

# IBM SPSS Advanced Statistics 19



*Note:* Before using this information and the product it supports, read the general information under Notices a pag. 165.

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

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

**© Copyright SPSS Inc. 1989, 2010.**

---

# Prefazione

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

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

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

## **Supporto tecnico**

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

## **Servizio clienti**

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

## ***Corsi di formazione***

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

## ***Pubblicazioni aggiuntive***

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

---

# Contenuto

## **1 Introduzione ad Advanced Statistics 1**

## **2 Analisi GLM multivariata 2**

GLM - Multivariata: Modello . . . . .	4
Costruisci termini. . . . .	5
Somma dei quadrati. . . . .	5
GLM - Multivariata: Contrasti . . . . .	6
Tipi di contrasto. . . . .	7
GLM - Multivariata: Profili. . . . .	8
GLM multivariata: Confronti post hoc . . . . .	9
GLM - Univariata: Salva . . . . .	11
GLM - Multivariata: Opzioni . . . . .	12
Opzioni aggiuntive del comando GLM . . . . .	14

## **3 GLM - Misure ripetute 15**

GLM - Misure ripetute: Definizione fattori . . . . .	18
GLM - Misure ripetute: Modello . . . . .	19
Costruisci termini. . . . .	20
Somma dei quadrati. . . . .	20
GLM - Misure ripetute: Contrasti. . . . .	21
Tipi di contrasto. . . . .	22
GLM - Misure ripetute: Profili . . . . .	23
GLM - Confronti post hoc delle misure ripetute . . . . .	24
GLM - Misure ripetute: Salva . . . . .	26
GLM - Misure ripetute: Opzioni . . . . .	27
Opzioni aggiuntive del comando GLM . . . . .	29

## **4 Analisi Componenti della varianza 30**

Componenti della varianza: Modello . . . . .	32
Costruisci termini. . . . .	32

Componenti della varianza: Opzioni . . . . .	33
Somma dei quadrati (Componenti della varianza) . . . . .	34
Componenti della varianza: Salva in un nuovo file . . . . .	35
Opzioni aggiuntive del comando VARCOMP . . . . .	35
<b>5   Modelli misti lineari</b>	<b>36</b>
Modelli misti lineari: specificare soggetti e ripetuti . . . . .	38
Effetti fissi nei modelli misti lineari . . . . .	40
Per costruire termini non nidificati . . . . .	40
Costruisci termini nidificati . . . . .	41
Somma dei quadrati . . . . .	41
Effetti casuali nei modelli misti lineari . . . . .	42
Stima dei modelli misti lineari . . . . .	44
Statistiche per modelli misti lineari . . . . .	45
Modelli misti lineari: medie marginali . . . . .	46
Modelli misti lineari: salva . . . . .	47
Opzioni aggiuntive del comando MIXED . . . . .	48
<b>6   Modelli lineari generalizzati</b>	<b>49</b>
Risposta dei modelli lineari generalizzati . . . . .	54
Categoria di riferimento dei modelli lineari generalizzati . . . . .	55
Predittori dei modelli lineari generalizzati . . . . .	56
Opzioni dei modelli lineari generalizzati . . . . .	57
Modello per i modelli lineari generalizzati . . . . .	58
Stima dei modelli lineari generalizzati . . . . .	60
Valori iniziali dei modelli lineari generalizzati . . . . .	62
Statistiche dei modelli lineari generalizzati . . . . .	63
Medie marginali dei modelli lineari generalizzati . . . . .	65
Salvataggio dei modelli lineari generalizzati . . . . .	67
Esportazione dei modelli lineari generalizzati . . . . .	69
Funzioni aggiuntive del comando GENLIN . . . . .	70

## **7 Equazioni di stima generalizzate** **72**

Tipo di modello per le equazioni di stima generalizzate . . . . .	76
Risposta delle equazioni di stima generalizzate . . . . .	80
Categoria di riferimento delle equazioni di stima generalizzate . . . . .	81
Predittori delle equazioni di stima generalizzate . . . . .	82
Opzioni per le equazioni di stima generalizzate . . . . .	83
Modello per le equazioni di stima generalizzate . . . . .	84
Stima delle equazioni di stima generalizzate . . . . .	86
Valori iniziali delle equazioni di stima generalizzate . . . . .	88
Statiche delle equazioni di stima generalizzate . . . . .	89
Medie marginali stimate delle equazioni di stima generalizzate . . . . .	91
Salvataggio delle equazioni di stima generalizzate . . . . .	93
Esportazione delle equazioni di stima generalizzate . . . . .	95
Funzioni aggiuntive del comando GENLIN . . . . .	96

## **8 Modelli misti lineari generalizzati** **98**

Per ottenere un modello misto lineare generalizzato . . . . .	100
Obiettivo . . . . .	101
Effetti fissi . . . . .	104
Aggiungi termine personalizzato . . . . .	105
Effetti casuali . . . . .	107
Blocco effetti casuali . . . . .	108
Peso e offset . . . . .	110
Opzioni di costruzione . . . . .	111
Medie stimate . . . . .	112
Salva . . . . .	114
Vista del modello . . . . .	115
Riepilogo modello . . . . .	115
Struttura dati . . . . .	116
Previsioni e osservazioni . . . . .	116
Classificazione . . . . .	116
Effetti fissi . . . . .	117
Coefficienti fissi . . . . .	117
Covarianze di effetti casuali . . . . .	118
Parametri di covarianza . . . . .	118
Medie stimate: effetti significativi . . . . .	119
Medie stimate: effetti personalizzati . . . . .	119

## **9 Selezione Analisi Loglineare** **121**

Analisi loglineare: Definisci intervallo . . . . .	122
Analisi loglineare: Modello . . . . .	123
Costruisci termini. . . . .	123
Selezione Analisi Loglineare: Opzioni . . . . .	124
Opzioni aggiuntive del comando HILOGLINEAR. . . . .	124

## **10 Analisi loglineare generale** **125**

Analisi loglineare generale: Modello. . . . .	127
Costruisci termini. . . . .	127
Analisi loglineare generale: Opzioni . . . . .	128
Analisi loglineare generale: Salva. . . . .	129
Opzioni aggiuntive del comando GENLOG . . . . .	129

## **11 Analisi loglineare Logit** **130**

Analisi loglineare Logit: Modello. . . . .	132
Costruisci termini. . . . .	133
Analisi loglineare Logit: Opzioni . . . . .	133
Analisi loglineare Logit: Salva. . . . .	134
Opzioni aggiuntive del comando GENLOG . . . . .	135

## **12 Tavole di sopravvivenza** **136**

Tavole di sopravvivenza: Definisci evento per variabile di stato. . . . .	138
Tavole di sopravvivenza: Definisci intervallo . . . . .	138
Tavole di sopravvivenza: Opzioni. . . . .	139
Opzioni aggiuntive del comando SURVIVAL. . . . .	139

## **13 Analisi di sopravvivenza di Kaplan-Meier** **141**

Kaplan-Meier: Definisci evento per variabile di stato . . . . .	143
Kaplan-Meier: Confronta i livelli del fattore. . . . .	143

Kaplan-Meier: Salva nuove variabili . . . . .	144
Kaplan-Meier: Opzioni . . . . .	145
Opzioni aggiuntive del comando KM . . . . .	145

## **14 Analisi della regressione di Cox 146**

Regressione di Cox: Definisci variabili categoriali . . . . .	148
Regressione di Cox: Grafici. . . . .	149
Regressione di Cox: Salva nuove variabili. . . . .	150
Regressione di Cox: Opzioni . . . . .	151
Regressione di Cox: Definisci evento per la variabile di stato . . . . .	152
Opzioni aggiuntive del comando COXREG . . . . .	152

## **15 Calcolo di covariate dipendenti dal tempo 153**

Per calcolare una covariata dipendente dal tempo . . . . .	154
Opzioni aggiuntive della regressione di Cox con covariate dipendenti dal tempo . . . . .	154

## **Appendici**

### **A Schemi di codifica delle variabili categoriali 155**

Deviazione . . . . .	155
Semplice . . . . .	156
Helmert . . . . .	157
Differenza . . . . .	157
Polinomiale. . . . .	158
Ripetuto . . . . .	158
Contrasto speciale . . . . .	159
Indicatore . . . . .	160

<b><i>B</i></b>	<b><i>Strutture della covarianza</i></b>	<b>161</b>
<b><i>C</i></b>	<b><i>Notices</i></b>	<b>165</b>
	<b><i>Indice</i></b>	<b>167</b>

# ***Introduzione ad Advanced Statistics***

Il modulo Advanced Statistics fornisce una serie di procedure che offrono opzioni di modellazione più avanzate rispetto a quelle disponibili nel modulo Statistics Base.

- L'Analisi GLM multivariata amplia il modello lineare generale fornito dall'Analisi GLM univariata per consentire l'uso di più variabili dipendenti. L'analisi GLM Misure ripetute, che rappresenta un ulteriore ampliamento rispetto al modello di base, consente di effettuare misurazioni ripetute su più variabili dipendenti.
- L'analisi Componenti della varianza è uno strumento specifico che consente di scomporre la variabilità di una variabile dipendente in componenti fissi e casuali.
- L'analisi Modelli misti lineari consente di ampliare il modello lineare generale in modo che i dati mostrino una variabilità correlata e non costante. La flessibilità del modello misto lineare, consente dunque di modellare, oltre al significato dei dati, anche le relative varianze e covarianze.
- L'analisi Modelli lineari generalizzati (GZLM) rilassa l'assunzione di normalità per il termine di errore e richiede solo che la variabile dipendente venga linearmente correlata ai predittori mediante una trasformazione o una funzione di collegamento. L'analisi Equazioni di stima generalizzate (GEE) amplia la procedura GZLM per permettere l'esecuzione di misurazioni ripetute.
- L'Analisi loglineare generale consente di adattare i modelli a dati dei conteggi classificati in modo incrociato, mentre l'Analisi loglineare con selezione del modello permette di scegliere tra due modelli.
- L'Analisi loglineare logit consente di adattare i modelli loglineari in modo da poter analizzare la relazione tra una variabile categoriale dipendente e uno o più predittori categoriali.
- L'Analisi di sopravvivenza disponibile tramite le Tavole di sopravvivenza permette di esaminare la distribuzione delle variabili tempo-evento con eventuali livelli della variabile fattore. L'Analisi di sopravvivenza di Kaplan-Meier consente di esaminare la distribuzione di variabili tempo-evento con gli eventuali livelli di una variabile fattore oppure di generare analisi distinte utilizzando i livelli di una variabile di stratificazione. La Regressione di Cox consente di creare modelli di tempo per un evento specifico utilizzando i valori delle covariate note.

# ***Analisi GLM multivariata***

La procedura GLM - Multivariata consente di eseguire l'analisi di regressione e l'analisi della varianza per più variabili dipendenti in base a una o più variabili fattore o covariate. Le variabili fattore suddividono la popolazione in gruppi. Con questa procedura di modello lineare generalizzato è possibile verificare le ipotesi nulle relative agli effetti delle variabili fattore sulle medie dei vari raggruppamenti di una distribuzione congiunta di variabili dipendenti. È possibile analizzare le interazioni tra fattori e gli effetti dei singoli fattori. È inoltre possibile includere gli effetti delle covariate e le interazioni tra covariate e fattori. Nell'analisi di regressione, le variabili indipendenti (stimatori) vengono specificate come covariate.

È possibile verificare sia modelli bilanciati che modelli non bilanciati. Un disegno è bilanciato se ciascuna cella del modello include lo stesso numero di casi. In un modello multivariato, la somma dei quadrati dovuta agli effetti del modello e la somma dei quadrati dell'errore sono in forma di matrice anziché in forma scalare come nell'analisi univariata. Tali matrici sono definite SSCP, acronimo dell'inglese Sums-of-Squares e Cross Products, ovvero somma dei quadrati e prodotti incrociati. Se si specifica più di una variabile dipendente, oltre all'analisi della varianza univariata per ciascuna variabile dipendente, viene eseguita anche l'analisi della varianza multivariata usando la traccia di Pillai, il lambda di Wilks, la traccia di Hotelling e la radice di Roy con statistica  $F$  approssimata. Oltre alla verifica delle ipotesi, la procedura GLM - Multivariata consente di ottenere stime dei parametri.

Per la verifica delle ipotesi sono disponibili i contrasti *a priori* usati di frequente. Dopo che da un test  $F$  globale è risultata una certa significatività, è inoltre possibile eseguire test post hoc per valutare le differenze tra medie specifiche. La procedura Medie marginali stimate consente di ottenere stime dei valori medi attesi per le celle incluse nel modello. I grafici di profilo, o grafici di interazione, di tali medie consentono di visualizzare in modo semplice alcune relazioni. I test per confronti multipli post hoc vengono eseguiti separatamente per ciascuna variabile dipendente.

Residui, valori attesi, distanza di Cook e valori d'influenza possono essere salvati come variabili nel file di dati per la verifica di ipotesi. Sono inoltre disponibili una matrice SSCP dei residui, ovvero una matrice quadrata di somme dei quadrati e prodotti incrociati dei residui, una matrice di covarianza dei residui, che corrisponde alla matrice SSCP dei residui divisa per il grado di libertà dei residui, e la matrice di correlazione dei residui, che corrisponde alla forma standardizzata della matrice di covarianza dei residui.

Minimi quadrati ponderati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni per un'analisi di minimi quadrati ponderati (WLS), in alcuni casi per compensare la diversa precisione della misura.

**Esempio.** Una ditta di prodotti in plastica misura tre proprietà del materiale plastico: resistenza, lucidezza e opacità. Dopo aver eseguito prove con tre tassi di estrusione e due diverse quantità di additivo, le tre proprietà vengono misurate in tutte le combinazioni possibili dei tassi di estrusione e delle quantità di additivo. Dall'analisi risulta che i tassi di estrusione e le quantità

di additivo producono risultati significativi se usati separatamente, mentre l'interazione tra i due fattori risulta non significativa.

**Metodi.** Per la valutazione di ipotesi diverse è possibile usare la somma dei quadrati Tipo I, Tipo II, Tipo III e Tipo IV. Il metodo predefinito è il Tipo III.

**Statistiche.** I seguenti test post hoc di intervalli e confronti multipli: Differenza meno significative (LSD), Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiplo basato su test  $F$ , Ryan-Einot-Gabriel-Welsch a intervallo multiplo, Student-Newman-Keuls, differenze significative di Tukey,  $b$  di Tukey, Duncan, Hochberg (GT2), Gabriel,  $t$  di Waller-Duncan, Dunnett (a una e a due vie), Tamhane (T2), Dunnett (T3), Games-Howell e  $C$  di Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle; test Levene per l'omogeneità della varianza; test  $M$  di Box per l'omogeneità delle matrici di covarianza delle variabili dipendenti e test di sfericità di Bartlett.

**Grafici.** Grafici di variabilità vs. densità, dei residui e di profilo (interazione).

**Dati.** Le variabili dipendenti devono essere quantitative. I fattori sono categoriali e possono essere associati a valori numerici o a valori stringa. Le covariate sono variabili quantitative correlate alla variabile dipendente.

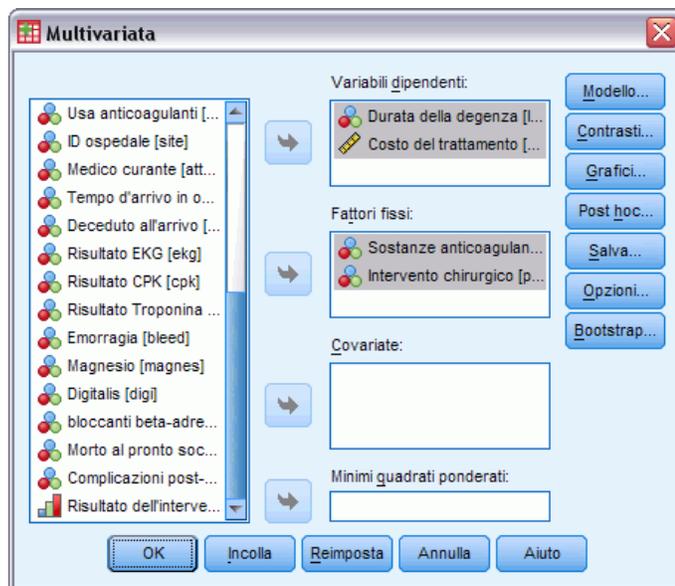
**Assunzioni.** Per le variabili dipendenti, i dati sono un campione casuale di vettori derivato da una popolazione normale multivariata nella quale le matrici di varianza-covarianza di tutte le celle sono uguali. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. Per la verifica delle assunzioni, è possibile usare i test di omogeneità della varianza, incluso il test  $M$  di Box, e i grafici di variabilità vs. densità. È inoltre possibile esaminare residui e grafici dei residui.

**Procedure correlate.** Prima di eseguire un'analisi di varianza, esaminare i dati tramite la procedura Esplora. Nel caso di una variabile dipendente singola, usare la procedura GLM univariata. Se le stesse variabili dipendenti sono state misurate in più occasioni per ciascun soggetto, usare la procedura GLM - Misure ripetute.

#### ***Per ottenere un'analisi della varianza GLM - Multivariata***

- Dai menu, scegliere:  
Analizza > Modello lineare generalizzato > Multivariata...

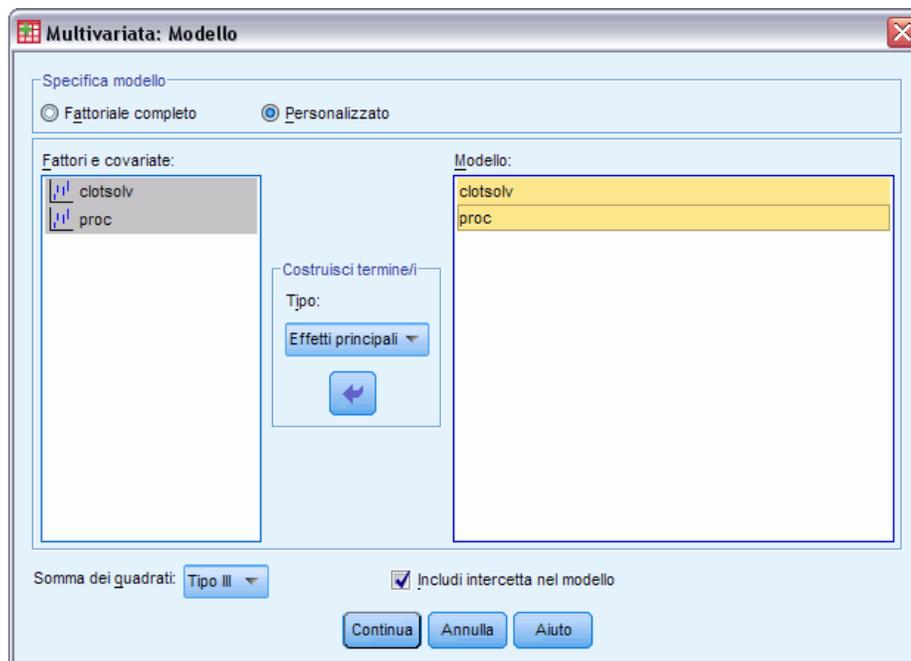
Figura 2-1  
Finestra di dialogo Multivariata



- Selezionare almeno due variabili dipendenti.  
Se necessario, è possibile specificare le opzioni Fattori fissi, Covariate e Minimi quadrati ponderati.

## GLM - Multivariata: Modello

Figura 2-2  
Finestra di dialogo Multivariata: Modello



**Specifica modello.** Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

**Fattori e covariate.** I fattori e le covariate sono elencati.

**Modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire.

**Somma dei quadrati.** Metodo per il calcolo della somma dei quadrati. Il metodo Somma dei quadrati in genere utilizzato con modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

**Includi l'intercetta nel modello.** L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

## **Costruisci termini**

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

## **Somma dei quadrati**

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo predefinito, è quello usato più di frequente.

**Tipo I.** Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

**Tipo II.** Questo metodo consente di calcolare le somme dei quadrati di un effetto del modello corretto per tutti gli altri effetti “appropriati”. È considerato appropriato un effetto corrispondente a tutti gli effetti che non includono l’effetto in esame. Il metodo Somma dei quadrati Tipo II è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato.
- Qualsiasi modello che include solo effetti principali del fattore.
- Qualsiasi modello di regressione.
- Un disegno nidificato in modo puro. Questo tipo di nidificazione può essere specificato tramite la sintassi.

**Tipo III.** Tipo predefinito. Questo metodo consente di calcolare la somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

**Tipo IV.** Questo metodo è specifico per situazioni con celle mancanti. Per qualsiasi effetto  $F$  del disegno, se  $F$  non è incluso in nessun altro effetto, allora Tipo IV = Tipo III = Tipo II. Se invece  $F$  è incluso in altri effetti, con il Tipo IV i contrasti creati tra i parametri in  $F$  vengono distribuiti equamente tra tutti gli effetti di livello superiore. Il metodo Somma dei quadrati Tipo IV viene in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato e non bilanciato contenente celle vuote.

## GLM - Multivariata: Contrasti

Figura 2-3  
Finestra di dialogo Multivariata: Contrasti



I contrasti consentono di verificare la significatività del grado di differenza tra i livelli di un effetto. È possibile specificare un contrasto per ciascun fattore del modello. I contrasti rappresentano combinazioni lineari dei parametri.

