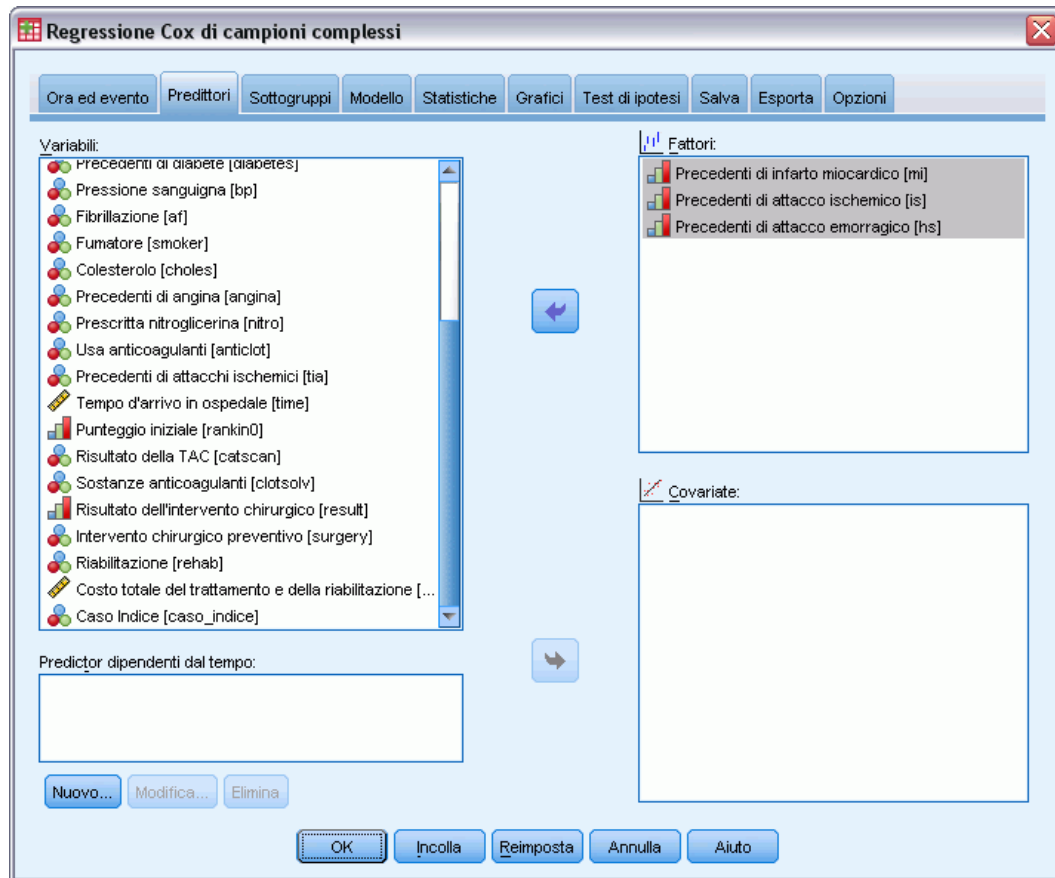
- ▶ Selezionare 4 Death (4 decesso) come valore che indica il verificarsi dell'evento terminale.
- ▶ Fare clic su Continua.

Figura 22-44
Finestra di dialogo regressione di Cox, scheda Ora ed evento



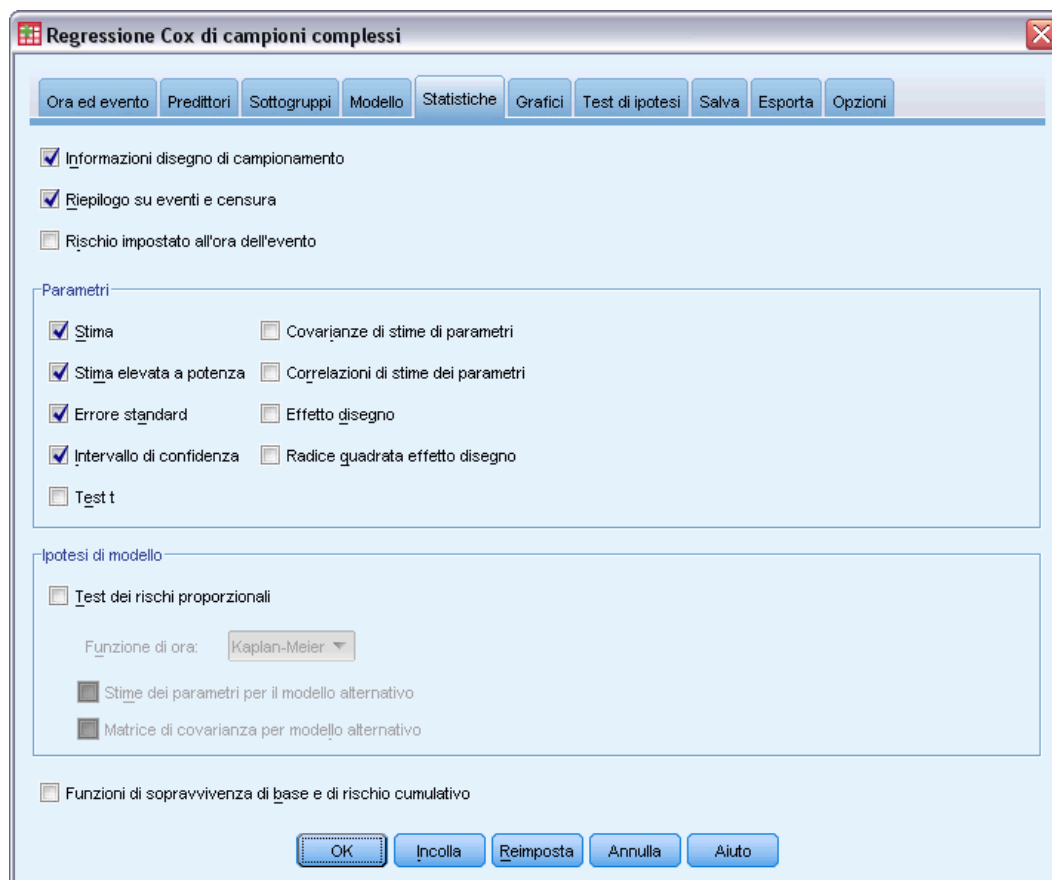
- ▶ Selezionare *Patient ID [patid]* (*ID paziente [idpaz]*) come identificatore del soggetto.
- ▶ Fare clic sulla scheda Predittori.