La verifica delle ipotesi è basata sull'ipotesi nulla  $\mathbf{LBM} = \mathbf{0}$ , dove  $\mathbf{L}$  è la matrice dei coefficienti di contrasto,  $\mathbf{M}$  è la matrice identità con dimensione pari al numero di variabili dipendenti e  $\mathbf{B}$  è il vettore dei parametri. Quando si specifica un contrasto, viene creata una matrice  $\mathbf{L}$  tale che le colonne corrispondenti al fattore corrispondono al contrasto. Le altre colonne vengono corrette in modo che la matrice  $\mathbf{L}$  possa essere stimata.

Oltre al test univariato con statistiche  $F$  e agli intervalli di confidenza simultanei tipo Bonferroni basati sulla distribuzione  $t$  di Student per le differenze tra contrasti in tutte le variabili dipendenti, sono disponibili anche i test multivariati in cui vengono usate la traccia di Pillai, il lambda di Wilks, la traccia di Hotelling e la radice di Roy.

Sono disponibili i contrasti deviazione, semplici, differenza, Helmert, ripetuti e polinomiali. Per i contrasti deviazione e i contrasti semplici, è possibile stabilire se la categoria di riferimento corrisponde alla prima o all'ultima categoria.

### **Tipi di contrasto**

**Deviazione.** Consente di confrontare la media di ciascun livello, a 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.

**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. Come categoria di riferimento, è possibile scegliere la prima o l'ultima categoria.

**Differenza.** Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Questo tipo di contrasto è a volte definito contrasto inverso di Helmert.

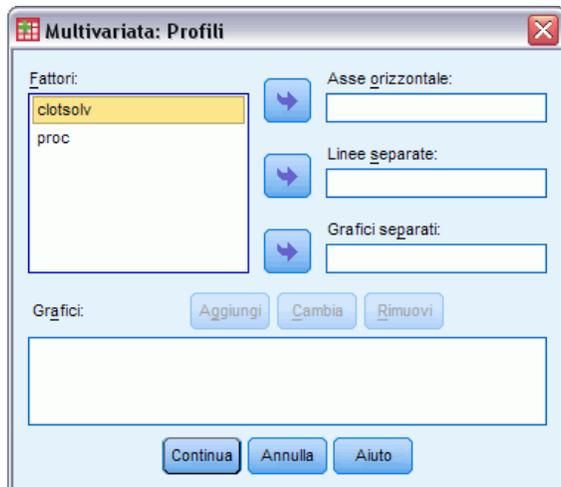
**Helmert.** Consente di confrontare la media di ciascun livello del fattore (a eccezione dell'ultimo) con la media dei livelli successivi.

**Ripetuto.** 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.

## GLM - Multivariata: Profili

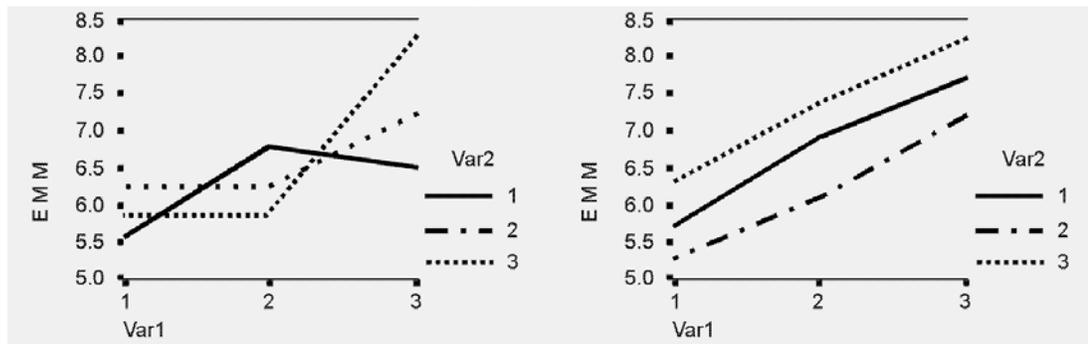
Figura 2-4  
Finestra di dialogo Multivariata: Profili



I profili, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un profilo è un grafico lineare in cui ciascun punto indica la media marginale stimata di una variabile dipendente (corretta per le covariate) in corrispondenza di un solo livello di un fattore. È possibile utilizzare i livelli di un secondo fattore per creare linee distinte. È possibile utilizzare ciascun livello di un terzo fattore per creare un grafico distinto. Tutti i fattori sono disponibili per i grafici. Questi grafici vengono creati per ciascuna variabile dipendente.

Il profilo di un fattore mostra se le medie marginali stimate aumentano o diminuiscono tra i vari livelli. Nel caso di due o più fattori, le linee parallele indicano che tra i fattori non esiste alcuna interazione, ovvero che è possibile analizzare i livelli di un solo fattore. Le linee che si incrociano indicano invece che esiste un'interazione.

Figura 2-5  
Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)

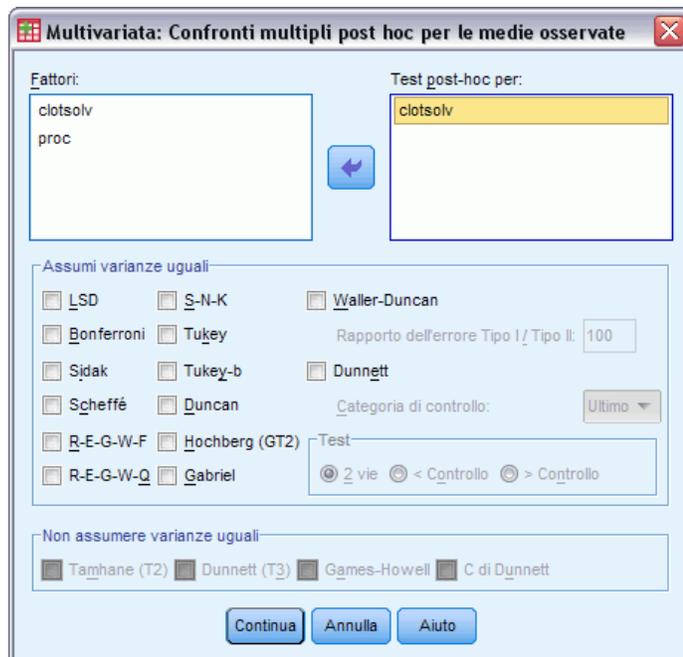


I grafici definiti tramite la selezione dei fattori per l'asse orizzontale e, se lo si desidera, dei fattori di linee e di grafici separati devono essere inclusi nell'elenco dei grafici.

## GLM multivariata: Confronti post hoc

Figura 2-6

Finestra di dialogo Confronti multipli post hoc multivariati per le medie osservate



**Test per confronti multipli post hoc.** Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stata apportata alcuna correzione. I test post hoc vengono eseguiti separatamente per ciascuna variabile dipendente.

I test per confronti multipli usati più di frequente sono il test di Bonferroni e il test delle differenze significative di Tukey. Il **test di Bonferroni**, basato sulla statistica  $t$  di Student, consente di correggere il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test  $t$  di Sidak** corregge inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test delle differenze significative di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti per coppie. Quando si eseguono test su un elevato numero di coppie di medie, il test delle differenze significative di Tukey risulta più efficace rispetto al test di Bonferroni. Nel caso di un numero limitato di coppie, risulta invece più efficace il test di Bonferroni.

Il test di **Hochberg (GT2)** è simile al test delle differenze significative di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test dei confronti a coppie di Gabriel** utilizza il modulo massimo studentizzato ed è in genere più indicativo del test di Hochberg (GT2) quando le dimensioni delle celle sono diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.

Il **test  $t$  per confronti multipli a coppie di Dunnett** confronta un insieme di trattamenti con una media di controllo singola. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. È inoltre possibile scegliere un test a 2

vie oppure a 1 via. Per verificare che la media in qualsiasi livello del fattore (a eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo, è necessario usare un test a due sensi. Per verificare se la media di qualsiasi livello del fattore è minore di quella della categoria di controllo, selezionare < Controllo. In modo analogo, per verificare se la media di qualsiasi livello del fattore è maggiore di quella della categoria di controllo, selezionare > Controllo.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test a intervalli decrescenti multipli. Le procedure a multipli decrescenti verificano in primo luogo se tutte le medie sono uguali. Se le medie non risultano tutte uguali, il test di uguaglianza viene eseguito su un sottoinsieme di medie. Il test **R-E-G-W F** è basato su un test  $F$ , mentre **R-E-G-W Q** è basato sull'intervallo studentizzato. Questi test risultano più efficaci rispetto ai test a intervallo multiplo di Duncan e Student-Newman-Keuls, che sono pure procedure a intervalli decrescenti multipli. È tuttavia consigliabile non usarli con celle di dimensioni non uguali.

Quando le varianze non sono uguali, è necessario utilizzare il test **Tamhane (T2)** (test per confronti a coppie conservativo basato su un test  $t$ ), il test di **Dunnnett T3** (test per confronti a coppie basato sul modulo studentizzato), il test per confronti a coppie di **Games-Howell** (a volte poco conservativo) o il test **C di Dunnnett** (test per confronti a coppie basato sull'intervallo studentizzato).

**Il test a intervallo multiplo di Duncan**, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test per intervallo che classificano le medie raggruppate e calcolano un valore di intervallo. Questi test sono usati meno frequentemente dei test descritti in precedenza.

**Il test t di Waller-Duncan** utilizza un approccio bayesiano. Si tratta di un test a intervallo che usa la media armonica della dimensione campione nel caso di dimensioni campione non uguali.

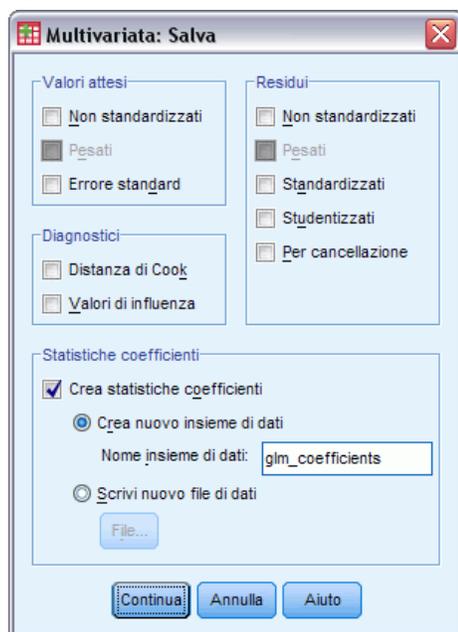
Il livello di significatività del test di **Scheffé** consente la verifica di tutte le possibili combinazioni lineari delle medie di gruppo e non dei soli confronti a coppie disponibili in questa funzione. Il test di Scheffé risulta pertanto più conservativo rispetto ad altri test, ovvero per ottenere un livello sufficiente di significatività, è richiesta una differenza maggiore tra le medie.

Il test per confronti a coppie multipli Differenza meno significativa o **LSD**, è equivalente a più test  $t$  tra tutte le coppie di gruppi. Lo svantaggio di questo test è che non viene eseguito alcun tentativo di correzione del livello di significatività osservata per confronti multipli.

**Test visualizzati.** I confronti a coppie sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), test C di Dunnnett e Dunnnett (T3). Per i test S-N-K,  $b$  di Tukey, Duncan, R-E-G-W  $F$ , R-E-G-W  $Q$  e Waller sono disponibili sottoinsiemi omogenei per test per intervallo. Il test delle differenze significative di Tukey, i test Hochberg (GT2), Gabriel e Scheffé sono sia test per confronti multipli che test a intervallo.

## GLM - Univariato: Salva

Figura 2-7  
Salva



È possibile salvare i valori attesi dal modello, le misure correlate e i residui come nuove variabili nell'Editor dei dati. Molte di queste variabili possono essere usate per l'esame di ipotesi sui dati. Per salvare i valori in modo da poterli usare in un'altra sessione IBM® SPSS® Statistics, è necessario salvare il file di dati corrente.

**Valori attesi.** Valori attesi dal modello per ciascun caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Pesati.** I valori attesi ponderati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Errore standard.** Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

**Diagnostici.** Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere una notevole influenza sul modello.

- **Distanza di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione ne modificherebbe sostanzialmente il valore.
- **Valori di influenza.** Valori di influenza non centrati. Una misura dell'influenza di ciascun caso sull'adattamento di un modello di regressione.

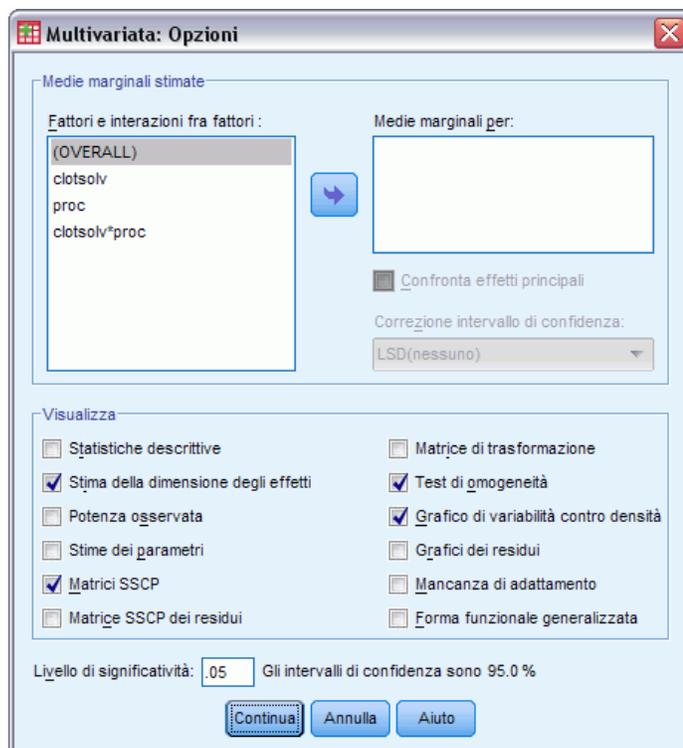
**Residui.** Un residuo non standardizzato corrisponde al valore effettivo della variabile dipendente diminuito del valore atteso dal modello. Sono inoltre disponibili residui standardizzati, studentizzati ed eliminati. Se è stata selezionata una variabile WLS, saranno inoltre disponibili residui non standardizzati pesati.

- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Pesati.** I residui ponderati non standardizzati. Disponibile solo se è stata selezionata una variabile WLS.
- **Standardizzati.** Il residuo diviso per una stima della deviazione standard. Il residuo standardizzato, conosciuto anche come residuo di Pearson, ha media 0 e deviazione standard 1.
- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza tra i valori assunti per questo caso dalle variabili indipendenti e le medie delle variabili indipendenti.
- **Per cancellazione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore stimato corretto.

**Statistiche dei coefficienti.** Scrive una matrice della varianza-covarianza delle stime dei parametri nel modello in un nuovo file di dati della sessione attiva o in un file di dati SPSS Statistics esterno. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime dei parametri, una riga di valori di significatività per le statistiche  $t$  corrispondenti alle stime dei parametri e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. Il file matrice può essere usato in altre procedure che leggono i file matrice.

## GLM - Multivariata: Opzioni

Figura 2-8  
Finestra di dialogo Multivariata: Opzioni



In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

**Medie marginali stimate.** Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono corrette per le eventuali covariate. Le interazioni sono disponibili solo se è stato impostato un modello personalizzato.

- **Confronta effetti principali.** Consente di eseguire confronti a coppie senza correzione tra le medie marginali stimate di qualsiasi effetto principale del modello, per fattori sia tra soggetti che entro soggetti. Questa opzione è disponibile solo se nell'elenco Medie marginali per sono stati selezionati effetti principali.
- **Correzione intervallo di confidenza.** Selezionare la differenza meno significativa (LSD), la correzione di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo comando è disponibile solo se è stato selezionato Confronta effetti principali.

**Visualizzazione.** Selezionare Statistiche descrittive per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione Stima della dimensione degli effetti fornisce un valore eta-quadrato parziale per ciascun effetto e per ciascuna stima dei parametri. La statistica eta-quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare Potenza osservata per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare Stime dei parametri per ottenere stime dei parametri, errori standard, test  $t$ , intervalli di confidenza e la potenza osservata per ciascun test. È possibile visualizzare le matrici SSCP di ipotesi ed errore e la matrice SSCP dei residui, nonché il test della sfericità di Bartlett della matrice di covarianza dei residui.

La funzione Test di omogeneità produce il test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori fra soggetti, solo per i fattori fra soggetti. È un test di omogeneità anche il test  $M$  di Box dell'omogeneità delle matrici di covarianza delle variabili dipendenti in tutte le possibili combinazioni di livello dei fattori tra soggetti. Le opzioni Grafici di variabilità vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare Grafici dei residui per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare Mancanza di adattamento per controllare che la relazione fra la variabile dipendente e le variabili indipendenti possa essere descritta in modo adeguato dal modello. La funzione Forma funzionale generalizzata consente di creare test di ipotesi personalizzati basati sulla forma funzionale generalizzata. Le righe di una matrice dei coefficienti di contrasto sono combinazioni lineari della forma funzionale generalizzata.

**Livello di significatività.** Potrebbe risultare utile correggere il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

## **Opzioni aggiuntive del comando GLM**

Queste opzioni sono valide per analisi univariate, multivariate e a misure ripetute. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati del disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).
- Specificare criteri EPS (tramite il sottocomando `CRITERIA`).
- Costruire una matrice **L**, **M** o **K** personalizzata (tramite il sottocomando `LMATRIX`, `MMATRIX` o `KMATRIX`).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando `CONTRAST`).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando `CONTRAST`).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando `POSTHOC`).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori per i fattori elencati (tramite il sottocomando `EMMEANS`).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando `SAVE`).
- Costruire un file di dati di matrici di correlazione (tramite il sottocomando `OUTFILE`).
- Costruire un file di dati di matrici contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando `OUTFILE`).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando `OUTFILE`).

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

## ***GLM - Misure ripetute***

La procedura GLM - Misure ripetute consente di eseguire un'analisi della varianza quando su ciascun soggetto o caso viene eseguita più volte la stessa misura. Se sono stati specificati fattori tra soggetti, la popolazione verrà suddivisa in gruppi. Questa procedura di modello lineare generalizzato consente di verificare ipotesi nulle relative agli effetti sia di fattori tra soggetti che di fattori entro soggetti. È possibile analizzare le interazioni tra fattori e gli effetti dei singoli fattori. È inoltre possibile includere gli effetti delle covariate costanti e delle interazioni tra covariate con i fattori tra soggetti.

In un disegno a misure ripetute doppiamente multivariato, le variabili dipendenti rappresentano le misure di più variabili per i vari livelli dei fattori entro soggetti. È possibile, ad esempio, aver misurato sia il battito cardiaco che la respirazione di ciascun soggetto in tre momenti diversi.

La procedura GLM - Misure ripetute consente di eseguire analisi sia univariate che multivariate per i dati di misure ripetute. È possibile verificare sia modelli bilanciati che modelli non bilanciati. Un disegno è bilanciato se ciascuna cella del modello include lo stesso numero di casi. In un modello multivariato, la somma dei quadrati dovuta agli effetti del modello e la somma dei quadrati dell'errore sono in forma di matrice anziché in forma scalare come nell'analisi univariata. Tali matrici sono definite SSCP, acronimo dell'inglese Sums-of-Squares e Cross Products, ovvero somma dei quadrati e prodotti incrociati. Oltre alla verifica delle ipotesi, la procedura GLM - Misure ripetute consente di ottenere stime dei parametri.

Per la verifica di ipotesi su fattori tra soggetti sono disponibili contrasti *a priori* usati di frequente. Dopo che da un test  $F$  globale è risultata una certa significatività, è inoltre possibile eseguire test post hoc per valutare le differenze tra medie specifiche. La procedura Medie marginali stimate consente di ottenere stime dei valori medi attesi per le celle incluse nel modello. I grafici di profilo, o grafici di interazione, di tali medie consentono di visualizzare in modo semplice alcune relazioni.

Residui, valori attesi, distanza di Cook e valori d'influenza possono essere salvati come variabili nel file di dati per la verifica di ipotesi. Sono inoltre disponibili una matrice SSCP dei residui, ovvero una matrice quadrata di somme dei quadrati e prodotti incrociati dei residui, una matrice di covarianza dei residui, che corrisponde alla matrice SSCP dei residui divisa per il grado di libertà dei residui, e la matrice di correlazione dei residui, che corrisponde alla forma standardizzata della matrice di covarianza dei residui.

Minimi quadrati ponderati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni per un'analisi di minimi quadrati ponderati (WLS), in alcuni casi per compensare la diversa precisione della misura.

**Esempio.** Dodici studenti vengono suddivisi tra gruppi di soggetti affetti da alta o bassa ansietà in base ai corrispondenti punteggi ottenuti da un test di stima dell'ansietà. La stima dell'ansietà è un fattore tra soggetti poiché determina la suddivisione dei soggetti in gruppi. A ciascun studente vengono assegnate quattro prove per un'attività di apprendimento e per ciascuna prova viene

registrato il numero di errori in variabili distinte. Per le quattro prove viene inoltre definito un fattore entro soggetti (prova) con quattro livelli. Dallo studio risulta che l'effetto di prova è significativo, mentre l'interazione tra prova e ansietà non lo è.

**Metodi.** Per la valutazione di ipotesi diverse è possibile usare la somma dei quadrati Tipo I, Tipo II, Tipo III e Tipo IV. Il metodo predefinito è il Tipo III.

**Statistiche.** I seguenti test post hoc di intervalli e confronti multipli (per fattori tra soggetti): Differenza meno significative (LSD), Bonferroni, Sidak, Scheffé, Ryan-Einot-Gabriel-Welsch multiplo basato su test  $F$ , Ryan-Einot-Gabriel-Welsch a intervallo multiplo, Student-Newman-Keuls, differenze significative di Tukey,  $b$  di Tukey, Duncan, Hochberg (GT2), Gabriel,  $t$  di Waller-Duncan, Dunnett (a una e a due vie), Tamhane (T2), Dunnett (T3), Games-Howell e  $C$  di Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle; il test di Levene per l'omogeneità della varianza; il test  $M$  di Box e il test di sfericità di Mauchly.

**Grafici.** Grafici di variabilità vs. densità, dei residui e di profilo (interazione).

**Dati.** Le variabili dipendenti devono essere quantitative. I fattori tra soggetti suddividono il campione in sottogruppi discreti, ad esempio maschio e femmina. Questi fattori sono categoriali e possono essere associati a valori numerici o a valori stringa. I fattori entro soggetti sono definiti nella finestra di dialogo Misure ripetute: Definizione fattori. Le covariate sono variabili quantitative correlate alla variabile dipendente. In un'analisi a misure ripetute, le covariate devono rimanere costanti a ciascun livello di una variabile entro soggetti.

Il file di dati deve includere un insieme di variabili per ciascun gruppo di misure eseguite sui soggetti. Nell'insieme, a ciascuna ripetizione della misura all'interno del gruppo è associata una variabile. Per il gruppo che include un numero di livelli pari al numero di ripetizioni viene definito un fattore entro soggetti. La misura del peso, ad esempio, potrebbe essere controllata in giorni diversi. Se le misure di una stessa proprietà sono state registrate per cinque giorni, il fattore entro soggetti può essere specificato come *giorno* con cinque livelli.

Per fattori entro soggetti multipli, il numero di misure di ciascun soggetto è uguale al prodotto del numero di livelli di ciascun fattore. Ad esempio, se le misure sono state prese in tre momenti diversi del giorno per quattro giorni, il numero totale di misure per ciascun soggetto sarà uguale a 12. I fattori entro soggetti possono essere specificati come *giorno* (4) e *ora* (3).

**Assunzioni.** Per eseguire un'analisi a misure ripetute è possibile adottare un approccio univariato o multivariato.

In base all'approccio univariato, definito anche approccio a split-plot o a modello misto, le variabili dipendenti vengono considerate come risposte ai livelli dei fattori entro soggetti. Le misure eseguite su un soggetto devono essere un campione di una distribuzione normale multivariata, mentre le matrici di varianza-covarianza sono uguali in tutte le celle formate dagli effetti tra soggetti. Sulla matrice di varianza-covarianza delle variabili dipendenti vengono formulate determinate ipotesi. La validità della statistica  $F$  usata nell'approccio univariato è assicurata se la matrice di varianza-covarianza è di forma circolare (Huynh e Mandeville, 1979).

Per verificare questa ipotesi, è possibile usare il test della sfericità di Mauchly, il base al quale il test di sfericità viene eseguito sulla matrice di varianza-covarianza di una variabile dipendente sottoposta a trasformazione ortonormalizzata. Per un'analisi a misure ripetute viene automaticamente visualizzato il test di Mauchly. Con dimensioni campionarie limitate, questo test risulta tuttavia poco efficace. Con dimensioni campionarie grandi può invece risultare

significativo anche quando il peso dello scostamento dei risultati è ridotto. Se il grado di significatività del test è elevato, è possibile considerare l'ipotesi della sfericità come vera. Se invece il grado di significatività è ridotto e l'ipotesi di sfericità sembra essere violata, è possibile apportare una correzione ai gradi di libertà a numeratore e a denominatore in modo da convalidare la statistica  $F$  univariata. Nella procedura GLM - Misure ripetute sono disponibili tre stime di questa correzione, definita **epsilon**. I gradi di libertà sia a numeratore che a denominatore devono essere moltiplicati per epsilon e la significatività del rapporto  $F$  deve essere valutata con il nuovo valore di gradi di libertà.

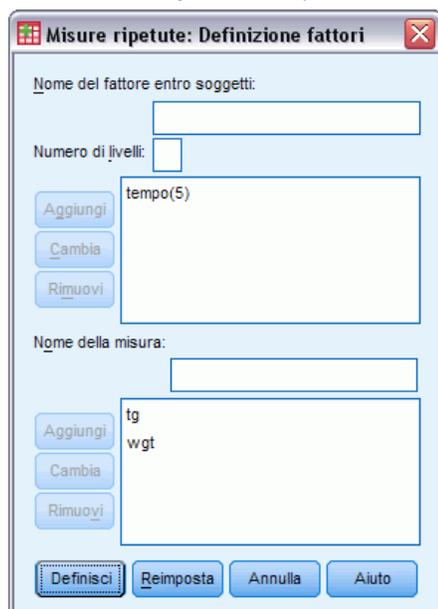
Con l'approccio multivariato, le misure eseguite su un soggetto vengono considerate come un campione derivato da una distribuzione normale multivariata e le matrici di varianza-covarianza risultano uguali in tutte le celle formate dagli effetti tra soggetti. Per verificare se le matrici di varianza-covarianza delle varie celle sono uguali, è possibile usare il test  $M$  di Box.

**Procedure correlate.** Prima di eseguire un'analisi di varianza, esaminare i dati tramite la procedura Esplora. Se *non* esistono misure ripetute in ciascun soggetto, usare la procedura GLM multivariata o GLM univariata. Se per ciascun soggetto sono disponibili solo due misure, ad esempio un pre-test e un post-test, e non sono disponibili fattori tra soggetti, è possibile usare la procedura Test t per campioni appaiati.

### Ottenere GLM a misure ripetute

- Dai menu, scegliere:  
Analizza > Modello lineare generalizzato > Misure ripetute...

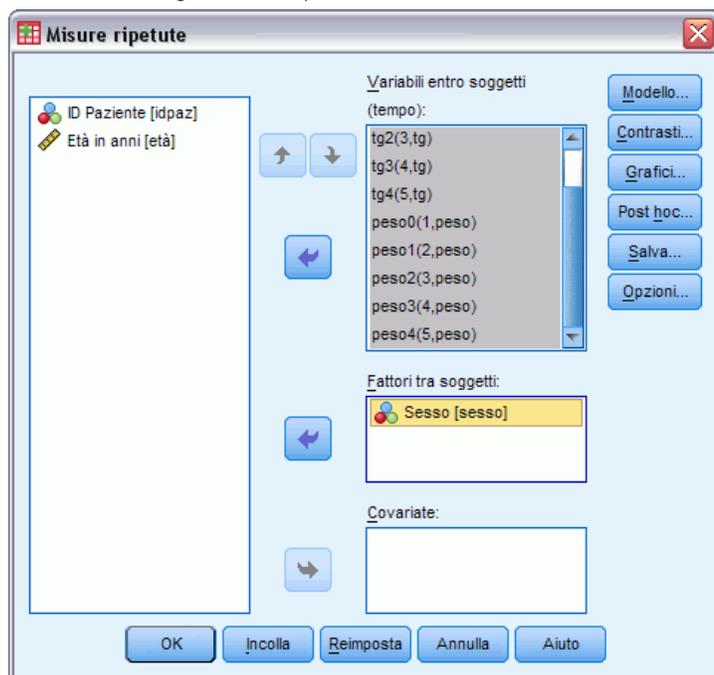
Figura 3-1  
Finestra di dialogo Misure ripetute: Definizione fattori



- Inserire il nome di un fattore entro soggetti e il corrispondente numero di livelli.
- Fare clic su Aggiungi.

- ▶ Ripetere la stessa operazione per ciascun fattore entro soggetti.  
Per definire i fattori misura di un disegno a misure ripetute doppiamente multivariato:
- ▶ Inserire il nome della misura.
- ▶ Fare clic su Aggiungi.  
Dopo aver definito tutti i fattori e le misure:
- ▶ Fare clic su Definisci.

Figura 3-2  
Finestra di dialogo Misure ripetute



- ▶ Selezionare dall'elenco una variabile dipendente corrispondente a ciascuna combinazione di fattori entro soggetti ed eventualmente a ciascuna misura.

Per spostare le variabili, usare le frecce verso l'alto e verso il basso.

Per modificare i fattori entro soggetti, è possibile visualizzare di nuovo la finestra di dialogo Misure ripetute: Definizione fattori senza chiudere la finestra principale. Se necessario, è possibile specificare i fattori tra soggetti e le covariate.

## **GLM - Misure ripetute: Definizione fattori**

La procedura GLM - Misure ripetute consente di analizzare gruppi di variabili dipendenti correlate che rappresentano misure diverse dello stesso attributo. In questa finestra di dialogo è possibile definire uno o più fattori entro soggetti da usare in una procedura GLM - Misure ripetute. Vedere [Figura 3-1](#) a pag. 17. L'ordine con cui vengono specificati i fattori entro soggetti è importante. Ciascun fattore rappresenta un livello incluso nel fattore precedente.

Per poter usare la procedura Misure ripetute, è necessario impostare i dati in modo corretto. In questa finestra di dialogo è necessario definire i fattori entro soggetti. Tali fattori non sono variabili esistenti nei dati, ma fattori definiti in questa finestra di dialogo.

**Esempio.** Si supponga che, in uno studio relativo alla perdita di peso, più persone vengano pesate ogni settimana per cinque settimane. Nel file di dati, ciascuna persona è un soggetto o un caso. Il peso viene registrato ogni settimana nelle variabili *peso1*, *peso2* e così via. Il sesso di ogni persona viene registrato in una variabile diversa. I pesi, misurati più volte per ciascun soggetto, possono essere raggruppati definendo un fattore entro soggetti. Il fattore potrebbe essere chiamato *settimana* e definito in modo da avere cinque livelli. Nella finestra di dialogo principale, le variabili da *peso1* a *peso5* vengono usate per assegnare i cinque livelli del fattore *settimana*. La variabile del file di dati che raggruppa maschi e femmine (*sess*) può essere specificata come un fattore tra soggetti per lo studio delle differenze tra maschi e femmine.

**Misure.** Definire le misure se sui soggetti sono stati eseguiti test di più misure alla volta. Su ciascun soggetto è possibile, ad esempio, misurare il battito cardiaco e la respirazione ogni giorno per un'intera settimana. Queste misure non sono disponibili in forma di variabili nel file di dati, ma vengono definite in questa finestra di dialogo. Un modello che include più di una misura è a volte definito modello a misure ripetute doppiamente multivariato.

## GLM - Misure ripetute: Modello

Figura 3-3  
Finestra di dialogo Misure ripetute: Modello

Misure ripetute: Modello

Specifica modello

Fattoriale completo  Personalizzato

Entro-soggetti: tempo

Modello entro soggetti: tempo

Fra soggetti: sesso, età

Modello tra soggetti: sesso, età

Costruisci termine/i

Effetti principali

Somma dei quadrati: Tipo III

Continua Annulla Aiuto

**Specifica modello.** Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

**Tra soggetti.** I fattori tra soggetti e le covariate sono elencati.

**Modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti e le interazioni entro soggetti nonché gli effetti e le interazioni tra soggetti desiderati per l'analisi.

**Somma dei quadrati.** Metodo per il calcolo della somma dei quadrati del modello tra soggetti. Il metodo Somma dei quadrati usato più di frequente per i modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

### ***Costruisci termini***

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

### ***Somma dei quadrati***

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo predefinito, è quello usato più di frequente.

**Tipo I.** Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

**Tipo II.** Questo metodo consente di calcolare le somme dei quadrati di un effetto del modello corretto per tutti gli altri effetti “appropriati”. È considerato appropriato un effetto corrispondente a tutti gli effetti che non includono l’effetto in esame. Il metodo Somma dei quadrati Tipo II è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato.
- Qualsiasi modello che include solo effetti principali del fattore.
- Qualsiasi modello di regressione.
- Un disegno nidificato in modo puro. Questo tipo di nidificazione può essere specificato tramite la sintassi.

**Tipo III.** Tipo predefinito. Questo metodo consente di calcolare la somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

**Tipo IV.** Questo metodo è specifico per situazioni con celle mancanti. Per qualsiasi effetto  $F$  del disegno, se  $F$  non è incluso in nessun altro effetto, allora Tipo IV = Tipo III = Tipo II. Se invece  $F$  è incluso in altri effetti, con il Tipo IV i contrasti creati tra i parametri in  $F$  vengono distribuiti equamente tra tutti gli effetti di livello superiore. Il metodo Somma dei quadrati Tipo IV viene in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I e il Tipo II.
- Qualsiasi modello bilanciato e non bilanciato contenente celle vuote.

## GLM - Misure ripetute: Contrasti

Figura 3-4  
Finestra di dialogo Misure ripetute: Contrasti



I contrasti consentono di verificare le differenze tra i livelli di un fattore tra soggetti. È possibile specificare un contrasto per ciascun fattore tra soggetti del modello. I contrasti rappresentano combinazioni lineari dei parametri.

La verifica di ipotesi è basata sull'ipotesi nulla  $\mathbf{LBM}=0$ , dove  $\mathbf{L}$  è la matrice dei coefficienti di contrasto,  $\mathbf{B}$  è il vettore dei parametri e  $\mathbf{M}$  è la matrice media che corrisponde alla trasformazione media della variabile dipendente. Per visualizzare la matrice di trasformazione è possibile selezionare Matrice di trasformazione nella finestra di dialogo Misure ripetute: Opzioni. Se per i fattori entro soggetti sono disponibili, ad esempio, quattro variabili dipendenti, un fattore entro soggetti di quattro livelli e contrasti polinomiali (default), la matrice  $\mathbf{M}$  sarà (0,5 0,5 0,5 0,5)'. Quando si specifica un contrasto, viene creata una matrice  $\mathbf{L}$  tale che le colonne corrispondenti al fattore tra soggetti corrispondono al contrasto. Le altre colonne vengono corrette in modo che la matrice  $\mathbf{L}$  possa essere stimata.

Sono disponibili i contrasti deviazione, semplici, differenza, Helmert, ripetuti e polinomiali. Per i contrasti deviazione e i contrasti semplici, è possibile stabilire se la categoria di riferimento corrisponde alla prima o all'ultima categoria.

Per i fattori entro soggetti deve essere selezionato un contrasto diverso da Nessuno.

### **Tipi di contrasto**

**Deviazione.** Consente di confrontare la media di ciascun livello, a 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.

**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. Come categoria di riferimento, è possibile scegliere la prima o l'ultima categoria.

**Differenza.** Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Questo tipo di contrasto è a volte definito contrasto inverso di Helmert.

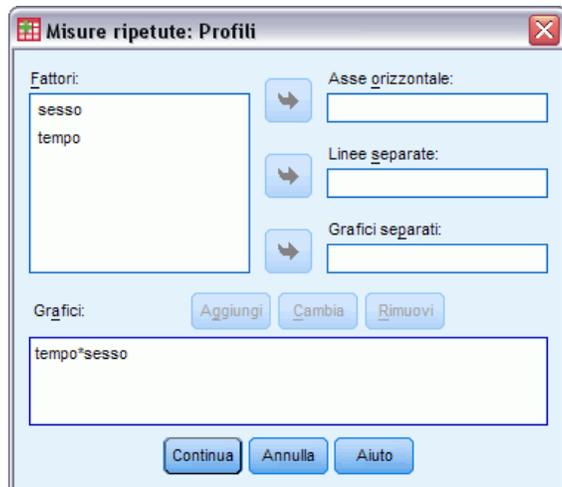
**Helmert.** Consente di confrontare la media di ciascun livello del fattore (a eccezione dell'ultimo) con la media dei livelli successivi.

**Ripetuto.** 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.

## GLM - Misure ripetute: Profili

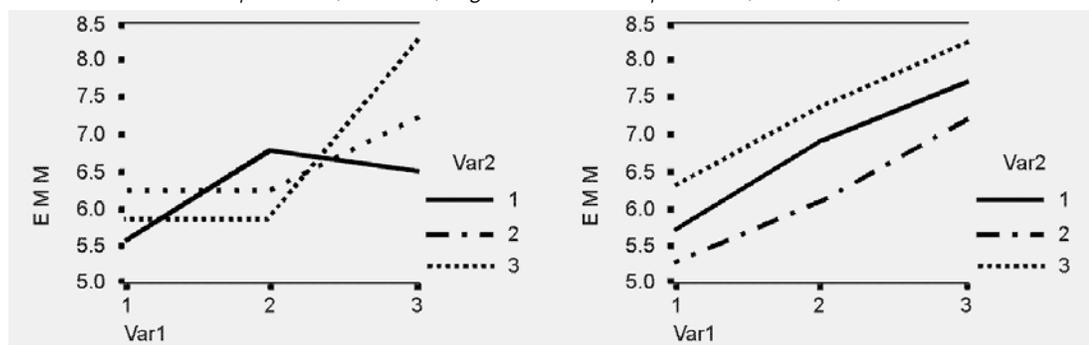
Figura 3-5  
Finestra di dialogo Misure ripetute: Profili



I profili, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un profilo è un grafico lineare in cui ciascun punto indica la media marginale stimata di una variabile dipendente (corretta per le covariate) in corrispondenza di un solo livello di un fattore. È possibile utilizzare i livelli di un secondo fattore per creare linee distinte. È possibile utilizzare ciascun livello di un terzo fattore per creare un grafico distinto. Tutti i fattori sono disponibili per i grafici. Questi grafici vengono creati per ciascuna variabile dipendente. Nei profili è possibile usare sia fattori tra soggetti che fattori entro soggetti.

Il profilo di un fattore mostra se le medie marginali stimate aumentano o diminuiscono tra i vari livelli. Nel caso di due o più fattori, le linee parallele indicano che tra i fattori non esiste alcuna interazione, ovvero che è possibile analizzare i livelli di un solo fattore. Le linee che si incrociano indicano invece che esiste un'interazione.

Figura 3-6  
Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)

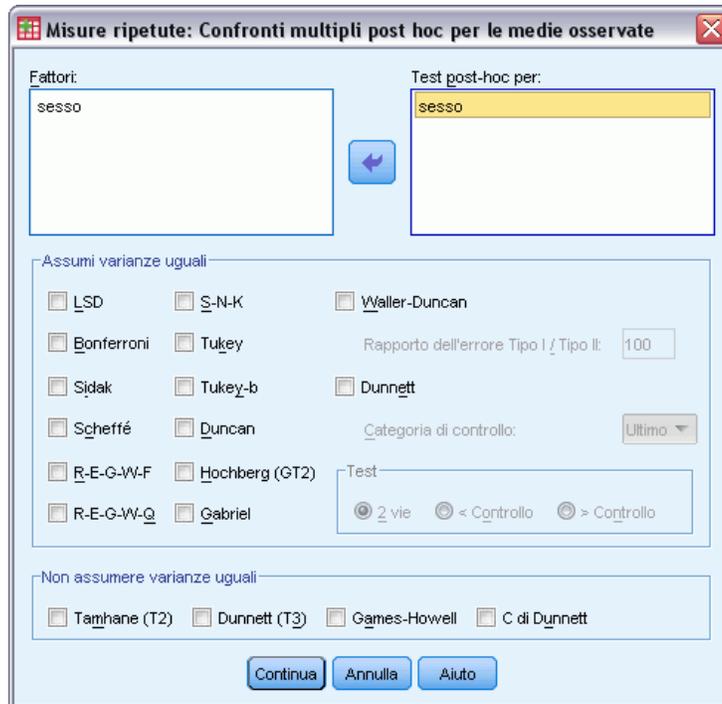


I grafici definiti tramite la selezione dei fattori per l'asse orizzontale e, se lo si desidera, dei fattori di linee e di grafici separati devono essere inclusi nell'elenco dei grafici.

## GLM - Confronti post hoc delle misure ripetute

Figura 3-7

Finestra di dialogo Confronti multipli post hoc delle misure ripetute per le medie osservate



**Test per confronti multipli post hoc.** Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervalli e i confronti a coppie multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stata apportata alcuna correzione. Questi test non sono disponibili se non sono presenti fattori tra soggetti e i test per confronti multipli post hoc vengono eseguiti per la media tra i livelli dei fattori entro soggetti.

I test per confronti multipli usati più di frequente sono il test di Bonferroni e il test delle differenze significative di Tukey. Il **test di Bonferroni**, basato sulla statistica  $t$  di Student, consente di correggere il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test  $t$  di Sidak** corregge inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test delle differenze significative di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti a coppie tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti per coppie. Quando si eseguono test su un elevato numero di coppie di medie, il test delle differenze significative di Tukey risulta più efficace rispetto al test di Bonferroni. Nel caso di un numero limitato di coppie, risulta invece più efficace il test di Bonferroni.

Il test di **Hochberg (GT2)** è simile al test delle differenze significative di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test dei confronti a coppie di Gabriel** utilizza il modulo massimo studentizzato ed è in genere più indicativo del test di Hochberg (GT2) quando le dimensioni delle celle sono diverse. Se la variabilità delle dimensioni delle celle risulta molto alta, il test di Gabriel può diventare poco conservativo.

Il **test t per confronti multipli a coppie di Dunnett** confronta un insieme di trattamenti con una media di controllo singola. L'ultima categoria è la categoria di controllo predefinita. In alternativa, è possibile scegliere la prima categoria. È inoltre possibile scegliere un test a 2 vie oppure a 1 via. Per verificare che la media in qualsiasi livello del fattore (a eccezione della categoria di controllo) non sia uguale a quella della categoria di controllo, è necessario usare un test a due sensi. Per verificare se la media di qualsiasi livello del fattore è minore di quella della categoria di controllo, selezionare  $<$  Controllo. In modo analogo, per verificare se la media di qualsiasi livello del fattore è maggiore di quella della categoria di controllo, selezionare  $>$  Controllo.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test a intervalli decrescenti multipli. Le procedure a multipli decrescenti verificano in primo luogo se tutte le medie sono uguali. Se le medie non risultano tutte uguali, il test di uguaglianza viene eseguito su un sottoinsieme di medie. Il test **R-E-G-W F** è basato su un test  $F$ , mentre **R-E-G-W Q** è basato sull'intervallo studentizzato. Questi test risultano più efficaci rispetto ai test a intervallo multiplo di Duncan e Student-Newman-Keuls, che sono pure procedure a intervalli decrescenti multipli. È tuttavia consigliabile non usarli con celle di dimensioni non uguali.