Figura 22-45
Finestra di dialogo *Regressione di Cox*, scheda *Predittori*



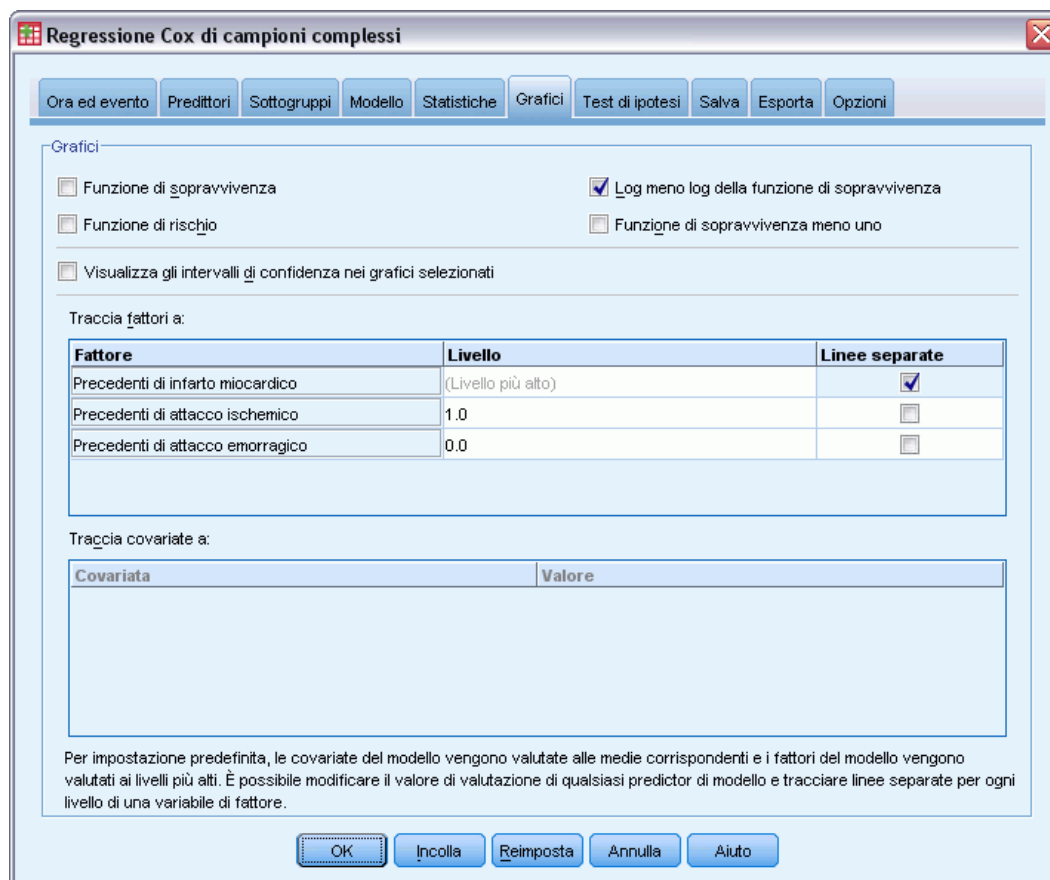
- ▶ Selezionare da *History of myocardial infarction [mi]* (*Anamnesi infarto del miocardio [inf]*) a *History of hemorrhagic stroke [hs]* (*Anamnesi emorragia celebrale [em]*) come fattori.
- ▶ Fare clic sulla scheda *Statistiche*.

Figura 22-46
Finestra di dialogo Regressione di Cox, scheda Statistiche



- ▶ Selezionare Stima, Stima elevata a potenza, Errore standard e Intervallo di confidenza nel gruppo Parametri.
- ▶ Fare clic sulla scheda Grafici.

Figura 22-47
Finestra di dialogo *Regressione di Cox*, scheda *Statistiche*



- ▶ Selezionare Log meno log della funzione di sopravvivenza.
- ▶ Selezionare Linee separate per *History of myocardial infarction* (*Anamnesi infarto del miocardio*).
- ▶ Selezionare 1,0 come livello per *History of ischemic stroke* (*Anamnesi dell'attacco ischemico*).
- ▶ Selezionare 0,0 come livello per *History of ischemic stroke* (*Anamnesi dell'attacco ischemico*).
- ▶ Fare clic sulla scheda Opzioni.

Figura 22-48
Finestra di dialogo Regressione di Cox, scheda Opzioni

Regressione Cox di campioni complessi

Or a ed evento | Predittori | Sottogruppi | Modello | Statistiche | Grafici | Test di ipotesi | Salva | Esporta | Opzioni

Stima

Max iterazioni: 100

Max dimezzamenti: 5

Limita iterazioni in base a variazioni di stima dei parametri

Modifica minima: 0.000001 Tipo: Relativo

Limita iterazioni in base a variazioni di verosimiglianza

Modifica minima: Tipo: Relativo

Visualizza cronologia iterazioni

Incremento: 1

Metodo di interruzione dei casi pari merito per la stima dei parametri:

Efron

Breslow

Intervallo di confidenza (%): 95

Funzioni di sopravvivenza

Metodo per stimare le funzioni di sopravvivenza di base:

Metodo Efron

Metodo Breslow

Metodo prodotto-limite

Intervalli di confidenza delle funzioni di sopravvivenza:

Calcola sulla base della funzione di sopravvivenza trasformata, quindi ritrasforma nelle unità originali

Trasformazione: Log

Calcola sulla base delle unità originali della funzione di sopravvivenza

Valori mancanti definiti dall'utente

Considera non valido

Considera valido

Questa impostazione si applica a tutte le variabili del modello categoriale e del disegno campione.

OK | Incolla | Reimposta | Annulla | Aiuto

- Selezionare Breslow come metodo di interruzione dei casi pari merito nel gruppo Stima.
- Fare clic su OK.

Informazioni disegno di campionamento

Figura 22-49
Informazioni disegno di campionamento

			N
Conteggi non ponderati	Valid	Soggetti	2421
		Casi	3310
	Casi non validi		0
	Casi totali		3310
Dimensione soggetti popolazione			2421,000
Stadio 1	Strati		1
	Unità		2421
Gradi di libertà del disegno campionario			2420

Questa tabella contiene le informazioni sul disegno di campionamento relative alla stima del modello.

- Sono presenti più casi per soggetto e tutti i 3.310 casi vengono utilizzati nell'analisi.
- Il disegno ha un singolo strato e 2.421 unità (una per ogni soggetto). I gradi di libertà del disegno di campionamento sono stimati in $2421 - 1 = 2420$.

Test degli effetti del modello

Figura 22-50
Test degli effetti del modello

Sorgente	df1	df2	F di Wald	Correzione per confronti multipli
mi	3.000	2418.000	452.873	.000
is	2.000	2419.000	1064.936	.000
hs	2.000	2419.000	739.197	.000

Variabile ora di sopravvivenza: Durata della degenza per riabilitazione,
Tempo trascorso per il primo evento
Variabile di stato evento: Primo evento dopo l'attacco = 4
Variabile ID soggetto: ID paziente
Modello: mi, is, hs

Il valore di significatività per ogni effetto è prossimo a 0 e questo suggerisce che tutti gli effetti contribuiscono al modello.

Stime di parametri

Figura 22-51
Stime dei parametri

Parametro	B	Errore standard	Intervallo di confidenza 95%		Exp(B)	Intervallo di confidenza 95% per Exp(B)	
			Inferiore	Superiore		Inferiore	Superiore
[mi=0]	-6.381	.283	-6.935	-5.827	.002	.001	.003
[mi=1]	-5.589	.284	-6.147	-5.032	.004	.002	.007
[mi=2]	-2.119	.344	-2.794	-1.445	.120	.061	.236
[mi=3]	.000 ^a	.	.	.	1.000	.	.
[is=1]	-6.421	.202	-6.817	-6.024	.002	.001	.002
[is=2]	-2.803	.222	-3.239	-2.366	.061	.039	.094
[is=3]	.000 ^a	.	.	.	1.000	.	.
[hs=0]	-6.148	.355	-6.844	-5.453	.002	.001	.004
[hs=1]	-2.232	.373	-2.963	-1.502	.107	.052	.223
[hs=2]	.000 ^a	.	.	.	1.000	.	.

Variabile ora di sopravvivenza: Durata della degenza per riabilitazione,
Tempo trascorso per il primo evento
Variabile di stato evento: Primo evento dopo l'attacco = 4
Variabile ID soggetto: ID paziente
Modello: mi, is, hs

a. Impostare su zero perché questo parametro è duplicato.

b. Metodo di interruzione dei casi pari merito: Breslow

La procedura utilizza l'ultima categoria di ogni fattore come categoria di riferimento; l'effetto delle altre categorie è relativo alla categoria di riferimento. Tenere presente che mentre la stima è utile per la verifica statistica, la stima elevata a potenza, $Exp(B)$, viene più facilmente interpretata come modifica prevista nel rischio relativamente alla categoria di riferimento.

- Il valore di $\text{Exp}(B)$ per $[mi=0]$ ($[inf=0]$) indica che il rischio di morte di un paziente che non ha avuto precedenti infarti del miocardio (inf) è 0,002 volte inferiore a quello di un paziente che ha avuto tre precedenti di infarto.
- Gli intervalli di $[mi=1]$ ($[inf=1]$) e $[mi=0]$ ($[inf=0]$) si sovrappongono e questo indica che il rischio per un paziente che ha avuto un solo precedente di infarto non è statisticamente distinguibile da quello di un paziente che invece non ha avuto alcun infarto precedente.
- Gli intervalli di confidenza di $[mi=0]$ ($[inf=0]$) e $[mi=1]$ ($[inf=1]$) non si sovrappongono con l'intervallo di $[mi=2]$ ($[inf=2]$) e nessuno di essi include il valore 0. Quindi, sembra che il rischio per i pazienti che non hanno avuto nessun precedente o un solo precedente di infarto del miocardio sia distinguibile dal rischio per i pazienti con due precedenti di infarto, che a sua volta è distinguibile dal rischio per i pazienti con tre precedenti di infarto.

Relazioni simili sono valide per i livelli di is e em , dove l'aumento del numero di episodi precedenti aumenta il rischio di morte.

Valori modello

Figura 22-52
Valori modello

		Intervallo ora di sopravvivenza				
		Inizio	Fine	Precedenti di infarto miocardico	Precedenti di attacchi ischemici	Precedenti di attacco emorragico
Modello di riferimento	1	.000		Tre	Tre	Due
Modello 1.1	1	.000		Nessuno	Uno	Nessuno
Modello 1.2	1	.000		Uno	Uno	Nessuno
Modello 1.3	1	.000		Due	Uno	Nessuno
Modello 1.4	1	.000		Tre	Uno	Nessuno