Quando le varianze non sono uguali, è necessario utilizzare il test **Tamhane (T2)** (test per confronti a coppie conservativo basato su un test  $t$ ), il test di **Dunnett T3** (test per confronti a coppie basato sul modulo studentizzato), il test per confronti a coppie di **Games-Howell** (a volte poco conservativo) o il test **C di Dunnett** (test per confronti a coppie basato sull'intervallo studentizzato).

Il **test a intervallo multiplo di Duncan**, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test per intervallo che classificano le medie raggruppate e calcolano un valore di intervallo. Questi test sono usati meno frequentemente dei test descritti in precedenza.

Il **test t di Waller-Duncan** utilizza un approccio bayesiano. Si tratta di un test a intervallo che usa la media armonica della dimensione campione nel caso di dimensioni campione non uguali.

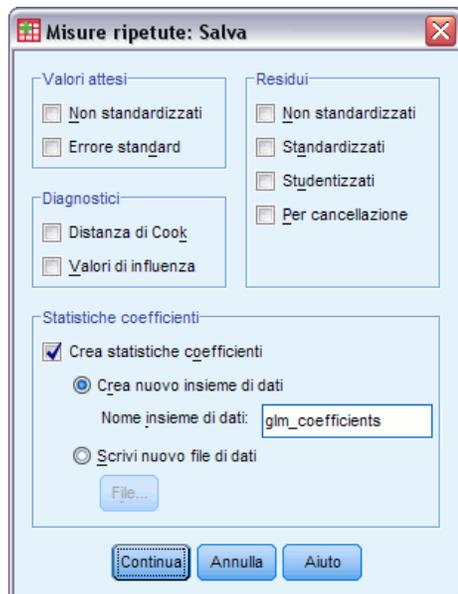
Il livello di significatività del test di **Scheffé** consente la verifica di tutte le possibili combinazioni lineari delle medie di gruppo e non dei soli confronti a coppie disponibili in questa funzione. Il test di Scheffé risulta pertanto più conservativo rispetto ad altri test, ovvero per ottenere un livello sufficiente di significatività, è richiesta una differenza maggiore tra le medie.

Il test per confronti a coppie multipli Differenza meno significativa o **LSD**, è equivalente a più test  $t$  tra tutte le coppie di gruppi. Lo svantaggio di questo test è che non viene eseguito alcun tentativo di correzione del livello di significatività osservata per confronti multipli.

**Test visualizzati.** I confronti a coppie sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), test  $C$  di Dunnett e Dunnett (T3). Per i test S-N-K,  $b$  di Tukey, Duncan, R-E-G-W  $F$ , R-E-G-W  $Q$  e Waller sono disponibili sottoinsiemi omogenei per test per intervallo. Il test delle differenze significative di Tukey, i test Hochberg (GT2), Gabriel e Scheffé sono sia test per confronti multipli che test a intervallo.

## GLM - Misure ripetute: Salva

Figura 3-8  
Finestra di dialogo Misure ripetute: Salva



È possibile salvare i valori attesi dal modello, le misure correlate e i residui come nuove variabili nell'Editor dei dati. Molte di queste variabili possono essere usate per l'esame di ipotesi sui dati. Per salvare i valori in modo da poterli usare in un'altra sessione IBM® SPSS® Statistics, è necessario salvare il file di dati corrente.

**Valori attesi.** Valori attesi dal modello per ciascun caso.

- **Non standardizzati.** I valori risultanti dal modello per la variabile dipendente e per ciascun caso.
- **Errore standard.** Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

**Diagnostici.** Misure per l'identificazione dei casi con combinazioni di valori insolite per le variabili indipendenti e dei casi che possono avere una notevole influenza sul modello. Sono disponibili la distanza di Cook e valori d'influenza non centrati.

- **Distanza di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione ne modificherebbe sostanzialmente il valore.
- **Valori di influenza.** Valori di influenza non centrati. Una misura dell'influenza di ciascun caso sull'adattamento di un modello di regressione.

**Residui.** Un residuo non standardizzato corrisponde al valore effettivo della variabile dipendente diminuito del valore atteso dal modello. Sono inoltre disponibili residui standardizzati, studentizzati ed eliminati.

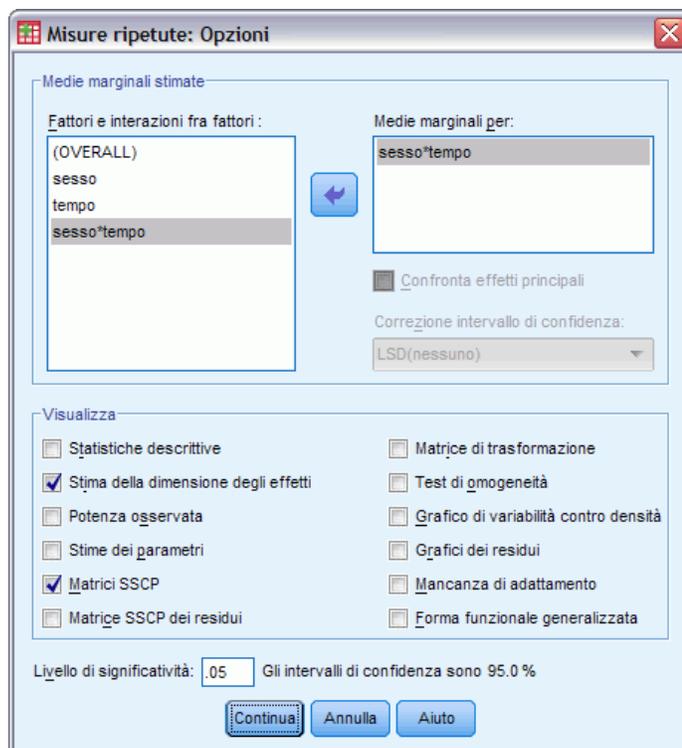
- **Non standardizzati.** La differenza tra un valore osservato e il valore stimato dal modello.

- **Standardizzati.** Il residuo diviso per una stima della deviazione standard. Il residuo standardizzato, conosciuto anche come residuo di Pearson, ha media 0 e deviazione standard 1.
- **Studentizzati.** Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza tra i valori assunti per questo caso dalle variabili indipendenti e le medie delle variabili indipendenti.
- **Per cancellazione.** Il residuo per un caso se quel caso venisse escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore stimato corretto.

**Statistiche dei coefficienti.** Consente di salvare una matrice di varianza-covarianza delle stime dei parametri in un file di dati. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime dei parametri, una riga di valori di significatività per le statistiche  $t$  corrispondenti alle stime dei parametri e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. Questi dati di matrici possono essere usati in altre procedure che eseguono la lettura di un file di matrici. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi degli insiemi di dati devono essere conformi alle regole dei nomi delle variabili.

## GLM - Misure ripetute: Opzioni

Figura 3-9  
Finestra di dialogo Misure ripetute: Opzioni



In questa finestra di dialogo sono disponibili statistiche opzionali. Le statistiche vengono calcolate tramite un modello di effetti fissi.

**Medie marginali stimate.** Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono corrette per le eventuali covariate. È possibile selezionare sia fattori entro soggetti che fattori tra soggetti.

- **Confronta effetti principali.** Consente di eseguire confronti a coppie senza correzione tra le medie marginali stimate di qualsiasi effetto principale del modello, per fattori sia tra soggetti che entro soggetti. Questa opzione è disponibile solo se nell'elenco Medie marginali per sono stati selezionati effetti principali.
- **Correzione intervallo di confidenza.** Selezionare la differenza meno significativa (LSD), la correzione di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo comando è disponibile solo se è stato selezionato Confronta effetti principali.

**Visualizzazione.** Selezionare Statistiche descrittive per produrre medie osservate, deviazioni standard e conteggi per tutte le variabili dipendenti di tutte le celle. La funzione Stima della dimensione degli effetti fornisce un valore eta-quadrato parziale per ciascun effetto e per ciascuna stima dei parametri. La statistica eta-quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore. Selezionare Potenza osservata per ottenere la potenza del test nel caso in cui l'ipotesi alternativa sia basata sul valore osservato. Selezionare Stime dei parametri per ottenere stime dei parametri, errori standard, test  $t$ , intervalli di confidenza e la potenza osservata per ciascun test. È possibile visualizzare le matrici SSCP di ipotesi ed errore e la matrice SSCP dei residui, nonché il test della sfericità di Bartlett della matrice di covarianza dei residui.

La funzione Test di omogeneità produce il test di Levene per l'omogeneità della varianza per ogni variabile dipendente su tutte le combinazioni di livello dei fattori fra soggetti, solo per i fattori fra soggetti. È un test di omogeneità anche il test  $M$  di Box dell'omogeneità delle matrici di covarianza delle variabili dipendenti in tutte le possibili combinazioni di livello dei fattori tra soggetti. Le opzioni Grafici di variabilità vs. densità e Grafici dei residui risultano utili per la verifica di ipotesi sui dati. Se non è disponibile alcun fattore, questa opzione risulta disattivata. Selezionare Grafici dei residui per ottenere un grafico dei residui osservati, attesi e standardizzati per ciascuna variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare Mancanza di adattamento per controllare che la relazione fra la variabile dipendente e le variabili indipendenti possa essere descritta in modo adeguato dal modello. La funzione Forma funzionale generalizzata consente di creare test di ipotesi personalizzati basati sulla forma funzionale generalizzata. Le righe di una matrice dei coefficienti di contrasto sono combinazioni lineari della forma funzionale generalizzata.

**Livello di significatività.** Potrebbe risultare utile correggere il livello di significatività usato nei test post hoc e il livello di confidenza usato per la costruzione degli intervalli di confidenza. Il valore specificato viene inoltre usato per il calcolo della potenza osservata per il test. Quando si specifica un livello di significatività, nella finestra di dialogo viene visualizzato il livello di intervalli di confidenza associato.

## **Opzioni aggiuntive del comando GLM**

Queste opzioni sono valide per analisi univariate, multivariate e a misure ripetute. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati del disegno (tramite il sottocomando `DESIGN`).
- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando `TEST`).
- Specificare contrasti multipli (tramite il sottocomando `CONTRAST`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).
- Specificare criteri EPS (tramite il sottocomando `CRITERIA`).
- Costruire una matrice **L**, **M** o **K** personalizzata (tramite il sottocomando `LMATRIX`, `MMATRIX` o `KMATRIX`).
- Per i contrasti deviazione e i contrasti semplici, specificare una categoria di riferimento intermedia (tramite il sottocomando `CONTRAST`).
- Specificare metrica per contrasti polinomiali (tramite il sottocomando `CONTRAST`).
- Specificare termini di errore per confronti post-hoc (tramite il sottocomando `POSTHOC`).
- Calcolare medie marginali stimate per qualsiasi fattore o interazione tra fattori per i fattori elencati (tramite il sottocomando `EMMEANS`).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando `SAVE`).
- Costruire un file di dati di matrici di correlazione (tramite il sottocomando `OUTFILE`).
- Costruire un file di dati di matrici contenente statistiche derivate dai dati della tabella ANOVA tra soggetti (tramite il sottocomando `OUTFILE`).
- Salvare la matrice del disegno in un nuovo file di dati (tramite il sottocomando `OUTFILE`).

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

# ***Analisi Componenti della varianza***

La procedura Componenti della varianza, per modelli con effetti misti, consente di stimare il contributo di ciascun effetto casuale in relazione alla varianza della variabile dipendente. Questa procedura è particolarmente utile per l'analisi di modelli misti, ad esempio disegni a misure ripetute univariate, a split-plot e di blocchi casuali. Il calcolo dei componenti della varianza consente di stabilire il punto da considerare per la riduzione della varianza.

Sono disponibili quattro metodi diversi per stimare i componenti della varianza: MINQUE (Minimum Nomr Quadratic Unbiased estimator), ANOVA (analisi della varianza), ML (massima verosimiglianza) e REML (massima verosimiglianza ristretta). A ciascun metodo sono associate specifiche diverse.

L'output predefinito di tutti i metodi include le stime dei componenti della varianza. Se si utilizza il metodo ML o REML, verrà inoltre visualizzata una tabella delle matrici di covarianza. L'output del metodo ANOVA può inoltre includere una tabella ANOVA e la media dei quadrati previsti, mentre l'output dei metodi ML e REML può includere una cronologia delle iterazioni. La procedura Componenti della varianza è del tutto compatibile con la procedura GLM - Univariata.

Minimi quadrati ponderati consente di specificare una variabile per l'assegnazione di pesi diversi alle osservazioni in un'analisi ponderata, in alcuni casi per compensare le variazioni nella precisione della misura.

**Esempio.** In una scuola agricola viene misurato l'aumento di peso dei maiali di sei diverse lettiera nel corso di un mese. La variabile lettiera è un fattore casuale a sei livelli. Le sei lettiera analizzate rappresentano un campione casuale di una numerosa popolazione di lettiera per maiali. Dallo studio risulta che la varianza dell'aumento di peso è attribuibile alla differenza tra lettiera in misura maggiore rispetto alla differenza tra i maiali di una lettiera.

**Dati.** La variabile dipendente è quantitativa. I fattori sono categoriali. Vi possono essere associati valori numerici o valori stringa composti da un massimo di otto byte. Almeno uno dei fattori deve essere casuale. Ovvero, i livelli del fattore devono essere un campione casuale di possibili livelli. Le covariate sono variabili quantitative correlate alla variabile dipendente.

**Assunzioni.** In tutti i metodi si presume che ai parametri del modello di un effetto casuale siano associate medie pari a zero e varianze di costanti finite e che i parametri non siano correlati. La mancanza di correlazione si riscontra anche nei parametri del modello derivati da effetti casuali diversi.

Il termine residuo ha anch'esso una media pari a zero e una varianza costante e finita. Non è correlato con i parametri di modello di alcun effetto casuale. Si presume che i termini residui di osservazioni diverse non siano correlati.

In base a queste ipotesi, le osservazioni del medesimo livello di un fattore casuale sono correlate. Ciò distingue un modello di componenti della varianza da un modello lineare generalizzato.

I metodi ANOVA e MINQUE non richiedono ipotesi di normalità. Consentono entrambi scostamenti irrilevanti dall'ipotesi di normalità.

Per poter distribuire in modo normale i metodi ML e REML, è necessario specificare il parametro del modello e il termine residuo.

**Procedure correlate.** Per esaminare i dati prima di eseguire l'analisi dei componenti della varianza, usare la procedura Esplora. Per la verifica delle ipotesi, usare le procedure GLM - Univariata, GLM - Multivariata e GLM - Misure ripetute.

### **Per ottenere un'analisi Componenti della varianza**

- Dai menu, scegliere:  
Analizza > Modello lineare generalizzato > Componenti della varianza...

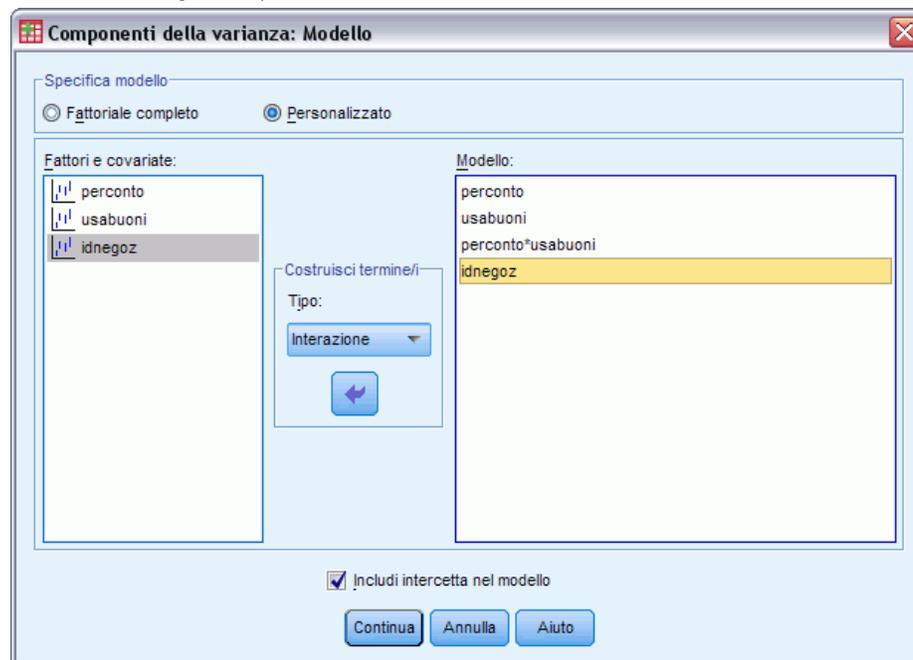
Figura 4-1  
Finestra di dialogo Componenti della varianza



- Selezionare una variabile dipendente.
- Selezionare le variabili per l'opzione Fattori fissi, Fattori casuali o Covariate, a seconda dei dati in uso. Per specificare una variabile di ponderazione, è necessario usare l'opzione Minimi quadrati ponderati.

## Componenti della varianza: Modello

Figura 4-2  
Finestra di dialogo Componenti della varianza: Modello



**Specifica modello.** Un modello fattoriale completo contiene tutti gli effetti principali dei fattori e delle covariate e tutte le interazioni fattore per fattore. Non contiene interazioni di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

**Fattori e covariate.** I fattori e le covariate sono elencati.

**Modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire. Il modello deve includere almeno un fattore casuale.

**Includi l'intercetta nel modello.** In genere l'intercetta viene inclusa nei modelli. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

### Costruisci termini

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

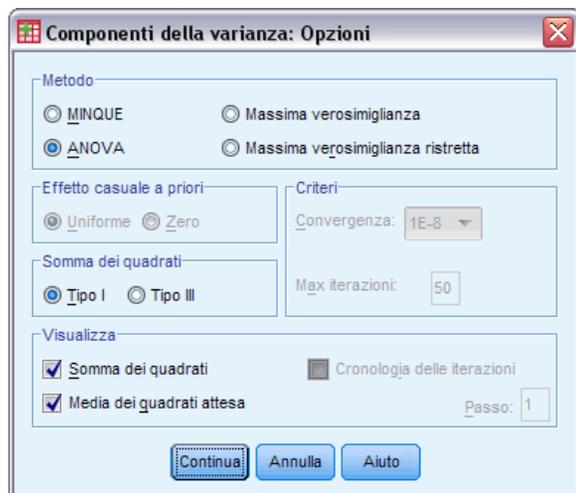
**Tutti 3-vie.** Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

**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.

## Componenti della varianza: Opzioni

Figura 4-3  
Finestra di dialogo Componenti della varianza: Opzioni



**Metodo.** Per la stima dei componenti della varianza sono disponibili quattro metodi.

- Il metodo MINQUE (Minimum Norm Quadratic Unbiased Estimator) consente di ottenere stime che non variano in relazione agli effetti fissi. Se i dati sono distribuiti normalmente e le stime sono corrette, con questo metodo si ottiene la varianza minore tra tutti gli stimatori non parziali. È possibile scegliere un metodo per i pesi degli effetti casuali.
- Il metodo ANOVA (analisi della varianza) consente di calcolare stime non distorte tramite somme dei quadrati Tipo I o Tipo III per ciascun effetto. A volte si ottengono stime di varianza negative, che indicano la presenza di un modello non corretto, l'uso di un metodo di stima non appropriato o un'insufficienza di dati.
- Il metodo Massima verosimiglianza (ML) consente di ottenere le stime che risultano più conformi ai dati effettivamente osservati tramite l'uso di iterazioni. Tali stime possono non essere imparziali. Questo metodo è asintoticamente normale. Le stime ML e REML non variano nel processo di conversione. Con questo metodo non vengono presi in considerazione i diversi gradi di libertà usati per la stima degli effetti fissi.
- Il metodo Massima verosimiglianza ristretta (REML) consente di ridurre le stime ANOVA per molti casi di dati bilanciati, se non per tutti. Inoltre, essendo corretto in funzione degli effetti fissi, questo metodo dovrebbe generare un numero di errori standard minore rispetto alla stima ML. Con questo metodo vengono presi in considerazione i diversi gradi di libertà usati per la stima degli effetti fissi.

**Pesi degli effetti casuali.** Uniforme implica che tutti gli effetti casuali e il termine residuo hanno un uguale impatto sulle osservazioni. Lo schema Zero equivale a presumere varianze di effetti casuali pari a zero. Disponibile solo con il metodo MINQUE.

**Somma dei quadrati.** Le somme dei quadrati Tipo I vengono adottate per il modello gerarchico, usato spesso per i componenti della varianza. Se si sceglie il Tipo III, ovvero l'impostazione predefinita in GLM, sarà possibile usare le stime di varianza nella procedura GLM - Univariata per la verifica di ipotesi con somme dei quadrati Tipo III. Disponibile solo con il metodo ANOVA .

**Criteri.** È possibile specificare il criterio di convergenza e il numero massimo di iterazioni. Disponibile solo per i metodi ML o REML.