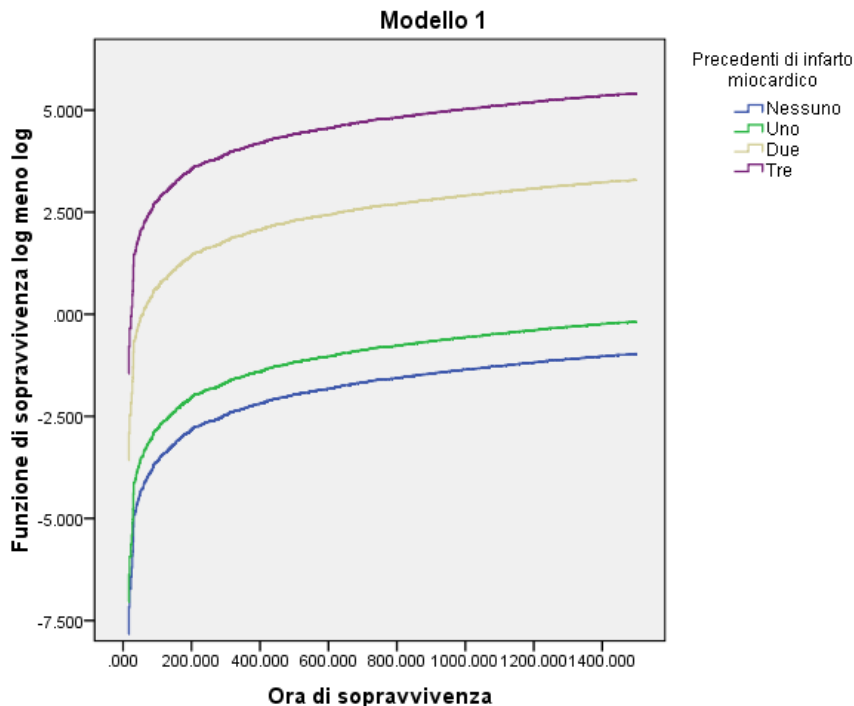
A un predictor non specificato è stato assegnato il valore di questo predictor nel modello di riferimento.
Ogni intervallo di ora di sopravvivenza è definito come Inizio < Ora di sopravvivenza <= Fine. Modello : mi, is, hs.

La tabella dei valori modello riporta i valori che definiscono ogni modello predittore. Oltre ai predittori nel modello, vengono mostrati i periodi iniziali e finali dell'intervallo di sopravvivenza. Per le analisi eseguite dalle finestre di dialogo, i periodi iniziali e finali saranno sempre rispettivamente 0 e illimitati; tramite la sintassi è possibile specificare i percorsi dei predittori costanti.

- Il modello di riferimento viene impostato sulla categoria di riferimento per ogni fattore e sul valore medio di ogni covariata (non vi sono covariate in questo modello). Per questo insieme di dati, la combinazione dei fattori mostrata per il modello di riferimento non può verificarsi, pertanto il grafico log meno log del modello di riferimento verrà ignorato.
- I modelli da 1,1 a 1,4 differiscono solo per il valore di *History of myocardial infarction* (*Anamnesi infarto del miocardio*). Viene creato un modello separato (e una riga separata nel grafico) per ogni valore di *History of myocardial infarction* (*Anamnesi infarto del miocardio*) mentre le altre variabili rimangono costanti.

Grafico log meno log

Figura 22-53
Grafico log meno log



Questo grafico mostra il log meno log della funzione di sopravvivenza, $\ln(-\ln(\text{sopravvivenza}))$, rispetto al tempo di sopravvivenza. Questo particolare grafico mostra una curva separata per ogni categoria di *History of myocardial infarction* (*Anamnesi infarto del miocardio*), con *History of ischemic stroke* (*Anamnesi dell'attacco ischemico*) il cui valore è 1 e *History of hemorrhagic stroke* (*Anamnesi emorragia cerebrale*) il cui valore è None. È inoltre un'utile rappresentazione dell'effetto di *History of myocardial infarction* (*Anamnesi infarto del miocardio*) sulla funzione di sopravvivenza. Come visto nella tabella delle stime dei parametri, sembra che la sopravvivenza dei pazienti che non hanno avuto precedenti di infarto del miocardio o con un solo precedente sia distinguibile rispetto alla sopravvivenza dei pazienti che hanno avuto due precedenti di infarto del miocardio, che a sua volta è distinguibile dalla sopravvivenza dei pazienti che hanno avuto tre precedenti di infarto.

Riepilogo

È stato stabilito un modello di Regressione di Cox relativamente alla sopravvivenza da infarto del miocardio che stima gli effetti del cambiamento dell'anamnesi nei pazienti che hanno subito infarto del miocardio. E questo è solo l'inizio, i ricercatori vorranno sicuramente includere nel modello altri predittori potenziali. Inoltre, in successive analisi dell'insieme di dati sarà possibile prendere in considerazione modifiche più significative alla struttura del modello. Ad esempio, il modello presume che l'effetto dell'evento che altera l'anamnesi del paziente possa essere

quantificato mediante un moltiplicatore del rischio di base. Invece, potrebbe essere ragionevole supporre che la forma del rischio di base venga alterata dal verificarsi di un evento non mortale. A tale scopo, è possibile stratificare l'analisi in base a *Event index (Indice evento)*.

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.
- **afatoxin.sav.** File di dati ipotetici che prende in esame il test di raccolti di mais con presenza di Aflatossina, un veleno la cui concentrazione varia notevolmente nei raccolti. Una macchina per la lavorazione dei cereali ha ricevuto 16 campioni da ciascuno degli otto raccolti di mais e ha misurato i livelli di Aflatossina in parti per miliardo (PPB).
- **anorectic.sav.** Per trovare una sintomatologia standardizzata del comportamento anoressico/bulimico, i ricercatori (Van der Ham, Meulman, Van Strien, e Van Engeland, 1997) hanno condotto uno studio basato su 55 adolescenti affetti da disordini alimentari conosciuti. Ogni paziente è stato visitato quattro volte in quattro anni, per un totale di 220 visite. Durante ogni visita, ai pazienti sono stati assegnati punteggi per ciascuno dei 16 sintomi. I punteggi relativi ai sintomi sono assenti per il paziente 71 alla visita 2, il paziente 76 alla visita 2 e il paziente 47 alla visita 3, con 217 osservazioni valide.
- **bankloan.sav.** File di dati ipotetici che prende in esame l'impegno di una banca nel tentativo di ridurre il tasso di inadempienza nel rimborso di un prestito. Il file contiene informazioni finanziarie e demografiche su 850 vecchi e potenziali clienti. I primi 700 casi riguardano i

clienti a cui sono stati concessi dei prestiti precedentemente. Gli ultimi 150 casi riguardano i potenziali clienti che la banca deve classificare come rischi di credito positivi o negativi.