**Visualizzazione.** Con il metodo ANOVA è possibile visualizzare le somme dei quadrati e le medie dei quadrati attese. Se è stato selezionato il metodo Massima verosimiglianza o Massima verosimiglianza ristretta, è possibile visualizzare una cronologia delle iterazioni.

### ***Somma dei quadrati (Componenti della varianza)***

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo predefinito, è quello usato più di frequente.

**Tipo I.** Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per i termini del modello che lo precedono. Il metodo Somma dei quadrati Tipo I viene in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

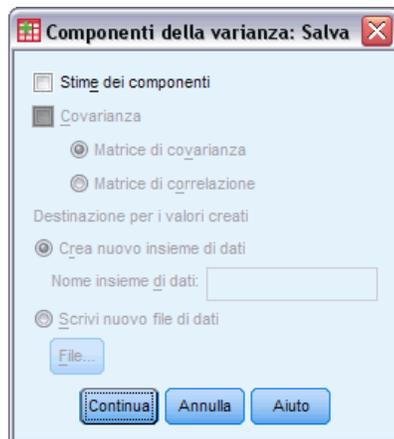
**Tipo III.** Tipo predefinito. Questo metodo consente di calcolare la somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato spesso utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

## Componenti della varianza: Salva in un nuovo file

Figura 4-4

Finestra di dialogo Componenti della varianza: Salva in un nuovo file



Alcuni dei risultati di questa procedura possono essere salvati in un nuovo file di dati di IBM® SPSS® Statistics.

**Stime dei componenti (Componenti della varianza).** Consente di salvare le stime dei componenti della varianza e le etichette delle stime in un file di dati. Questi dati possono quindi essere usati per il calcolo di altre statistiche o per altre analisi nelle procedure GLM. È possibile, ad esempio, usarli per il calcolo degli intervalli di confidenza o per la verifica di ipotesi.

**Covarianza.** Consente di salvare una matrice di varianza-covarianza o una matrice di correlazione in un file di dati o insieme di dati. È disponibile solo se è stato selezionato il metodo Massima verosimiglianza o Massima verosimiglianza ristretta.

**Destinazione per i valori creati.** Consente di specificare un nome per il file di dati o un nome di file esterno per il file contenente le stime dei componenti della varianza e/o la matrice. I file di dati possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali alla fine della sessione. I nomi degli insiemi di dati devono essere conformi alle regole dei nomi delle variabili.

Tramite il comando `MATRIX` è possibile estrarre i dati desiderati dal file di dati e calcolare quindi gli intervalli di confidenza o eseguire test.

## Opzioni aggiuntive del comando `VARCOMP`

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati del disegno (tramite il sottocomando `DESIGN`).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando `MISSING`).
- Specificare criteri EPS (tramite il sottocomando `CRITERIA`).

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

## ***Modelli misti lineari***

La procedura Modelli misti lineari consente di espandere il modello lineare generale affinché i dati mostrino una variabilità correlata e non costante. La flessibilità del modello misto lineare, consente dunque di modellare, oltre al significato dei dati, anche le relative varianze e covarianze.

La procedura Modelli misti lineari è inoltre uno strumento flessibile per l'adattamento di modelli formulati come modelli misti lineari. Tali modelli includono i modelli a più livelli, i modelli lineari gerarchici e i modelli a coefficiente casuale.

**Esempio.** Poniamo il caso di una catena di drogherie che sia interessata a conoscere gli effetti di diversi coupon sulla spesa dei clienti. Considerando un campione casuale dei propri clienti abituali, potranno seguire l'andamento della spesa di ciascun cliente nell'arco di dieci settimane. Ogni settimana verrà inviato ai clienti un coupon diverso. La funzionalità Modelli misti lineari consentirà di stimare l'effetto di coupon diversi sulla spesa e di adattare al tempo stesso la correlazione in base alle osservazioni ripetute su ciascun oggetto per dieci settimane.

**Metodi.** Stima della massima verosimiglianza (ML) e della massima verosimiglianza ristretta (REML).

**Statistiche.** Statistiche descrittive: dimensioni campione; medie e deviazione standard della variabile dipendente, covariate per ciascuna combinazione di livello dei fattori. Informazioni a livello di fattore: valori ordinati dei livelli di ciascun fattore e delle relative frequenze. Stima dei parametri e intervalli di confidenza per gli effetti fissi, test di Wald e intervalli di confidenza per i parametri delle matrici di covarianza. Per la valutazione di ipotesi diverse è possibile utilizzare la somma dei quadrati Tipo I e Tipo III. Il metodo predefinito è il Tipo III.

**Dati.** La variabile dipendente deve essere quantitativa. I fattori devono essere categoriali ed essere associati a valori numerici o a valori stringa. Le covariate e la variabile di ponderazione devono essere quantitative. I soggetti e le variabili ripetute possono essere di qualsiasi tipo.

**Assunzioni.** Si presume che la variabile dipendente sia correlata in modo lineare ai fattori fissi, ai fattori casuali e alle covariate. Gli effetti fissi determinano la media della variabile dipendente. Gli effetti casuali determinano la struttura della covarianza della variabile dipendente. Gli effetti casuali multipli vengono considerati indipendenti l'uno dall'altro e per ciascuno di essi vengono calcolate matrici di covarianza distinte. I termini del modello specificati per lo stesso effetto casuale possono tuttavia essere correlati. Il modello a misure ripetute determina la struttura della covarianza dei residui. Si presume inoltre che la variabile dipendente provenga da una distribuzione normale.

**Procedure correlate.** Per esaminare i dati prima di eseguire un'analisi, usare la procedura Esplora. Se si ritiene che non esista una variabilità correlata o non costante, è possibile usare le procedure GLM - Univariata o GLM - Misure ripetute. In alternativa, se gli effetti casuali presentano una

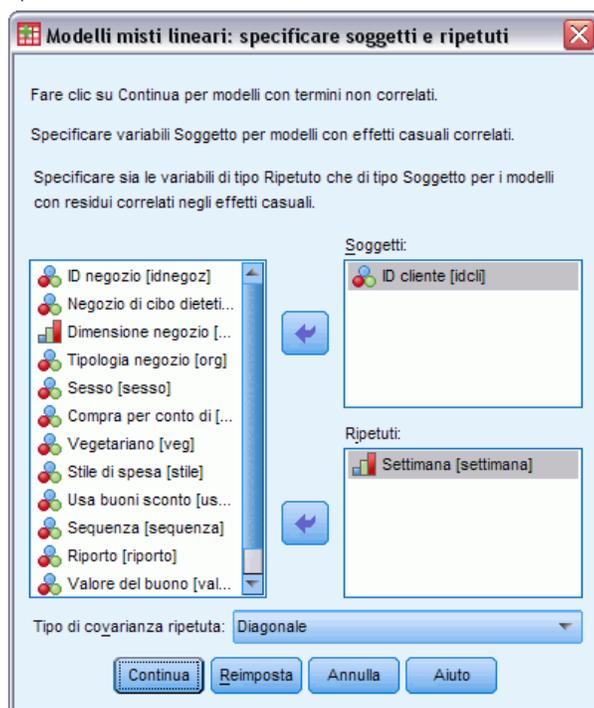
struttura di covarianza con componenti di varianza e non esistono misure ripetute, è possibile utilizzare la procedura di analisi dei componenti della varianza.

### **Per ottenere un'analisi dei modelli misti lineari**

- Dai menu, scegliere:  
Analizza > Modelli misti > Lineare...

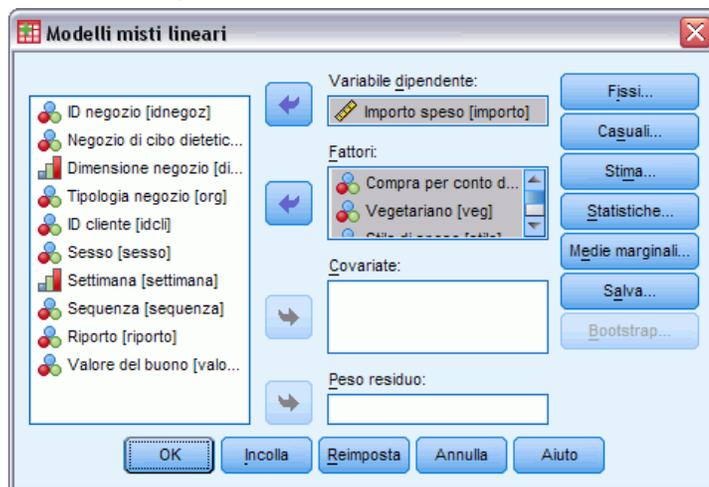
Figura 5-1

*Finestra di dialogo Modelli misti lineari - Selezione Variabili di definizione di soggetti e osservazioni ripetute*



- È possibile selezionare una o più variabili dei soggetti.
- È possibile selezionare una o più variabili di osservazioni ripetute.
- È possibile selezionare una struttura della covarianza residua.
- Fare clic su Continua.

Figura 5-2  
Finestra di dialogo Modelli misti lineari



- ▶ Selezionare una variabile dipendente.
- ▶ Selezionare almeno un fattore o una covarianza.
- ▶ Fare clic su Fisso o Casuale, quindi specificare almeno un modello degli effetti fissi o casuali.  
Se necessario, selezionare una variabile di ponderazione.

## ***Modelli misti lineari: specificare soggetti e ripetuti***

Questa finestra di dialogo consente di selezionare le variabili che definiscono i soggetti e le osservazioni ripetute e di scegliere una struttura di covarianza per i residui. Vedere [Figura 5-1](#) a pag. 37.

**Soggetti.** Un soggetto è un'unità di osservazione che può essere considerata indipendente rispetto ad altri soggetti. Ad esempio, i risultati della misurazione della pressione sanguigna di un paziente in uno studio medico sono da considerarsi indipendenti rispetto alle misurazioni effettuate sugli altri pazienti. La definizione dei soggetti è particolarmente importante nel caso in cui esistano misurazioni ripetute per ciascun soggetto e si desideri determinare una correlazione tra queste osservazioni. Ad esempio, è possibile stabilire se esiste una correlazione tra le misurazioni della pressione sanguigna di un paziente rilevate nel corso di più visite mediche.

I soggetti possono inoltre essere definiti in base alla combinazione livello/fattore di più variabili. Ad esempio, è possibile specificare *Sesso* e *Categoria età* come variabili soggetto per definire una situazione di questo tipo: le variabili *uomini di età superiore ai 65 anni* sono simili tra loro, ma indipendenti dalle variabili *uomini di età inferiore ai 65 anni* e *donne*.

Tutte le variabili specificate nell'elenco Soggetti vengono utilizzate per definire i soggetti per la struttura della covarianza residua. Per definire i soggetti per la struttura della covarianza ad effetti casuali è possibile utilizzare solo alcune o tutte le variabili.

**Ripetuto.** Le variabili specificate in questo elenco vengono utilizzate per identificare le osservazioni ripetute. Ad esempio, la variabile singola *Settimana* potrà identificare le 10 settimane di osservazione in uno studio medico, mentre le variabili *Mese* e *Giorno* utilizzate contemporaneamente potranno identificare le osservazioni giornaliere nel corso di un anno.

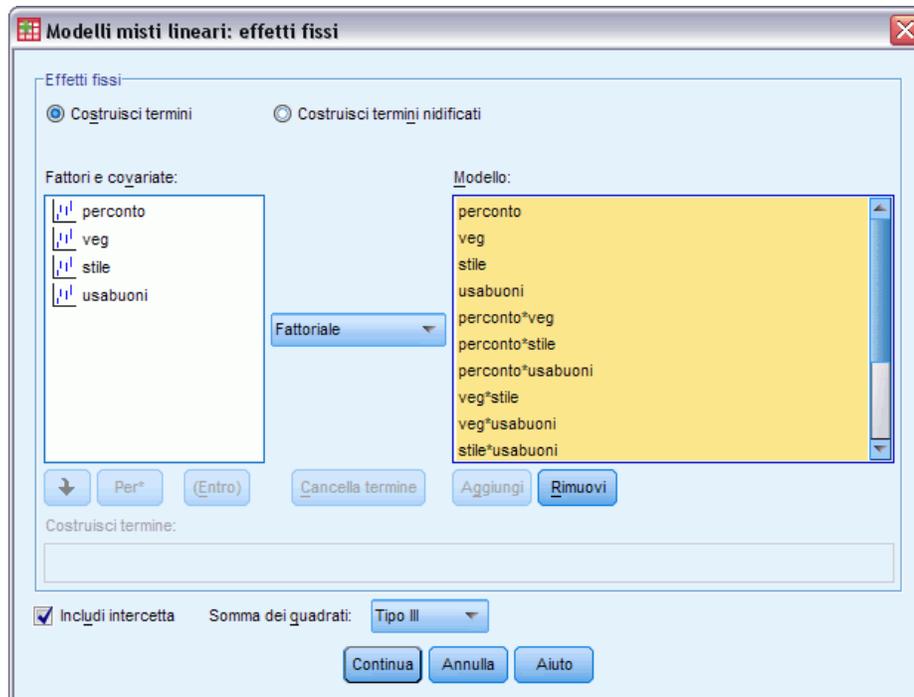
**Tipo di covarianza ripetuta.** Questo tipo di covarianza specifica la struttura della covarianza per i residui. Le strutture disponibili sono:

- Ante dipendenza: primo ordine
- AR(1)
- AR(1): eterogenea
- ARMA(1,1)
- Simmetria composta
- Simmetria composta: metrica di correlazione
- Simmetria composta: eterogenea
- Diagonale
- Analitica fattoriale: primo ordine
- Analitica fattoriale: primo ordine, eterogenea
- Huynh-Feldt
- Identità scalata
- Toeplitz
- Toeplitz: eterogenea
- Senza struttura
- Non strutturata: Correlazioni

Per ulteriori informazioni, vedere l'argomento Strutture della covarianza in l'appendice B a pag. 161.

## Effetti fissi nei modelli misti lineari

Figura 5-3  
Finestra di dialogo Modelli misti lineari: effetti fissi



**Effetti fissi.** Non esiste alcun modello predefinito, pertanto è necessario specificare esplicitamente gli effetti fissi. In alternativa, è possibile creare termini nidificati o non nidificati.

**Includi 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.

**Somma dei quadrati.** Metodo per il calcolo della somma dei quadrati. Per i modelli senza celle mancanti, il metodo Tipo III è quello di uso più comune.

### Per costruire termini non nidificati

Per i fattori e le covariate selezionati:

**Fattoriale.** Consente di creare tutte le possibili interazioni e gli effetti principali delle variabili selezionate. È l'impostazione predefinita.

**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.

### **Costruisci 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 l'andamento della 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 inoltre possibile includere effetti di interazione o 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  $A(X)$ .

### **Somma dei quadrati**

Per il modello è possibile scegliere un tipo di somma dei quadrati. Il Tipo III, il tipo predefinito, è quello usato più di frequente.

**Tipo I.** Questo metodo è definito anche scomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene corretto solo per il termine del modello che lo precede. Il metodo Somma dei quadrati Tipo I è in genere usato con i seguenti elementi:

- Un modello ANOVA bilanciato in cui gli effetti principali vengono specificati prima degli effetti di interazione di ordine 1, ciascuno dei quali viene a sua volta specificato prima degli effetti di interazione di ordine 2 e così via.
- Un modello di regressione polinomiale in cui qualsiasi termine di ordine più basso è specificato prima dei termini di ordine più elevato.
- Un modello nidificato in modo puro in cui il primo effetto specificato è nidificato nel secondo, il quale è a sua volta nidificato nel terzo e così via. Questo tipo di nidificazione può essere specificato esclusivamente tramite la sintassi.

**Tipo III.** Tipo predefinito. Questo metodo consente di calcolare la somma dei quadrati di un effetto del disegno come la somma dei quadrati corretta per qualsiasi altro effetto che non lo include e ortogonale rispetto agli eventuali effetti che lo contengono. Il vantaggio associato a questo tipo di somme dei quadrati è che non varia al variare delle frequenze di cella, a condizione che la forma generale di stimabilità rimanga costante. È pertanto considerato utile per modelli non bilanciati privi di celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale

alla tecnica dei quadrati delle medie ponderate di Yates. Il metodo Somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- I modelli elencati per il Tipo I.
- Qualsiasi modello bilanciato o non bilanciato e privo di celle vuote.

## ***Effetti casuali nei modelli misti lineari***

Figura 5-4  
Finestra di dialogo Modelli misti lineari: effetti casuali

**Tipo di covarianza.** Consente di specificare la struttura della covarianza per il modello degli effetti casuali. Per ogni effetto casuale viene stimata una matrice di covarianza distinta. Le strutture disponibili sono:

- Ante dipendenza: primo ordine
- AR(1)
- AR(1): eterogenea
- ARMA(1,1)
- Simmetria composta
- Simmetria composta: metrica di correlazione

- Simmetria composta: eterogenea
- Diagonale
- Analitica fattoriale: primo ordine
- Analitica fattoriale: primo ordine, eterogenea
- Huynh-Feldt
- Identità scalata
- Toeplitz
- Toeplitz: eterogenea
- Senza struttura
- Non strutturata: metrica di correlazione
- Componenti della varianza

Per ulteriori informazioni, vedere l'argomento [Strutture della covarianza in l'appendice B a pag. 161](#).

**Effetti casuali.** Non esiste alcun modello predefinito ed è pertanto necessario specificare esplicitamente gli effetti casuali. In alternativa, è possibile creare termini nidificati o non nidificati. È inoltre possibile includere un termine intercetta nel modello degli effetti casuali.

È possibile specificare più modelli per gli effetti casuali. Dopo aver costruito il primo modello, fare clic su **Successivo** per creare il modello successivo. Se si desidera tornare indietro e scorrere i modelli esistenti, fare clic su **Precedente**. Si presume che ciascun modello degli effetti casuali sia indipendente dagli altri. Ciò significa che per ciascuno di essi vengono calcolate matrici di covarianza distinte. I termini specificati nello stesso modello di effetti casuali possono essere correlati.

**Raggruppamenti di soggetti.** Le variabili elencate sono quelle selezionate come variabili soggetto nella finestra di dialogo **Modelli misti lineari**: specificare soggetti e ripetuti. Sceglierne alcune o tutte, per definire i soggetti per il modello degli effetti casuali.

## Stima dei modelli misti lineari

Figura 5-5  
Finestra di dialogo Modelli misti lineari: stima

**Metodo.** Selezionare la stima della massima verosimiglianza o della massima verosimiglianza ristretta.

### Iterazioni:

- **Max iterazioni.** Specificare un intero non negativo.
- **Dimezzamenti massimi.** 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.
- **Stampa cronologia iterazioni per ogni n passi.** Consente di visualizzare una tabella contenente il valore della funzione di verosimiglianza log e le stime dei parametri per ciascuna iterazione  $n$  che inizia con l'iterazione  $0^{\circ}$  (stime iniziali). Quando si stamperà la cronologia dell'iterazione, verrà stampata sempre l'ultima iterazione, a prescindere dal valore  $n$ .

**Convergenza verosimiglianza.** La convergenza viene presunta se la modifica assoluta o la modifica relativa della funzione del logaritmo di verosimiglianza sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è 0.

**Convergenza dei parametri.** La convergenza viene presunta se la modifica assoluta massima o la modifica relativa massima delle stime dei parametri sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è 0.

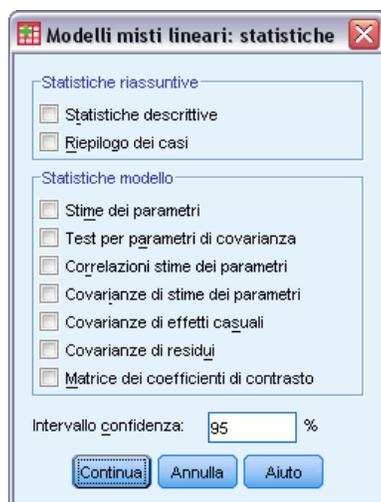
**Convergenza hessiana.** Per la specifica Assoluta, la convergenza viene presunta se la statistica basata sulla convergenza hessiana è inferiore al valore specificato. Per la specifica Relativa, la convergenza viene presunta se la statistica è inferiore al prodotto del valore specificato e del valore assoluto della verosimiglianza. Il criterio non viene utilizzato se il valore specificato è 0.

**Max passi di punteggio.** È necessario utilizzare l' algoritmo del punteggio Fisher fino al numero di iterazione  $n$ . Specificare un intero positivo.

**Tolleranza della singolarità.** Questo valore viene utilizzato come tolleranza per il controllo della singolarità. Specificare un valore positivo.

## Statistiche per modelli misti lineari

Figura 5-6  
Finestra di dialogo Modelli misti lineari: statistiche



**Statistiche riassuntive.** Consente di creare tabelle per:

- **Statistiche descrittive.** Se specificata, consente di visualizzare dimensioni campione, medie e deviazioni standard della variabile dipendente e delle covariate (se specificate). Le statistiche vengono visualizzate per ciascuna combinazione di livello dei fattori.
- **Riepilogo dei casi.** Consente di visualizzare i valori dei fattori ordinati, le variabili delle misure ripetute, gli oggetti delle misure ripetute, nonché i soggetti con effetti casuali con relative frequenze.

**Statistiche modello.** Consente di creare tabelle per:

- **Stime dei parametri.** Consente di visualizzare le stime di parametro per effetti casuali e fissi con gli errori standard approssimati.
- **Test per parametri di covarianza.** Consente di visualizzare gli errori standard asintotici e i test di Wald per i parametri della covarianza.

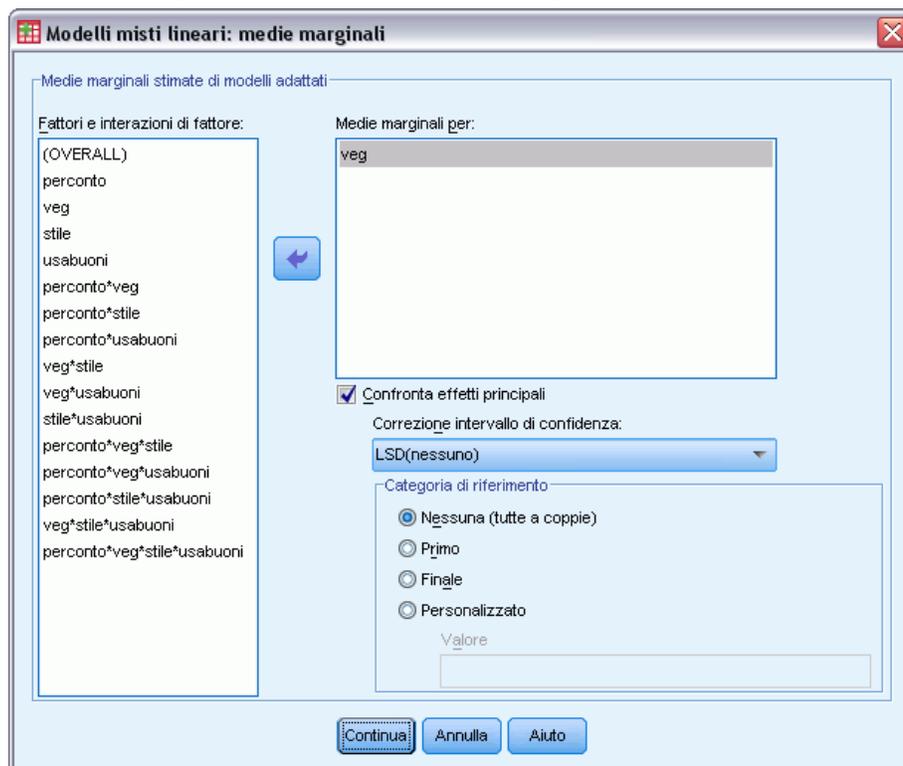
- **Correlazioni di stime di parametri.** Consente di visualizzare la matrice di correlazione asintotica delle stime dei parametri degli effetti fissi.
- **Covarianze di stime di parametri.** Consente di visualizzare la matrice di covarianza asintotica delle stime del parametro degli effetti fissi.
- **Covarianze di effetti casuali.** Consente di visualizzare la matrice di covarianza stimata degli effetti casuali. Questa opzione è disponibile solo se è stato specificato almeno un effetto casuale. Se per un effetto casuale è stata specificata una variabile soggetto, verrà visualizzato il blocco comune.
- **Covarianze di residui.** Consente di visualizzare la matrice di covarianza residua stimata. Questa opzione è disponibile solo se è stata specificata una variabile ripetuta. Se è stata specificata una variabile soggetto, verrà visualizzato il blocco comune.
- **Matrice dei coefficienti di contrasto.** Questa opzione consente di visualizzare le funzioni stimabili utilizzate per il test degli effetti fissi e delle ipotesi personalizzate.

**Intervallo di confidenza.** Questo valore viene utilizzato ogni volta che si costruisce un intervallo di confidenza. Specificare un valore maggiore o uguale a 0 e minore di 100. Il valore predefinito è 95.

## Modelli misti lineari: medie marginali

Figura 5-7

Finestra di dialogo Modelli misti lineari: medie marginali

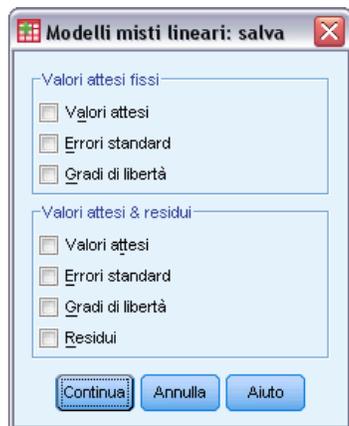


**Medie marginali stimate di modelli adattati.** Questo gruppo consente di richiedere le medie marginali stimate previste in base a un modello per la variabile dipendente nelle celle e i relativi errori standard per i fattori specificati. È in genere possibile richiedere il confronto dei livelli di fattore degli effetti principali.

- **Fattori e interazioni tra fattori.** Questo elenco contiene i fattori e le interazioni tra fattori specificati nella finestra di dialogo Fisso, oltre a un termine OVERALL. L'elenco non include i termini del modello creati da covariate.
- **Medie marginali per.** La procedura calcola le medie marginali stimate per i fattori e le interazioni tra fattori selezionati nell'elenco. Se è selezionato OVERALL, vengono visualizzate le medie marginali stimate della variabile dipendente e accorpati tutti i fattori. Si noti che tutti i fattori o le interazioni tra fattori selezionati rimangono selezionati, a meno che una variabile associata non sia stata rimossa dall'elenco Fattori della finestra di dialogo principale.
- **Confronta effetti principali.** Questa opzione consente di richiedere confronti a coppie dei livelli degli effetti principali selezionati. La correzione dell'intervallo di confidenza consente di applicare una correzione agli intervalli di confidenza e ai valori di significatività per tenere conto di confronti multipli. I metodi disponibili sono: LSD (nessuna correzione), Bonferroni e Sidak. Per ciascun fattore è infine possibile selezionare una categoria di riferimento in base alla quale effettuare i confronti. Se non si seleziona una categoria di riferimento, verranno creati tutti i confronti a coppie. Le opzioni per la categoria di riferimento sono la prima, l'ultima o una personalizzata. In quest'ultimo caso, inserire il valore della categoria di riferimento.

## Modelli misti lineari: salva

Figura 5-8  
Finestra di dialogo Modelli misti lineari: salva



Questa finestra di dialogo consente di salvare diversi risultati del modello nel file di lavoro.

**Valori attesi fissi.** Salva le variabili correlate alle medie di regressione senza gli effetti.

- **Valori attesi.** Medie di regressione senza gli effetti casuali.
- **Errori standard.** Errori standard delle stime.
- **Gradi di libertà.** Gradi di libertà associati alle stime.

**Valori attesi e residui.** Salva le variabili correlate al valore adattato al modello.

- **Valori attesi.** Valore adattato al modello.
- **Errori standard.** Errori standard delle stime.
- **Gradi di libertà.** Gradi di libertà associati alle stime.
- **Residui.** Valore dei dati meno il valore atteso.

## ***Opzioni aggiuntive del comando MIXED***

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare test di effetti vs. una combinazione lineare di effetti o un valore (tramite il sottocomando TEST).
- Includere valori mancanti definiti dall'utente (tramite il sottocomando MISSING).
- Calcolare medie marginali stimate per i valori di covariate specificati (tramite la parola chiave WITH del sottocomando EMMEANS).
- Confrontare effetti principali semplici di interazioni (tramite il sottocomando EMMEANS).

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

# ***Modelli lineari generalizzati***

Il modello lineare generalizzato amplia il modello lineare generale in modo che la variabile dipendente venga linearmente correlata ai fattori e alle covariate tramite una funzione di collegamento specifica. Inoltre, il modello consente alla variabile dipendente di avere una distribuzione non normale. Il modello lineare generalizzato include modelli statistici ampiamente usati come la regressione lineare per le risposte distribuite normalmente, i modelli logistici per i dati binari, i modelli loglineari per i dati dei conteggi, i modelli log-log complementari per dati di sopravvivenza troncati a intervalli, nonché molti altri modelli statistici. Ciò è possibile grazie al fatto che il modello formulato è molto generale.

**Esempi.** Una compagnia di navigazione può utilizzare modelli lineari generalizzati per adattare una regressione di Poisson ai conteggi dei danni relativi a vari tipi di navi costruite in periodi differenti e il modello risultante può aiutare a determinare quali tipi di navi sono più soggetti a subire danni.

Una compagnia di assicurazioni auto può utilizzare modelli lineari generalizzati per adattare una regressione gamma alle richieste di risarcimento danni e il modello risultante può aiutare a determinare i fattori che contribuiscono maggiormente all'ammontare del risarcimento.

I ricercatori di medicina possono utilizzare modelli lineari generalizzati per adattare una regressione log-log complementare per i dati di sopravvivenza troncati a intervalli per prevedere la ricorrenza di una condizione medica.

**Dati.** La risposta può essere una scala, un conteggio, binaria o equivalente a un numero di eventi. Si presume che i fattori siano categoriali e che le covariate, il peso della scala e l'offset siano variabili scala.

**Assunzioni.** Si presume che i casi siano osservazioni indipendenti.

## ***Per ottenere un modello lineare generalizzato***

Dai menu, scegliere:

Analizza > Modelli lineari generalizzati > Modelli lineari generalizzati...

Figura 6-1  
 Scheda Modelli lineari generalizzati: Tipo di modello

- ▶ Specificare una distribuzione e una funzione di collegamento (vedere la sezione che segue per informazioni sulle varie opzioni).
- ▶ Nella scheda **Risposta**, selezionare una variabile dipendente.
- ▶ Nella scheda **Predittori**, selezionare i fattori e le covariate per utilizzarli nella previsione della variabile dipendente.
- ▶ Nella scheda **Model**, specificare gli effetti del modello utilizzando i fattori e le covariate selezionati.

La scheda Type of Model consente di specificare la distribuzione e la funzione di collegamento per il proprio modello e offre collegamenti per vari modelli comuni che sono classificati per tipo di risposta.

### ***Tipi di modello***

#### **Risposta scala.**

- **Lineare.** Specifica Normale come distribuzione e Identità come funzione di collegamento.
- **Gamma con collegamento log.** Specifica Gamma come distribuzione e Log come funzione di collegamento.

#### **Risposta ordinale.**

- **Logistica ordinale.** Specifica Multinomiale (ordinali) come distribuzione e Logit cumulativa come funzione di collegamento.
- **Probit ordinale.** Specifica Multinomiale (ordinali) come distribuzione e Probit cumulativa come funzione di collegamento.

#### **Conteggi.**

- **Loglineare Poisson.** Specifica Poisson come distribuzione e Log come funzione di collegamento.
- **Binomiale negativa con collegamento log.** Specifica una distribuzione negativa binomiale (con un valore di 1 per il parametro ausiliario) come distribuzione e Log come funzione di collegamento. Affinché, tramite la procedura, venga stimato il valore del parametro ausiliario, specificare un modello personalizzato con distribuzione binomiale negativa e selezionare Stima valore nel gruppo dei parametri.

#### **Risposta binaria o dati eventi/prove.**

- **Logistica binaria.** Specifica Binominale come distribuzione e Logit come funzione di collegamento.
- **Probit binaria.** Specifica Binominale come distribuzione e Probit come funzione di collegamento.
- **Sopravvivenza di censura per intervallo.** Specifica Binominale come distribuzione e Log-log complementare come funzione di collegamento.

#### **Combinazione.**

- **Tweedie con collegamento log.** Specifica Tweedie come distribuzione e Log come funzione di collegamento.
- **Tweedie con collegamento identità.** Specifica Tweedie come distribuzione e Identità come funzione di collegamento.

**Personalizzata.** Specificare la propria combinazione di funzioni di distribuzione e di collegamento.

#### **Distribuzione**

Questa selezione specifica la distribuzione della variabile dipendente. La possibilità di specificare una distribuzione non normale e una funzione di collegamento senza identità è uno dei principali vantaggi offerti dal modello lineare generalizzato rispetto a quello generale. Dal momento che è possibile combinare più distribuzioni e funzioni di collegamento e che molte di queste sono adatte a qualsiasi combinazione di dati, è generalmente consigliabile fare una valutazione teorica a priori oppure selezionare la combinazione che si ritiene possa essere più adatta.

- **Binomiale.** Questa distribuzione è indicata solo per le variabili che rappresentano una risposta binaria o un numero di eventi.

- **Gamma.** Questa distribuzione è indicata per le variabili con valori di scala positivi che presentano una distribuzione asimmetrica in direzione dei valori positivi più alti. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Gaussiana inversa.** Questa distribuzione è indicata per le variabili con valori di scala positivi che presentano una distribuzione asimmetrica in direzione dei valori positivi più alti. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Binomiale negativo.** Questa distribuzione può essere considerata come il numero di prove necessario per osservare  $k$  successi ed è indicata per variabili con valori interi non negativi. Se il valore dei dati è un numero non intero, inferiore a 0 o mancante, il caso corrispondente non viene usato nell'analisi. Il valore del parametro ausiliario della distribuzione binomiale negativa può essere qualsiasi numero maggiore o uguale a 0; è possibile impostarlo su un valore fisso o consentirne la stima tramite la procedura. Se il parametro ausiliario viene impostato su 0, il risultato di questa distribuzione è uguale a quello della distribuzione di Poisson.
- **Normale.** Questa distribuzione è indicata per le variabili di scala i cui valori presentano una distribuzione simmetrica a forma di campana intorno al valore centrale (medio). La variabile dipendente deve essere numerica.
- **Poisson.** Questa distribuzione può essere considerata equivalente al numero di occorrenze di un evento desiderato in un intervallo di tempo fisso ed è indicata per le variabili con valori interi non negativi. Se il valore dei dati è un numero non intero, inferiore a 0 o mancante, il caso corrispondente non viene usato nell'analisi.
- **Tweedie.** Questa distribuzione è indicata per le variabili che possono essere rappresentate da combinazioni di Poisson di distribuzioni gamma; la distribuzione è "mista" nel senso che combina le proprietà di distribuzioni continue (assumono valori reali non negativi) e discrete (massa di probabilità positiva con un unico valore, 0). La variabile dipendente deve essere numerica, con i valori dei dati maggiori o uguali a zero. Se un valore dei dati è minore di zero o mancante, il caso corrispondente non viene usato nell'analisi. Il valore fisso del parametro della distribuzione Tweedie può essere qualsiasi numero maggiore di uno e minore di due.
- **Multinomiale.** Questa distribuzione è indicata per le variabili che rappresentano una risposta ordinale. La variabile dipendente può essere di tipo numerico o stringa e deve avere almeno due valori di dati distinti validi.

### **Funzioni di collegamento**

La funzione Collegamento è la trasformazione della variabile dipendente che permette di stimare il modello. Sono disponibili le seguenti funzioni:

- **Identità.**  $f(x)=x$ . La variabile dipendente non viene trasformata. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.
- **Log-log complementare.**  $f(x)=\log(-\log(1-x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Cauchit cumulativa.**  $f(x) = \tan(\pi (x - 0.5))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.

- **Log-log complementare cumulativa.**  $f(x)=\ln(-\ln(1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Logit cumulativa.**  $f(x)=\ln(x / (1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Log-log negativa cumulativa.**  $f(x)=-\ln(-\ln(x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Probit cumulativa.**  $f(x)=\Phi^{-1}(x)$ , applicata alla probabilità cumulata di ciascuna categoria della risposta, dove  $\Phi^{-1}$  è la funzione di distribuzione cumulata normale standard inversa. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Log.**  $f(x)=\log(x)$ . Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.
- **Complemento log.**  $f(x)=\log(1-x)$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Binomiale negativa.**  $f(x)=\log(x / (x+k^{-1}))$ , dove  $k$  è il parametro ausiliario della distribuzione binomiale negativa. Questa funzione è indicata solo per la distribuzione binomiale negativa.
- **Log-log negativo.**  $f(x)=-\log(-\log(x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Potenza odd.**  $f(x)=[(x/(1-x))^{\alpha}-1]/\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  è la specifica del numero richiesto e deve essere un numero reale. Questa funzione è indicata solo per la distribuzione binomiale.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , dove  $\Phi^{-1}$  è la funzione di distribuzione cumulata normale standard inversa. Questa funzione è indicata solo per la distribuzione binomiale.
- **Potenza.**  $f(x)=x^{\alpha}$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  è la specifica del numero richiesto e deve essere un numero reale. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.

## Risposta dei modelli lineari generalizzati

Figura 6-2  
Finestra di dialogo Modelli lineari generalizzati

In molti casi è possibile specificare semplicemente una variabile dipendente. Tuttavia, le variabili che possono assumere solo due valori e le risposte che registrano gli eventi durante le prove richiedono più attenzione.

- Risposta binaria.** Se la variabile dipendente può assumere solo due valori, è possibile specificare la [categoria di riferimento](#) per la stima dei parametri. Le variabili di risposta binaria possono essere di tipo stringa o numeriche.
- Numero di eventi occorsi in un insieme di prove.** Se la risposta rappresenta il numero di eventi di un insieme di prove, la variabile dipendente contiene il numero di eventi ed è possibile selezionare un'ulteriore variabile che contenga il numero di prove. In alternativa, se il numero di prove è lo stesso per tutti i soggetti, è possibile specificare le prove utilizzando un valore fisso. Il numero di prove deve essere maggiore o uguale al numero di eventi di ciascun caso. Gli eventi devono essere numeri interi non negativi, mentre le prove devono essere numeri interi positivi.

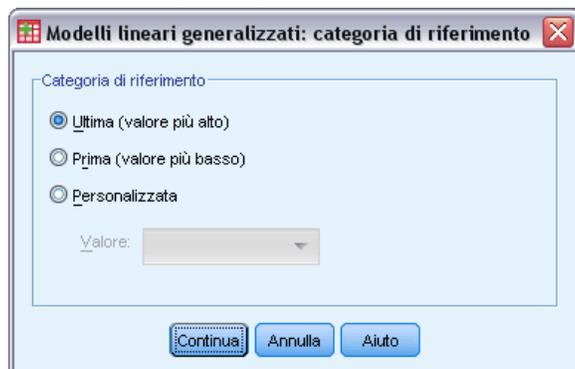
Per i modelli multinomiali ordinali è possibile specificare l'ordine delle categorie della risposta, che può essere: crescente, decrescente o dati. L'ordine dati significa che il primo valore trovato nei dati definisce la prima categoria e l'ultimo valore trovato definisce l'ultima categoria.

**Peso della scala.** Il parametro scala è un parametro del modello stimato correlato alla varianza della risposta. I pesi della scala sono valori "noti" che possono variare a seconda delle osservazioni. Se la variabile del peso della scala è stata specificata, il parametro scala, che è correlato alla varianza della risposta, viene diviso per la suddetta variabile per ciascuna osservazione. Per l'analisi non vengono usati i casi con valori di pesi scala inferiori o uguali a 0 o mancanti.

### ***Categoria di riferimento dei modelli lineari generalizzati***

Figura 6-3

*Finestra di dialogo Categoria di riferimento dei modelli lineari generalizzati*

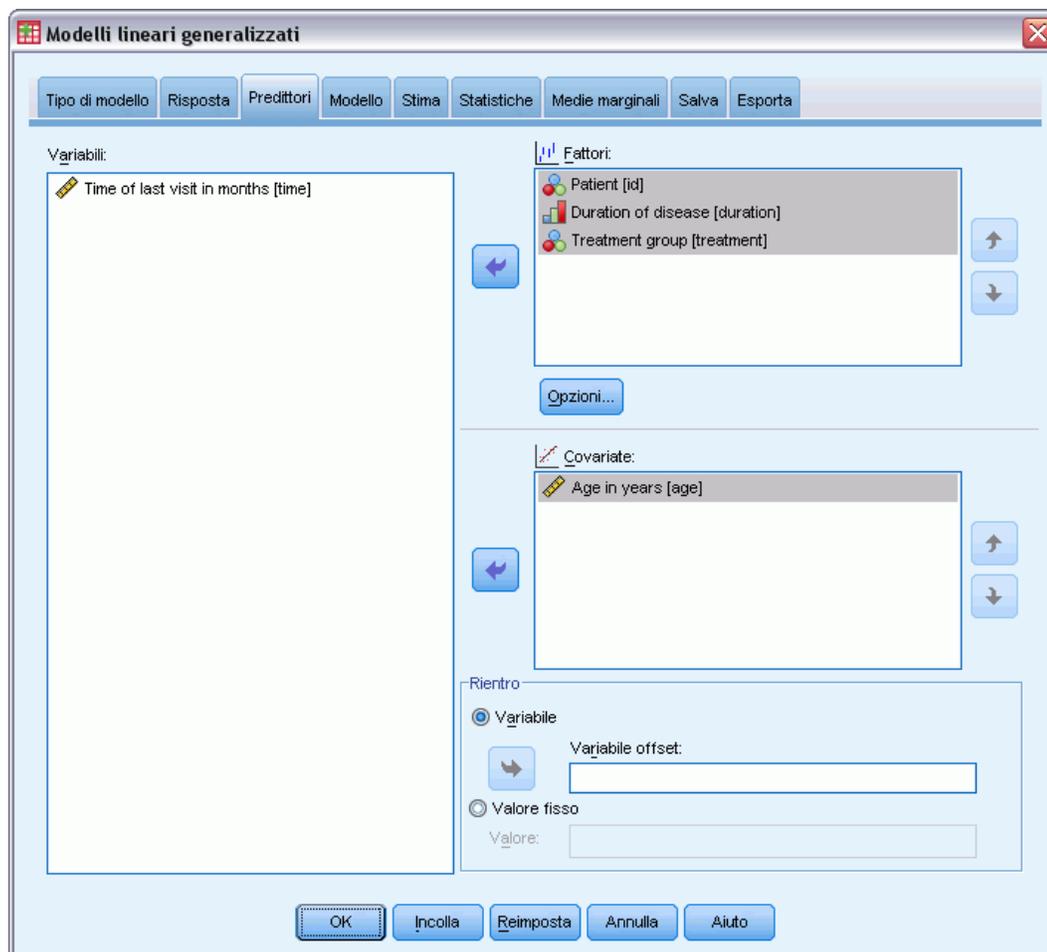


Per la risposta binaria, è possibile scegliere la categoria di riferimento per la variabile dipendente. Questa operazione può influire su un determinato output, ad esempio quello delle stime dei parametri e dei valori salvati, ma non dovrebbe modificare l'adattamento del modello. Ad esempio, se la risposta binaria assume i valori 0 e 1:

- Per impostazione predefinita, la procedura imposta l'ultima categoria (dal valore più alto), o 1, come categoria di riferimento. In questa situazione, le probabilità salvate del modello stimano la probabilità che un determinato caso assuma il valore 0, mentre le stime dei parametri dovrebbero essere interpretate come relative alla verosimiglianza della categoria 0.
- Se si specifica la prima categoria (dal valore più basso), oppure 0, come categoria di riferimento, le probabilità salvate del modello stimano la probabilità che un determinato caso assuma il valore 1.
- Se si specifica la categoria personalizzata e la variabile ha delle etichette definite, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Questo può risultare conveniente quando, durante un'operazione di specifica di un modello, non ci si ricorda esattamente del modo in cui è stata codificata una determinata variabile.