- **bankloan_binning.sav.** File di dati ipotetici che contiene informazioni finanziarie e demografiche su 5000 vecchi clienti.
- **behavior.sav.** In un classico esempio (Prezzo e Bouffard, 1974), 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 (Green e Rao, 1972), 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 (McCullagh e Nelder, 1989) 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 (Green e Wind, 1973), 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 (Kennedy, Riquier, e Sharp, 1996). 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*.
- **debate.sav.** File di dati ipotetici che prende in esame le risposte appaite 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" (Rickman, Mitchell, Dingman, e Dalen, 1974). 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 (Blake e Merz, 1998) 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 (Bell, 1961) ha presentato una tabella per illustrare i possibili gruppi sociali. Guttman (Guttman, 1968) 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 (Rosenberg e Kim, 1975) 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 15×15 , 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 *sexo*, *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., durate 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/. 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, (Breiman e Friedman, 1985) e (Hastie e Tibshirani, 1990), 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.

- **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 (Hartigan, 1975).
- **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 (McCullagh et al., 1989) 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.
- **stocks.sav** Questo file di dati ipotetici contiene i prezzi e i volumi delle scorte riferiti a un anno.
- **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 (Collett, 2003).

- **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 (Collett et al., 2003).
- **verd1985.sav.** Questo file di dati prende in esame un'indagine (Verdegaal, 1985). 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 (Ware, Dockery, Spiro III, Speizer, e Ferris Jr., 1984). I dati contengono misure binarie ripetute del livello di asma dei bambini della città di Steubenville, Ohio, di 7, 8, 9 e 10 anni. I dati indicano anche se la mamma dei bambini era fumatrice durante il primo anno dello studio.
- **workprog.sav.** File di dati ipotetici che prende in esame un programma di lavoro governativo il cui obiettivo è fornire attività più adatte alle persone diversamente abili. È stato seguito un campione di potenziali partecipanti al programma, alcuni dei quali sono stati selezionati in modo casuale e altri no. Ogni caso rappresenta un diverso partecipante al programma.
- **worldsales.sav** Questo file di dati ipotetici contiene i ricavi suddivisi per continenti e prodotti.

Note legali

Queste informazioni sono state preparate per prodotti e servizi offerti in tutto il mondo.

IBM potrebbe non offrire i prodotti, i servizi o le funzionalità di cui si tratta nel presente documento in altri paesi. Contattare il rappresentante IBM locale per informazioni sui prodotti e i servizi attualmente disponibili nella propria zona. Qualsiasi riferimento a un prodotto, programma o servizio IBM non intende dichiarare o implicare che sia possibile utilizzare esclusivamente tale prodotto, programma o servizio IBM. Potrà invece essere utilizzato qualsiasi prodotto, programma o servizio con funzionalità equivalente e che non violi i diritti di proprietà intellettuale di IBM. Tuttavia, è responsabilità dell'utente valutare e verificare il funzionamento di qualsiasi prodotto, programma o servizio non IBM.

IBM può essere titolare di brevetti o domande di brevetto relativi alla materia oggetto del presente documento. La consegna del presente documento non conferisce alcuna licenza rispetto a questi brevetti. Rivolgere per iscritto i quesiti sulle licenze a:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

Per richieste di informazioni sulle licenze riguardanti il set di caratteri a byte doppio (DBCS), contattare l'Intellectual Property Department di IBM del proprio paese, oppure inviare le richieste in forma scritta all'indirizzo:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Giappone.

Il seguente paragrafo non si applica per il Regno Unito o altri paesi in cui le presenti disposizioni non sono conformi alle leggi locali: INTERNATIONAL BUSINESS MACHINES FORNISCE QUESTA PUBBLICAZIONE “COSÌ COM'È” SENZA GARANZIA DI ALCUN TIPO, SIA ESSA ESPRESSA O IMPLICITA, INCLUSE, MA NON LIMITATE A, LE GARANZIE IMPLICITE DI NON VIOLAZIONE, COMMERCIALIZZABILITÀ O IDONEITÀ A UNO SCOPO SPECIFICO. Alcuni stati non consentono limitazioni di garanzie espresse o implicite in determinate transazioni, pertanto quanto sopra potrebbe non essere applicabile.

Le presenti informazioni possono includere imprecisioni tecniche o errori tipografici. Le modifiche periodiche apportate alle informazioni contenute in questa pubblicazione verranno inserite nelle nuove edizioni della pubblicazione. IBM può apportare miglioramenti e/o modifiche al/ai prodotto/i e/o al/ai programma/i descritti nella presente pubblicazione in qualsiasi momento senza preavviso.

Qualsiasi riferimento nelle presenti informazioni a siti Web non IBM viene fornito esclusivamente per facilitare la consultazione e non rappresenta in alcun modo un'approvazione o sostegno da parte nostra di tali siti Web. I materiali contenuti in tali siti Web non fanno parte dei materiali di questo prodotto IBM e il loro utilizzo è esclusivamente a rischio dell'utente.

IBM può utilizzare o distribuire eventuali informazioni fornite dall'utente nei modi che ritiene appropriati senza incorrere in alcun obbligo nei confronti dell'utente.

I licenziatari del programma che desiderassero informazioni su di esso allo scopo di abilitare: (i) lo scambio di informazioni tra programmi creati indipendentemente e altri programmi (questo compreso) e (ii) l'utilizzo in comune delle informazioni scambiate, dovranno rivolgersi a:

IBM Software Group, All'attenzione di: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Tali informazioni saranno fornite in conformità ai termini e alle condizioni in vigore e, in alcuni casi, dietro pagamento.