## Predittori dei modelli lineari generalizzati

Figura 6-4  
Generalized Linear Models: scheda Predittori



La scheda Predittori consente di specificare i fattori e le covariate usati per definire gli effetti del modello nonché di definire un offset opzionale.

**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.

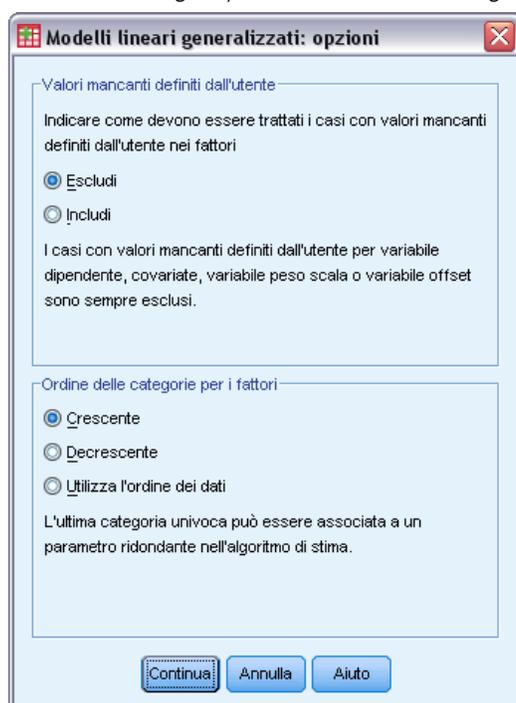
*Nota:* se la risposta è binomiale ed ha un formato binario, la procedura calcola le statistiche di devianza e di bontà d'adattamento chi-quadrato in base alle sottopopolazioni, che si basano sulla classificazione incrociata dei valori osservati per i fattori e le covariate selezionati. È generalmente consigliabile usare lo stesso insieme di predittori nelle varie esecuzioni della procedura per mantenere un numero di sottopopolazioni uniforme.

**Offset.** Questo termine identifica un predittore “strutturale”. Il suo coefficiente non viene stimato utilizzando dal modello ma ritenuto sempre equivalente a 1. Quindi, i valori dell'offset vengono semplicemente aggiunti al predittore lineare della variabile dipendente. Ciò è particolarmente

utile nei modelli di regressione di Poisson, in cui ciascun caso può avere vari livelli di esposizione all'evento desiderato. Ad esempio, quando si creano modelli per calcolare la percentuale di incidenti per i singoli conducenti, esiste una sostanziale differenza tra il conducente che ha avuto un incidente in tre anni di guida e uno che ne ha avuto uno in 25 anni. È possibile creare un modello del numero di incidenti come risposta di Poisson, se l'esperienza del conducente è inclusa come termine di offset.

## Opzioni dei modelli lineari generalizzati

Figura 6-5  
Finestra di dialogo Opzioni dei modelli lineari generalizzati



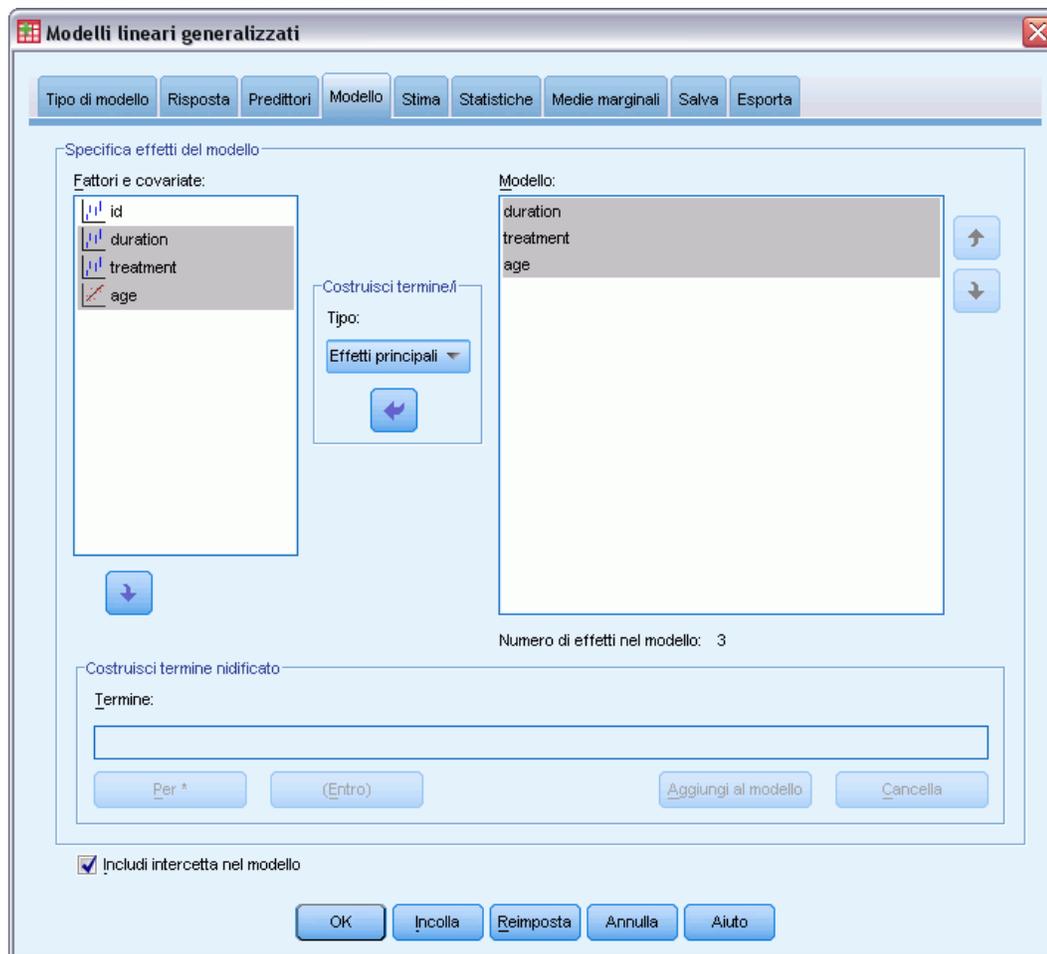
Queste opzioni vengono applicate a tutti i fattori selezionati nella scheda Predittori.

**Valori mancanti definiti dall'utente.** Affinché un caso possa essere incluso nell'analisi, è necessario che i fattori abbiano valori validi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito delle variabili dei fattori.

**Ordine delle categorie.** Questa opzione è importante per stabilire l'ultimo livello di un fattore che può essere associato a un parametro ridondante nell'algoritmo di stima. La modifica dell'ordine delle categorie influisce sui valori degli effetti a livello di fattore poiché le stime di questi parametri vengono calcolate in funzione dell'“ultimo” livello. I fattori possono essere disposti in ordine crescente (dal valore più basso a quello più alto), in ordine decrescente (dal valore più alto a quello più basso) oppure in base all'ordine dei dati. Ciò significa che il primo valore trovato nei dati definisce la prima categoria, e l'ultimo valore univoco trovato definisce l'ultima categoria.

## Modello per i modelli lineari generalizzati

Figura 6-6  
Generalized Linear Models: scheda Modello



**Specifica modello effetti.** Il modello predefinito prevede solo l'uso dell'intercetta, quindi è necessario specificare esplicitamente altri effetti per il modello. In alternativa, è possibile creare termini nidificati o non nidificati.

### Termini non nidificati

Per i fattori e le covariate selezionati:

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

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

**Fattoriale.** Consente di creare tutte le possibili interazioni e gli effetti principali delle variabili selezionate.

**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'intercetta può essere esclusa.

I modelli con la distribuzione ordinale multinomiale non dispongono di un singolo termine di intercetta; invece, vi sono parametri di soglia che definiscono i punti di transizione tra categorie adiacenti. Le soglie vengono sempre incluse nel modello.

## Stima dei modelli lineari generalizzati

Figura 6-7  
Generalized Linear Models: scheda Stima

The screenshot shows the 'Stima' (Estimation) tab of the 'Modelli lineari generalizzati' software. The interface includes a menu bar with options: Tipo di modello, Risposta, Predittori, Modello, Stima, Statistiche, Medie marginali, Salva, and Esporta. The main area is divided into two sections: 'Stima dei parametri' and 'Iterazioni'.

**Stima dei parametri:**

- Metodo:** Ibrido (dropdown menu)
- Iterazioni punteggio massimo Fisher:** 1 (text input)
- Metodo parametro scala:** Valore fisso (dropdown menu)
- Valore:** 1 (text input)
- Matrice di covarianza:**
  - Stimatore basato sul modello
  - Stimatore robusto
- Recupera valori iniziali per le stime dei parametri da un insieme di dati
- 

**Iterazioni:**

- Massimo numero di iterazioni:** 100 (text input)
- Max dimezzamenti:** 5 (text input)
- Verifica per la separazione dei punti di dati
- Iterazione iniziale:** 20 (text input)

**Criteri di convergenza:**

Deve essere specificato almeno un criterio di convergenza con un minimo maggiore di 0.

	Minimo:	Tipo:
<input checked="" type="checkbox"/> Modifica nelle stime dei parametri	1E-006	Assoluto
<input type="checkbox"/> Modifica in verosimiglianza		Assoluto
<input type="checkbox"/> Convergenza hessiana		Assoluto

**Tolleranza della singolarità:** 1E-012 (dropdown menu)

Buttons at the bottom: OK, Incolla, Reimposta, Annulla, Aiuto.

**Stima dei parametri.** I controlli di questo gruppo consentono di specificare i metodi di stima e di definire i valori iniziali per la stima dei parametri.

- **Metodo.** È possibile selezionare un metodo di stima per i parametri. In particolare è possibile scegliere il metodo di Newton-Raphson, il punteggio Fisher o un metodo ibrido in cui si eseguono le iterazioni per il calcolo del punteggio 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.
- **Metodo basato sul parametro scala.** È possibile selezionare il metodo di stima basato sui parametri scala. La massima verosimiglianza stima i parametri di scala utilizzando gli effetti del modello, benché questa opzione non sia valida se la risposta ha, una distribuzione multinomiale, binomiale, di Poisson o binomiale negativa. Le opzioni relative alla devianza e il chi-quadrato di Pearson stimano i parametri scala utilizzando i valori risultanti dalle relative statistiche. In alternativa, è possibile specificare un valore fisso per il parametro scala.

- **Valori iniziali.** La procedura calcola automaticamente i valori iniziali dei parametri. In alternativa, è possibile specificare i **valori iniziali** per le stime del parametro.
- **Matrice di covarianza.** Lo stimatore basato sul modello è il valore negativo dell'inverso generalizzato della matrice hessiana. Lo stimatore robusto (chiamato anche Huber/White/sandwich) è lo stimatore basato sul modello "corretto" che fornisce una stima uniforme della covarianza anche quando le specifiche della varianza e la funzione di collegamento sono errate.

#### Iterazioni.

- **Max 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.
- **Verificare la 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. Questa opzione è disponibile per risposte multinomiali e risposte binomiali con formato binario.

#### Criteri di convergenza.

- **Convergenza 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.
- **Convergenza 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.
- **Convergenza hessiana.** In caso di specifica Assoluta, la convergenza viene presunta se la statistica basata sulla convergenza hessiana è inferiore al valore positivo specificato. Per la specifica relativa, la convergenza viene presunta se la statistica è inferiore al prodotto del valore positivo specificato e del valore assoluto del logaritmo di verosimiglianza.

**Tolleranza della singolarità.** Le matrici singole (o non reversibili) hanno colonne linearmente dipendenti che possono creare gravi problemi all'algoritmo di stima. Poiché anche le matrici quasi singolari possono generare risultati inadeguati, la procedura considera le matrici i cui determinanti sono inferiori alla tolleranza come singolari. Specificare un valore positivo.

## Valori iniziali dei modelli lineari generalizzati

Figura 6-8  
Finestra di dialogo Valori iniziali dei modelli lineari generalizzati



Se si specificano valori iniziali, è necessario indicarli per tutti i parametri del modello, compresi quelli ridondanti. Nell'insieme dei dati, l'ordine delle variabili da sinistra a destra deve essere: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., dove *RowType\_* e *VarName\_* sono variabili di tipo stringa e *P1*, *P2*, ... sono variabili numeriche che corrispondono a un elenco ordinato dei parametri.

- I valori iniziali vengono forniti in un record con valore *EST* per la variabile *RowType\_*; i valori iniziali effettivi sono specificati nelle variabili *P1*, *P2*, .... La procedura ignora tutti i record per i quali *RowType\_* ha un valore diverso da *EST*, nonché qualsiasi record oltre la prima occorrenza di *RowType\_* uguale a *EST*.
- L'intercetta, se inclusa nel modello, o i parametri di soglia, se la risposta ha una distribuzione multinomiale, devono essere i primi valori iniziali elencati.
- Il parametro di scala *e*, se la risposta ha una distribuzione binomiale negativa, il parametro binomiale negativo, devono essere gli ultimi valori iniziali specificati.
- Se è stata abilitata l'opzione *Dividi*, le variabili devono iniziare con le variabili dei file, suddivise nell'ordine specificato durante la creazione del file suddiviso, seguite da *RowType\_*, *VarName\_*, *P1*, *P2*, ... come indicato in precedenza. Le divisioni nell'insieme di dati specificato devono verificarsi nello stesso ordine in cui si verificano nell'insieme di dati originale.

*Nota:* i nomi delle variabili *P1*, *P2*, ... non sono obbligatori; la procedura accetta qualsiasi nome di variabile valido per i parametri poiché l'associazione delle variabili ai parametri viene effettuato in base alla posizione della variabile, non in base al nome. Tutte le variabili successive all'ultimo parametro vengono ignorate.

La struttura del file per i valori iniziali è la stessa di quella usata per l'esportazione del modello sotto forma di dati; quindi, è possibile usare i valori finali di un'esecuzione precedente della procedura come input per un'esecuzione successiva.

## Statistiche dei modelli lineari generalizzati

Figura 6-9  
Generalized Linear Models: scheda Statistiche

### Effetti del modello.

- Tipo di analisi.** Specificare il tipo di analisi che si desidera ottenere. Le analisi di tipo I sono generalmente indicate nei casi in cui ci siano ragioni a priori per ordinare i predittori nel modello, mentre le analisi di tipo III sono applicabili a contesti più generali. Le statistiche di Wald o le statistiche del rapporto di verosimiglianza vengono calcolate in base alla selezione effettuata nel gruppo Statistiche chi-quadrato.
- Intervalli di confidenza.** Specificare un livello di confidenza superiore a 50 e inferiore a 100. Gli intervalli di Wald si basano sul presupposto che i parametri abbiano una distribuzione normale asintotica; gli intervalli basati sulla verosimiglianza dei profili sono più precisi ma possono richiedere operazioni di calcolo complesse. Il livello di tolleranza per gli intervalli

basati sulla verosimiglianza dei profili è il criterio utilizzato per arrestare l'algoritmo iterativo impiegato per il calcolo degli intervalli.

- **Funzione di verosimiglianza.** Controlla il formato di visualizzazione della funzione di verosimiglianza. La funzione completa comprende un ulteriore termine che è costante rispetto alle stime dei parametri; non ha alcun effetto sulla stima dei parametri e viene escluso dalla visualizzazione in alcuni prodotti software.

**Stampa.** L'output disponibile è descritto di seguito.

- **Riepilogo dell'elaborazione dei casi.** Visualizza il numero e la percentuale di casi inclusi ed esclusi nella e dall'analisi e la tabella di riepilogo dei dati correlati.
- **Statistiche descrittive.** Visualizza le statistiche descrittive e le informazioni di riepilogo sulla variabile dipendente, sulle covariate e sui fattori.
- **Informazioni sul modello.** Visualizza il nome dell'insieme dei dati, della variabile dipendente, degli eventi, delle variabili delle prove, della variabile di offset, della variabile dei pesi della scala, della distribuzione di probabilità e della funzione di collegamento.
- **Statistiche di bontà dell'adattamento.** Visualizza la devianza e la devianza scalata, il chi-quadrato di Pearson, il chi-quadrato di Pearson scalato, la verosimiglianza, il criterio di informazione di Akaike (AIC), il criterio di informazione di Akaike corretto per il campione finito (AICC), il criterio di informazione bayesiano (BIC) e il criterio di informazione di Akaike uniforme (CAIC).
- **Statistiche riassuntive del modello.** Visualizza i test di adattamento del modello comprese le statistiche del rapporto di verosimiglianza per il test di adattamento del modello Omnibus e le statistiche per i contrasti di tipo I o II per ciascun effetto.
- **Stime dei parametri.** Visualizza le stime dei parametri, le statistiche corrispondenti e gli intervalli di confidenza. È possibile anche visualizzare le stime dei parametri elevati a potenza oltre alle stime dei parametri grezzi.
- **Matrice di covarianza per le stime dei parametri.** Visualizza la matrice di covarianza residua dei parametri stimata.
- **Matrice di correlazione per le stime dei parametri.** Visualizza la matrice di correlazione dei parametri stimata.
- **Matrici dei coefficienti di contrasto (L).** Visualizza i coefficienti di contrasto per gli effetti predefiniti e per le medie marginali stimate se si seleziona l'opzione corrispondente nella scheda Medie marginali.
- **Funzioni funzionali generali.** Visualizza le matrici per la generazione delle matrici dei coefficienti di contrasto (L).
- **Cronologia iterazioni.** Visualizza la cronologia delle iterazioni per le stime dei parametri e la verosimiglianza e stampa l'ultima valutazione del vettore del gradiente e la matrice hessiana. La tabella Cronologia iterazioni visualizza le stime dei parametri per tutte le  $n^{\circ}$  iterazioni a partire dalla 0-esima iterazione (stime iniziali), dove  $n$  è il valore dell'intervallo di stampa. Se la cronologia delle iterazioni è richiesta, l'ultima iterazione viene sempre visualizzata indipendentemente da  $n$ .
- **Test del moltiplicatore di Lagrange.** Visualizza le statistiche del test del moltiplicatore di Lagrange che consentono di valutare la validità di un parametro scala calcolato usando la devianza o il chi-quadrato di Pearson o di un insieme corrispondente a un numero fisso per le

distribuzioni normale, gamma, gaussiana inversa e Tweedie. Per la distribuzione binomiale negativa, viene verificato il parametro ausiliario fisso.

## Medie marginali dei modelli lineari generalizzati

Figura 6-10  
Generalized Linear Models: scheda Medie marginali

**Fattori e interazioni:**

M	Termine
<input checked="" type="checkbox"/>	duration
<input checked="" type="checkbox"/>	treatment
<input type="checkbox"/>	id

**Visualizza medie per:**

Termine	Contrasto	Categoria di riferim...
duration	Semplice	2
treatment	None	
id	None	

**Scala**

Calcola medie per risultato

Calcola medie per il predittore lineare

**Correzione per confronti multipli:**

Differenza meno significativa

Visualizza media stimata complessiva

OK Incolla Reimposta Annulla Aiuto

Questa scheda consente di visualizzare le medie marginali stimate per i livelli e le interazioni tra fattori. È possibile anche visualizzare la media stimata globale. Le medie marginali stimate non sono disponibili per i modelli multinomiali ordinali.

**Fattori e interazioni.** Questo elenco contiene i fattori specificati nella scheda Predittori e le interazioni tra fattori specificate nella scheda Modello. L'elenco non comprende le covariate. I termini possono essere selezionati direttamente dall'elenco oppure combinati in un termine di interazione utilizzando il pulsante Per \*.

**Medie marginali per.** Le medie stimate vengono calcolate per i fattori e le interazioni tra fattori selezionate. Il contrasto determina la modalità di impostazione dei test di ipotesi per il confronto delle medie stimate. Per il contrasto semplice è necessaria una categoria di riferimento o un livello di fattore da usare per il confronto.

- **Paiwise (Analisi dei dati mancanti).** I confronti a coppie vengono calcolati per le combinazioni dei fattori specificati o implicati di tutti i livelli. Questa è l'unico contrasto disponibile per le interazioni tra fattori.
- **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.** Ciascun livello di fattore viene confrontato con la media principale. I contrasti di deviazione non sono ortogonali.
- **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.
- **Ripetuto.** 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.

**Scala.** È possibile calcolare le medie marginali stimate per la risposta o il predittore lineare, utilizzando rispettivamente la scala originale della variabile indipendente e la variabile dipendente trasformata dalla funzione di collegamento.

**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.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.
- **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.
- **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.