Il programma concesso in licenza descritto nel presente documento e tutto il materiale correlato disponibile sono forniti da IBM in base ai termini del contratto di licenza cliente IBM, del contratto di licenza internazionale IBM o del contratto equivalente esistente tra le parti.

le informazioni relative a prodotti non IBM sono state ottenute dai fornitori di tali prodotti, da loro annunci pubblicati e da altre fonti disponibili al pubblico. IBM non ha verificato tali prodotti e non può confermare l'accuratezza delle prestazioni, la compatibilità o qualsiasi altra dichiarazione relativa a prodotti non IBM. Eventuali domande in merito alle funzionalità dei prodotti non IBM vanno indirizzate ai fornitori di tali prodotti.

Le presenti informazioni includono esempi di dati e report utilizzati in operazioni aziendali quotidiane. Per fornire una descrizione il più possibile esaustiva, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e ogni somiglianza a nomi e indirizzi utilizzati da aziende reali è puramente casuale.

Per chi visualizza queste informazioni a video: le fotografie e le illustrazioni a colori potrebbero non essere disponibili.

Marchi commerciali

IBM, il logo IBM, ibm.com e SPSS sono marchi di IBM Corporation, registrati in numerose giurisdizioni nel mondo. Un elenco aggiornato dei marchi IBM è disponibile sul Web all'indirizzo <http://www.ibm.com/legal/copytrade.shtml>.

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

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

Java e tutti i marchi e i logo basati su Java sono marchi di Sun Microsystems, Inc. negli Stati Uniti e/o negli altri paesi.

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

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

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Questo prodotto utilizza WinWrap Basic, Copyright 1993-2007, Polar Engineering and Consulting, <http://www.winwrap.com>.

Altri nomi di prodotti e servizi possono essere marchi commerciali di IBM o di altre aziende.

Le schermate dei prodotti Adobe sono state ristampate su autorizzazione di Adobe Systems Incorporated.

Le schermate dei prodotti Microsoft sono state ristampate su autorizzazione di Microsoft Corporation.



Bibliografia

- Bell, E. H. 1961. *Social foundations of human behavior: Introduction to the study of sociology*. New York: Harper & Row.
- Blake, C. L., e C. J. Merz. 1998. "UCI Repository of machine learning databases." Available at <http://www.ics.uci.edu/~mlearn/MLRepository.html>.
- Breiman, L., e J. H. Friedman. 1985. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80, .
- Cochran, W. G. 1977. *Sampling Techniques*, 3rd ed. New York: John Wiley and Sons.
- Collett, D. 2003. *Modelling survival data in medical research*, 2 ed. Boca Raton: Chapman & Hall/CRC.
- Cox, D. R., e E. J. Snell. 1989. *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.
- Green, P. E., e V. Rao. 1972. *Applied multidimensional scaling*. Hinsdale, Ill.: Dryden Press.
- Green, P. E., e Y. Wind. 1973. *Multiattribute decisions in marketing: A measurement approach*. Hinsdale, Ill.: Dryden Press.
- Guttman, L. 1968. A general nonmetric technique for finding the smallest coordinate space for configurations of points. *Psychometrika*, 33, .
- Hartigan, J. A. 1975. *Clustering algorithms*. New York: John Wiley and Sons.
- Hastie, T., e R. Tibshirani. 1990. *Generalized additive models*. London: Chapman and Hall.
- Kennedy, R., C. Riquier, e B. Sharp. 1996. Practical applications of correspondence analysis to categorical data in market research. *Journal of Targeting, Measurement and Analysis for Marketing*, 5, .
- Kish, L. 1965. *Survey Sampling*. New York: John Wiley and Sons.
- Kish, L. 1987. *Statistical Design for Research*. New York: John Wiley and Sons.
- McCullagh, P., e J. A. Nelder. 1989. *Generalized Linear Models*, 2nd ed. London: Chapman & Hall.
- McFadden, D. 1974. Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Economics*, P. Zarembka, ed. New York: Academic Press.
- Murthy, M. N. 1967. *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.
- Nagelkerke, N. J. D. 1991. A note on the general definition of the coefficient of determination. *Biometrika*, 78:3, .
- Prezzo, R. H., e D. L. Bouffard. 1974. Behavioral appropriateness and situational constraints as dimensions of social behavior. *Journal of Personality and Social Psychology*, 30, .
- Rickman, R., N. Mitchell, J. Dingman, e J. E. Dalen. 1974. Changes in serum cholesterol during the Stillman Diet. *Journal of the American Medical Association*, 228, .
- Rosenberg, S., e M. P. Kim. 1975. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 10, .
- Särndal, C., B. Swensson, e J. Wretman. 1992. *Model Assisted Survey Sampling*. New York: Springer-Verlag.

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

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

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

- ampiezza popolazione
 - in Campionamento guidato, 12
 - in descrittive campioni complessi, 34
 - in frequenze campioni complessi, 30, 163
 - in rapporti campioni complessi, 43
 - in tavole di contingenza campioni complessi, 39
- avvisi
 - in Regressione ordinale campioni complessi, 218
- Bonferroni
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- Bonferroni sequenziale corretta
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- campionamento
 - disegno complesso, 4
 - campionamento casuale semplice
 - in Campionamento guidato, 8
 - campionamento complesso
 - piano analisi, 19
 - piano di campionamento, 4
- Campionamento guidato di Campioni complessi, 98
 - campionamento PPS, 128
 - procedure correlate, 144
 - struttura di campionamento, completa, 98
 - struttura di campionamento, parziale, 110
 - tabella riassuntiva, 108, 140
- campionamento PPS
 - in Campionamento guidato, 8
- campionamento sequenziale
 - in Campionamento guidato, 8
- campionamento sistematico
 - in Campionamento guidato, 8
- Campioni complessi
 - opzioni, 32, 36, 41, 44
 - test di ipotesi, 50, 61, 72
 - valori mancanti, 31, 40
- categoria di riferimento
 - in Modello lineare generalizzato campioni complessi, 52
 - in Regressione logistica campioni complessi, 57
- categorie previste
 - in Regressione logistica campioni complessi, 63
 - in Regressione ordinale campioni complessi, 74
- chi-quadrato
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- chi-quadrato corretto
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- coefficiente di rischio
 - in tavole di contingenza campioni complessi, 39, 170, 174-175
- coefficiente di variazione (COV)
 - in descrittive campioni complessi, 34
 - in frequenze campioni complessi, 30
 - in rapporti campioni complessi, 43
 - in tavole di contingenza campioni complessi, 39
- conteggio non pesato
 - in descrittive campioni complessi, 34
 - in frequenze campioni complessi, 30
 - in rapporti campioni complessi, 43
 - in tavole di contingenza campioni complessi, 39
- contrasti
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti di deviazione
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti di Helmert
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti differenza
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti polinomiali
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti ripetuti
 - in Modello lineare generalizzato campioni complessi, 52
- contrasti semplici
 - in Modello lineare generalizzato campioni complessi, 52
- convergenza parametri
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- convergenza verosimiglianza
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- correlazioni di stime di parametri
 - in Modello lineare generalizzato campioni complessi, 49
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70
- covarianze di stime di parametri
 - in Modello lineare generalizzato campioni complessi, 49
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70
- cronologia iterazioni
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- dati pubblici
 - in descrittive campioni complessi, 165
 - in Preparazione guidata dell'analisi, 145
- Descrittive campioni complessi, 33, 165
 - dati pubblici, 165
 - procedure correlate, 169
 - statistiche, 34, 168
 - statistiche per sottopopolazione, 168
 - valori mancanti, 35
- differenza di rischio
 - in tavole di contingenza campioni complessi, 39
- differenza meno significativa (LSD)
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89

- dimensione campione
 - in Campionamento guidato, 10, 12
- dimezzamenti
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- effetto disegno
 - in descrittive campioni complessi, 34
 - in frequenze campioni complessi, 30
 - in Modello lineare generalizzato campioni complessi, 49
 - in rapporti campioni complessi, 43
 - in Regressione Cox di campioni complessi, 86
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70
 - in tavole di contingenza campioni complessi, 39
- errore standard
 - in descrittive campioni complessi, 34, 168
 - in frequenze campioni complessi, 30, 163
 - in Modello lineare generalizzato campioni complessi, 49
 - in rapporti campioni complessi, 43
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70
 - in tavole di contingenza campioni complessi, 39
- file del piano, 2
- file di esempio
 - posizione, 270
- Frequenze campioni complessi, 29, 160
 - procedure correlate, 164
 - statistiche, 30
 - tabella di frequenza, 163
 - tabella di frequenza per sottopopolazione, 163
- gradi di libertà
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- grafico log meno log
 - in Regressione Cox di campioni complessi, 268
- informazioni disegno di campionamento
 - in Regressione Cox di campioni complessi, 86, 232, 265
- input pesi campione
 - in Campionamento guidato, 6
- intervalli di confidenza
 - in descrittive campioni complessi, 34, 168
 - in frequenze campioni complessi, 30, 163
 - in Modello lineare generalizzato campioni complessi, 49, 54
 - in rapporti campioni complessi, 43
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70
 - in tavole di contingenza campioni complessi, 39
- iterazioni
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- livello di confidenza
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- marchi commerciali, 281
- media
 - in descrittive campioni complessi, 34, 168
- medie marginali
 - in GLM univariato, 190
- medie marginali stimate
 - in Modello lineare generalizzato campioni complessi, 52
- metodo di campionamento
 - in Campionamento guidato, 8
- metodo di campionamento di Brewer
 - in Campionamento guidato, 8
- metodo di campionamento di Murthy
 - in Campionamento guidato, 8
- metodo di campionamento di Sampford
 - in Campionamento guidato, 8
- Metodo di stima Breslow
 - in Regressione Cox di campioni complessi, 94
- Metodo di stima Efron
 - in Regressione Cox di campioni complessi, 94
- metodo Newton-Raphson
 - in Regressione ordinale campioni complessi, 75
- misura di dimensione
 - in Campionamento guidato, 8
- modelli predittore
 - in Regressione Cox di campioni complessi, 267
- modello cumulativo generalizzato
 - in Regressione ordinale campioni complessi, 215
- Modello lineare generalizzato campioni complessi, 46, 183
 - funzioni aggiuntive del comando, 55
 - medie marginali, 190
 - medie stimate, 52
 - modello, 48
 - opzioni, 54
 - procedure correlate, 192
 - riepilogo del modello, 188
 - salvataggio di variabili, 53
 - statistiche, 49
 - stime dei parametri, 189
 - test degli effetti del modello, 188
- note legali, 280
- percentuali di colonna
 - in tavole di contingenza campioni complessi, 39
- percentuali di riga
 - in tavole di contingenza campioni complessi, 39
- percentuali di tabella
 - in frequenze campioni complessi, 30, 163
 - in tavole di contingenza campioni complessi, 39
- pesi di campione
 - in Campionamento guidato, 12
 - in Preparazione guidata dell'analisi, 20

- piano analisi, 19
 piano di campionamento, 4
 predittore dipendente dal tempo
 in Regressione Cox di campioni complessi, 82, 221
 predittori costanti, dipendenti dal tempo
 in Regressione Cox di campioni complessi, 237
 Preparazione guidata dell'analisi Campioni complessi, 145
 dati pubblici, 145
 pesi di campionamento non disponibili, 148
 procedure correlate, 159
 tabella riassuntiva, 148, 158
 probabilità cumulate
 in Regressione ordinale campioni complessi, 74
 probabilità di inclusione
 in Campionamento guidato, 12
 probabilità di risposta
 in Regressione ordinale campioni complessi, 68
 probabilità prevista
 in Regressione logistica campioni complessi, 63
 in Regressione ordinale campioni complessi, 74
 proporzione campione
 in Campionamento guidato, 12
 Punteggio Fisher
 in Regressione ordinale campioni complessi, 75
- radice quadrata dell'effetto disegno
 in descrittive campioni complessi, 34
 in frequenze campioni complessi, 30
 in Modello lineare generalizzato campioni complessi, 49
 in rapporti campioni complessi, 43
 in Regressione Cox di campioni complessi, 86
 in Regressione logistica campioni complessi, 60
 in Regressione ordinale campioni complessi, 70
 in tavole di contingenza campioni complessi, 39
 raggruppamenti
 in Campionamento guidato, 6
 in Preparazione guidata dell'analisi, 20
 rapporti
 in rapporti campioni complessi, 180
 Rapporti campioni complessi, 42, 177
 procedure correlate, 181
 rapporti, 180
 statistiche, 43
 valori mancanti, 44
 rapporti odd
 in Regressione logistica campioni complessi, 62, 201
 in Regressione ordinale campioni complessi, 73, 213
 in tavole di contingenza campioni complessi, 39, 170
 Regressione Cox di campioni complessi, 221
 analisi di Kaplan-Meier, 77
 definizione evento, 80
 esportazione del modello, 92
 grafici, 88
 grafico log meno log, 268
 informazioni disegno di campionamento, 232, 265
 modello, 84
 opzioni, 94
 predittore dipendente dal tempo, 82, 221
 predittori, 81
 predittori costanti, dipendenti dal tempo, 237
 salvataggio di variabili, 90
 sottogruppi, 83
 statistiche, 86
 stime dei parametri, 237, 266
 test degli effetti del modello, 233, 236, 266
 test dei rischi proporzionali, 233
 test di ipotesi, 89
 valori modello, 267
 variabili di data e ora, 77
 Regressione logistica campioni complessi, 56, 194
 categoria di riferimento, 57
 funzioni aggiuntive del comando, 65
 modello, 58
 opzioni, 64
 procedure correlate, 203
 rapporti odd, 62, 201
 salvataggio di variabili, 63
 statistiche, 60
 statistiche pseudo R^2 , 198
 stime dei parametri, 200
 tabelle di classificazione, 199
 test degli effetti del modello, 200
 Regressione ordinale campioni complessi, 66, 204
 avvisi, 218
 modello, 68
 modello cumulativo generalizzato, 215
 opzioni, 75
 probabilità di risposta, 68
 procedure correlate, 220
 rapporti odd, 73, 213
 salvataggio di variabili, 74
 statistiche, 70
 statistiche pseudo R^2 , 209, 219
 stime dei parametri, 210
 tabelle di classificazione, 212
 test degli effetti del modello, 210
 residui
 in Modello lineare generalizzato campioni complessi, 53
 residui aggregati
 in Regressione Cox di campioni complessi, 90
 residui corretti
 in tavole di contingenza campioni complessi, 39
 residui di devianza
 in Regressione Cox di campioni complessi, 90
 residui martingala
 in Regressione Cox di campioni complessi, 90
 residui parziali di Schoenfeld
 in Regressione Cox di campioni complessi, 90
 residui punteggio
 in Regressione Cox di campioni complessi, 90
 residuo
 in tavole di contingenza campioni complessi, 39
 residuo di Cox-Snell
 in Regressione Cox di campioni complessi, 90

- separazione
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75
- Sidak corretta
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- Sidak sequenziale corretta
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- somma
 - in descrittive campioni complessi, 34
- sottopopolazione
 - in Regressione Cox di campioni complessi, 83
- statistica F
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- statistica F corretta
 - in campioni complessi, 50, 61, 72
 - in Regressione Cox di campioni complessi, 89
- statistica R^2
 - in Modello lineare generalizzato campioni complessi, 49, 188
- statistiche pseudo R^2
 - in Regressione logistica campioni complessi, 60, 198
 - in Regressione ordinale campioni complessi, 70, 209, 219
- stima di campionamento
 - in Preparazione guidata dell'analisi, 22
- stime dei parametri
 - in Modello lineare generalizzato campioni complessi, 49, 189
 - in Regressione Cox di campioni complessi, 86
 - in Regressione logistica campioni complessi, 60, 200
 - in Regressione ordinale campioni complessi, 70, 210
- strati di base
 - in Regressione Cox di campioni complessi, 83
- stratificazione
 - in Campionamento guidato, 6
 - in Preparazione guidata dell'analisi, 20
- struttura di campionamento, completa
 - in Campionamento guidato, 98
- struttura di campionamento, parziale
 - in Campionamento guidato, 110

- tabella riassuntiva
 - in Campionamento guidato, 108, 140
 - in Preparazione guidata dell'analisi, 148, 158
- tabella tavola di contingenza
 - in tavole di contingenza campioni complessi, 173
- tabelle di classificazione
 - in Regressione logistica campioni complessi, 60, 199
 - in Regressione ordinale campioni complessi, 70, 212
- Tavole di contingenza campioni complessi, 37, 170
 - coefficiente di rischio, 170, 174–175
 - procedure correlate, 175
 - statistiche, 39
 - tabella tavola di contingenza, 173
- test degli effetti del modello
 - in Modello lineare generalizzato campioni complessi, 188
 - in Regressione Cox di campioni complessi, 266
 - in Regressione logistica campioni complessi, 200
 - in Regressione ordinale campioni complessi, 210
- test dei rischi proporzionali
 - in Regressione Cox di campioni complessi, 86, 233
- test delle linee parallele
 - in Regressione ordinale campioni complessi, 70, 215
- test t
 - in Modello lineare generalizzato campioni complessi, 49
 - in Regressione logistica campioni complessi, 60
 - in Regressione ordinale campioni complessi, 70

- valori attesi
 - in Modello lineare generalizzato campioni complessi, 53
 - in tavole di contingenza campioni complessi, 39
- valori cumulati
 - in frequenze campioni complessi, 30
- valori mancanti
 - in campioni complessi, 31, 40
 - in descrittive campioni complessi, 35
 - in Modello lineare generalizzato campioni complessi, 54
 - in rapporti campioni complessi, 44
 - in Regressione logistica campioni complessi, 64
 - in Regressione ordinale campioni complessi, 75