## Salvataggio dei modelli lineari generalizzati

Figura 6-11  
Generalized Linear Models: scheda Salva

Salva	Elemento da salvare	Nome variabile o nome radice	Categorie da salvare
<input type="checkbox"/>	Valore atteso della media della risposta	MeanPredicted	
<input type="checkbox"/>	Limite inferiore dell'intervallo di confidenza per la media della risposta	CIMeanPredictedLower	
<input type="checkbox"/>	Limite superiore dell'intervallo di confidenza per la media della risposta	CIMeanPredictedUpper	
<input type="checkbox"/>	Categoria prevista	Valoreprevisto	
<input type="checkbox"/>	Valore atteso del predittore lineare	XBPredicted	
<input type="checkbox"/>	Errore standard stimato del valore atteso del predittore lineare	XBStandardError	
<input type="checkbox"/>	Distanza di Cook	CooksDistance	
<input type="checkbox"/>	Valore di influenza	Leverage	
<input type="checkbox"/>	Residuo	Residuo	
<input type="checkbox"/>	Residuo di Pearson	PearsonResidual	
<input type="checkbox"/>	Residuo di Pearson standardizzato	StdPearsonResidual	
<input type="checkbox"/>	Residuo di devianza	DevianceResidual	
<input type="checkbox"/>	Residuo della devianza standardizzata	StdDevianceResidual	
<input type="checkbox"/>	Residuo della verosimiglianza	LikelihoodResidual	

Variabile esistente con lo stesso nome

Aggiungi suffisso al nome della nuova variabile (applicabile solo ai nomi predefiniti)  
 Sostituisci variabile esistente (applicabile ai nomi predefiniti e definiti dall'utente)

Se si forniscono nomi specifici per le variabili, assicurarsi che non siano in conflitto con le variabili esistenti nell'insieme di dati attivo. Selezionare l'opzione Sostituisci variabile esistente per utilizzare i nomi forniti dall'utente per sostituire le variabili esistenti.

OK   Incolla   Reimposta   Annulla   Aiuto

Gli elementi selezionati vengono salvati con il nome specificato; è possibile scegliere di sovrascrivere le variabili esistenti con lo stesso nome oppure evitare possibili conflitti aggiungendo suffissi che rendano i nomi delle nuove variabili univoci.

- Valore atteso della media della risposta.** Salva i valori previsti per il modello per ciascun caso presente nella metrica di risposta originale. Se la distribuzione della risposta è binomiale e la variabile dipendente è binaria, la procedura salva le probabilità previste. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Probabilità prevista cumulativa e la procedura salva la probabilità prevista cumulativa per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- Lower bound of confidence interval for mean of response.** Salva il limite inferiore dell'intervallo di confidenza per la media della risposta. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Correzione inferiore dell'intervallo di confidenza per la probabilità prevista cumulativa e la procedura salva il limite inferiore per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.

- **Upper bound of confidence interval for mean of response.** Salva il limite superiore dell'intervallo di confidenza per la media della risposta. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Correzione superiore dell'intervallo di confidenza per la probabilità prevista cumulativa e la procedura salva il limite superiore per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- **Categoria prevista.** Per i modelli con la distribuzione binomiale e la variabile dipendente binaria oppure con la distribuzione multinomiale, salva la categoria della risposta prevista per ogni caso. Questa opzione non è disponibile per altre distribuzioni della risposta.
- **Valore atteso del predittore lineare.** Salva i valori previsti per il modello per ciascun caso nella metrica nel predittore lineare (risposta trasformata tramite la funzione di collegamento specifica). Se la distribuzione della risposta è multinomiale, la procedura salva il valore previsto per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- **Errore standard stimato del valore atteso del predittore lineare.** Se la distribuzione della risposta è multinomiale, la procedura salva l'errore standard stimato per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.

I seguenti elementi non sono disponibili quando la distribuzione della risposta è multinomiale.

- **Distanza di Cook.** Una misura di quanto cambierebbero i residui di tutti i casi se un particolare caso fosse escluso dal calcolo dei coefficienti di regressione. Valori alti indicano che l'esclusione di un caso dal calcolo dei coefficienti di regressione ne modificherebbe sostanzialmente il valore.
- **Valore di influenza.** Una misura dell'influenza di un dato sull'adattamento della regressione. L'influenza centrata varia da 0 (nessuna influenza sull'adattamento) a  $(N-1)/N$ .
- **residuo grezzo.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Residuo di Pearson.** Radice quadrata del contributo di un caso alla statistica chi-quadrato di Pearson con il segno del residuo grezzo.
- **Residuo standardizzato di Pearson.** Residuo di Pearson moltiplicato per la radice quadrata dell'inverso del prodotto del parametro scala e dell'influenza  $1-$  per il caso.
- **Residuo di devianza** Radice quadrata del contributo di un caso alla statistica di devianza con il segno del residuo grezzo.
- **Residuo della devianza standardizzato.** Residuo di devianza moltiplicato per la radice quadrata dell'inverso del prodotto del parametro scala e dell'influenza  $1-$  per il caso.
- **Residuo di verosimiglianza.** Radice quadrata di una media ponderata (basata sull'influenza del caso) dei quadrati dei residui di Pearson e della devianza standardizzati, con il segno del residuo grezzo.

## Esportazione dei modelli lineari generalizzati

Figura 6-12  
Generalized Linear Models: scheda Esporta

**Esporta modello come dati.** 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.

- **Variabili di distinzione.** Se utilizzata, qualsiasi variabile definisce le divisioni.
- **RowType\_** Assume valori (ed etichette di 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 stimati (ad eccezione dei parametri di scala o dei parametri binomiali negativi), per i tipi di riga *COV* or *CORR*, con etichette dei valori corrispondenti alle stringhe

dei parametri mostrate nella tabella 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 (inclusi i parametri di scala e i parametri binomiali negativi, se necessario), con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella 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.

Per il parametro di scala, le covarianze, le correlazioni, il livello di significatività e i gradi di libertà sono impostati sul valore mancante di sistema. Se il parametro di scala viene stimato tramite la massima verosimiglianza, viene generato l'errore standard; altrimenti, tale parametro viene impostato sul valore mancante di sistema.

Per il parametro binomiale negativo, le covarianze, le correlazioni, il livello di significatività e i gradi di libertà sono impostati sul valore mancante di sistema. Se il parametro binomiale negativo viene stimato tramite la massima verosimiglianza, viene generato l'errore standard; altrimenti, tale parametro viene impostato sul valore mancante di sistema.

Se vi sono divisioni, l'elenco di parametri deve essere accumulato in tutte le divisioni. In una determinata divisione, alcuni parametri possono risultare irrilevanti; cosa ben diversa dai parametri ridondanti. Per i parametri irrilevanti, tutte le covarianze o le correlazioni, le stime dei parametri, gli errori standard, i livelli di significatività e i gradi di libertà sono impostati sul valore mancante di sistema.

Questo file matrice può essere usato per definire i valori iniziali e stimare altri modelli. Si noti che questo file non può essere immediatamente usato 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. Nonostante ciò, è necessario accertarsi che tutti i parametri presenti in questo file di matrice abbiano lo stesso significato per la procedura che esegue la lettura del file.

**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.

## ***Funzioni aggiuntive del comando GENLIN***

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i valori iniziali per le stime dei parametri come elenco di numeri (utilizzando il sottocomando `CRITERIA`).
- Fissare le covariate sui valori diversi dalle relative medie durante il calcolo delle medie marginali stimate (utilizzando il sottocomando `EMMEANS`).

- Specificare contrasti polinomiali personalizzati per le medie marginali stimate (utilizzando il sottocomando `EMMEANS`).
- Specificare un sottoinsieme dei fattori che consente di visualizzare le medie marginali stimate e di confrontarle utilizzando il tipo di contrasto specificato (utilizzando le parole chiave `TABLES` e `COMPARE` del sottocomando `EMMEANS`).

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

## ***Equazioni di stima generalizzate***

La procedura Equazioni di stima generalizzate estende il modello lineare generalizzato per consentire l'analisi di misurazioni ripetute o altre osservazioni correlate come i dati raggruppati.

**Esempio.** I responsabili della sanità pubblica possono utilizzare equazioni di stima generalizzate per adattare una regressione logistica con misurazioni ripetute per lo studio degli effetti dell'inquinamento atmosferico sui bambini.

**Dati.** La risposta può essere una scala, un conteggio, binaria o equivalente a un numero di eventi. Si presume che i fattori siano categoriali e che le covariate, il peso della scala e l'offset siano variabili scala. Le variabili usate per definire i soggetti o le misurazioni ripetute entro i soggetti non possono essere usate per definire la risposta, ma possono assumere altri ruoli nel modello.

**Assunzioni.** Si presume che i casi siano dipendenti entro i soggetti e indipendenti tra i soggetti. La matrice di correlazione che rappresenta le dipendenze entro i soggetti vengono stimate insieme al modello.

### ***Come ottenere equazioni di stima generalizzate***

Dai menu, scegliere:

Analizza > Modelli lineari generalizzati > Equazioni di stima generalizzate...

Figura 7-1  
Equazioni di stima generalizzate: scheda Ripetute

- Selezionare una o più variabili soggetto (vedere la sezione che segue per informazioni sulle opzioni disponibili).

La combinazione di valori delle variabili specificate deve definire in modo univoco i **soggetti** nell'insieme di dati. Ad esempio la variabile singola *ID paziente* è generalmente sufficiente per definire i pazienti di un ospedale. Tuttavia, se i codici di identificazione dei pazienti non sono univoci tra gli ospedali, può essere necessario usare la combinazione di variabili *ID ospedale* e *ID paziente*. Se sono previste più misurazioni, vengono registrate più osservazioni per ciascun soggetto, quindi è possibile che ciascun soggetto occupi più casi nell'ambito dell'insieme dei dati.

- Nella scheda **Tipo di modello**, specificare una distribuzione e una funzione di collegamento.
- Nella scheda **Risposta**, selezionare una variabile dipendente.
- Nella scheda **Predittori**, selezionare i fattori e le covariate per utilizzarli nella previsione della variabile dipendente.
- Nella scheda **Model**, specificare gli effetti del modello utilizzando i fattori e le covariate selezionati.

In alternativa, è possibile anche usare la scheda Ripetute per specificare:

**Variabil entro i soggetti.** La combinazione dei valori delle variabili entro i soggetti definisce l'ordine delle misurazioni entro i soggetti. Quindi, la combinazione di variabili entro i soggetti e soggetto definisce in modo univoco ciascuna misurazione. Ad esempio la combinazione di *Periodo*, *ID ospedale* e *ID paziente* definisce, per ciascun caso, una visita specifica effettuata dal paziente in un ospedale specifico.

Se l'insieme dei dati è già ordinato in modo che le misurazioni ripetute di ciascun soggetto si verificano in un blocco contiguo di casi o nell'ordine corretto, non è necessario specificare la variabile entro i soggetti ed è possibile deselezionare Ordina i casi in base alle variabili soggetto e entro i soggetti in modo da risparmiare tempo e poter eseguire più rapidamente l'ordinamento temporaneo. L'uso delle variabili entro i soggetti è generalmente utile per ordinare in modo corretto le misurazioni.

Le variabili soggetto e entro i soggetti non possono essere usate per definire la risposta, ma svolgono comunque altre funzioni nel modello. La variabile *ID ospedale* ad esempio può essere usata come fattore nel modello.

**Matrice di covarianza.** Lo stimatore basato sul modello è il valore negativo dell'inverso generalizzato della matrice hessiana. Lo stimatore robusto (chiamato anche stimatore di Huber/White/sandwich) è lo stimatore basato sul modello "corretto" che fornisce una stima uniforme della covarianza anche nel caso in cui la matrice di correlazione di lavoro non sia stata specificata correttamente. Questa specifica si applica ai parametri nella parte del modello lineare delle equazioni di stima generalizzate, mentre la specifica nella scheda [Stima](#) si applica soltanto al modello lineare generalizzato iniziale.

**Matrice di correlazione di lavoro.** Questa matrice di correlazione rappresenta le dipendenze entro i soggetti. La sua dimensione è influenzata dal numero di misurazioni e conseguentemente anche dalla combinazione di valori delle variabili entro i soggetti. È possibile specificare una delle seguenti strutture:

- **Indipendente.** Le misure ripetute non vengono correlate.
- **AR(1).** Le misure ripetute hanno una relazione autoregressiva del primo ordine. La correlazione tra due elementi qualsiasi è uguale a  $\rho$  per gli elementi adiacenti, a  $\rho^2$  per gli elementi separati da un terzo elemento e così via.  $\rho$  viene limitato in modo che  $-1 < \rho < 1$ .
- **Scambiabile.** Questa struttura è caratterizzata da correlazioni omogenee tra gli elementi. Viene chiamata anche struttura simmetrica composta.
- **Dipendente da M.** Le misurazioni consecutive hanno un coefficiente di correlazione comune, le coppie di misurazioni separate da un terzo elemento hanno un coefficiente di correlazione comune e così via, nelle coppie di misurazioni separate da altre misurazioni  $m-1$ . Le misurazioni con un grado di separazione più ampio vengono considerate non correlate. Se si seleziona questa struttura, è necessario scegliere un valore  $m$  inferiore all'ordine della matrice di correlazione di lavoro.
- **Non strutturato.** Questa struttura corrisponde a una matrice di correlazione completamente generale.

Per impostazione predefinita, la procedura corregge le stime di correlazione in base al numero di parametri non ridondanti. Tuttavia, può essere utile rimuovere la correzione se si desidera che le stime non subiscano variazioni a causa delle variazioni di replica a livello di soggetto nei dati.

- **Max iterazioni.** Il numero massimo di iterazioni che verrà eseguito dall'algoritmo delle equazioni di stima generalizzate. Specifica un intero non negativo. Questa specifica si applica ai parametri nella parte del modello lineare delle equazioni di stima generalizzate, mentre la specifica nella scheda [Stima](#) si applica soltanto al modello lineare generalizzato iniziale.
- **Aggiorna matrice.** Gli elementi nella matrice di correlazione di lavoro vengono stimati in base alle stime dei parametri, che vengono aggiornate al termine di ciascuna iterazione dell'algoritmo. Se la matrice di correlazione di lavoro non viene aggiornata, il processo di stima utilizza la matrice di correlazione di lavoro iniziale. Se la matrice viene aggiornata, è possibile specificare l'intervallo delle iterazioni in cui devono essere aggiornati gli elementi della matrice di correlazione di lavoro. L'impostazione di un valore superiore a 1 consente di ridurre i tempi di elaborazione.

**Criteri di convergenza.** Queste specifiche si applicano ai parametri nella parte del modello lineare delle equazioni di stima generalizzate, mentre la specifica nella scheda [Stima](#) si applica soltanto al modello lineare generalizzato iniziale.

- **Convergenza 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.
- **Convergenza hessiana.** La convergenza viene presunta se la statistica basata sulla statistica hessiana è inferiore al valore specificato che deve essere positivo.

## Tipo di modello per le equazioni di stima generalizzate

Figura 7-2  
Equazioni di stima generalizzate: Scheda Tipo di modello

La scheda Tipo di modello consente di specificare la distribuzione e la funzione di collegamento per il proprio modello e offre collegamenti per vari modelli comuni che sono classificati per tipo di risposta.

### Tipi di modello

#### Risposta scala.

- **Lineare.** Specifica Normale come distribuzione e Identità come funzione di collegamento.
- **Gamma con collegamento log.** Specifica Gamma come distribuzione e Log come funzione di collegamento.

#### Risposta ordinale.

- **Logistica ordinale.** Specifica Multinomiale (ordinali) come distribuzione e Logit cumulativa come funzione di collegamento.
- **Probit ordinale.** Specifica Multinomiale (ordinali) come distribuzione e Probit cumulativa come funzione di collegamento.

#### Conteggi.

- **Loglineare Poisson.** Specifica Poisson come distribuzione e Log come funzione di collegamento.
- **Binomiale negativa con collegamento log.** Specifica una distribuzione negativa binomiale (con un valore di 1 per il parametro ausiliario) come distribuzione e Log come funzione di collegamento. Affinché, tramite la procedura, venga stimato il valore del parametro ausiliario, specificare un modello personalizzato con distribuzione binomiale negativa e selezionare Stima valore nel gruppo dei parametri.

#### Risposta binaria o dati eventi/prove.

- **Logistica binaria.** Specifica Binominale come distribuzione e Logit come funzione di collegamento.
- **Probit binaria.** Specifica Binominale come distribuzione e Probit come funzione di collegamento.
- **Sopravvivenza di censura per intervallo.** Specifica Binominale come distribuzione e Log-log complementare come funzione di collegamento.

#### Combinazione.

- **Tweedie con collegamento log.** Specifica Tweedie come distribuzione e Log come funzione di collegamento.
- **Tweedie con collegamento identità.** Specifica Tweedie come distribuzione e Identità come funzione di collegamento.

**Personalizzata.** Specificare la propria combinazione di funzioni di distribuzione e di collegamento.

#### Distribuzione

Questa selezione specifica la distribuzione della variabile dipendente. La possibilità di specificare una distribuzione non normale e una funzione di collegamento senza identità è uno dei principali vantaggi offerti dal modello lineare generalizzato rispetto a quello generale. Dal momento che è possibile combinare più distribuzioni e funzioni di collegamento e che molte di queste sono adatte a qualsiasi combinazione di dati, è generalmente consigliabile fare una valutazione teorica a priori oppure selezionare la combinazione che si ritiene possa essere più adatta.

- **Binomiale.** Questa distribuzione è indicata solo per le variabili che rappresentano una risposta binaria o un numero di eventi.
- **Gamma.** Questa distribuzione è indicata per le variabili con valori di scala positivi che presentano una distribuzione asimmetrica in direzione dei valori positivi più alti. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.

- **Gaussiana inversa.** Questa distribuzione è indicata per le variabili con valori di scala positivi che presentano una distribuzione asimmetrica in direzione dei valori positivi più alti. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Binomiale negativo.** Questa distribuzione può essere considerata come il numero di prove necessario per osservare  $k$  successi ed è indicata per variabili con valori interi non negativi. Se il valore dei dati è un numero non intero, inferiore a 0 o mancante, il caso corrispondente non viene usato nell'analisi. Il valore del parametro ausiliario della distribuzione binomiale negativa può essere qualsiasi numero maggiore o uguale a 0; è possibile impostarlo su un valore fisso o consentirne la stima tramite la procedura. Se il parametro ausiliario viene impostato su 0, il risultato di questa distribuzione è uguale a quello della distribuzione di Poisson.
- **Normale.** Questa distribuzione è indicata per le variabili di scala i cui valori presentano una distribuzione simmetrica a forma di campana intorno al valore centrale (medio). La variabile dipendente deve essere numerica.
- **Poisson.** Questa distribuzione può essere considerata equivalente al numero di occorrenze di un evento desiderato in un intervallo di tempo fisso ed è indicata per le variabili con valori interi non negativi. Se il valore dei dati è un numero non intero, inferiore a 0 o mancante, il caso corrispondente non viene usato nell'analisi.
- **Tweedie.** Questa distribuzione è indicata per le variabili che possono essere rappresentate da combinazioni di Poisson di distribuzioni gamma; la distribuzione è "mista" nel senso che combina le proprietà di distribuzioni continue (assumono valori reali non negativi) e discrete (massa di probabilità positiva con un unico valore, 0). La variabile dipendente deve essere numerica, con i valori dei dati maggiori o uguali a zero. Se un valore dei dati è minore di zero o mancante, il caso corrispondente non viene usato nell'analisi. Il valore fisso del parametro della distribuzione Tweedie può essere qualsiasi numero maggiore di uno e minore di due.
- **Multinomiale.** Questa distribuzione è indicata per le variabili che rappresentano una risposta ordinale. La variabile dipendente può essere di tipo numerico o stringa e deve avere almeno due valori di dati distinti validi.

### **Funzione di collegamento**

La funzione Collegamento è la trasformazione della variabile dipendente che permette di stimare il modello. Sono disponibili le seguenti funzioni:

- **Identità.**  $f(x)=x$ . La variabile dipendente non viene trasformata. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.
- **Log-log complementare.**  $f(x)=\log(-\log(1-x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Cauchit cumulativa.**  $f(x) = \tan(\pi (x - 0.5))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Log-log complementare cumulativa.**  $f(x)=\ln(-\ln(1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Logit cumulativa.**  $f(x)=\ln(x / (1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.

- **Log-log negativa cumulativa.**  $f(x)=-\ln(-\ln(x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Probit cumulativa.**  $f(x)=\Phi^{-1}(x)$ , applicata alla probabilità cumulata di ciascuna categoria della risposta, dove  $\Phi^{-1}$  è la funzione di distribuzione cumulata normale standard inversa. Questa funzione è indicata solo per la distribuzione multinomiale.
- **Log.**  $f(x)=\log(x)$ . Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.
- **Complemento log.**  $f(x)=\log(1-x)$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Binomiale negativa.**  $f(x)=\log(x / (x+k^{-1}))$ , dove  $k$  è il parametro ausiliario della distribuzione binomiale negativa. Questa funzione è indicata solo per la distribuzione binomiale negativa.
- **Log-log negativo.**  $f(x)=-\log(-\log(x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Potenza odd.**  $f(x)=[(x/(1-x))^\alpha-1]/\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  è la specifica del numero richiesto e deve essere un numero reale. Questa funzione è indicata solo per la distribuzione binomiale.
- **Probit.**  $f(x)=\Phi^{-1}(x)$ , dove  $\Phi^{-1}$  è la funzione di distribuzione cumulata normale standard inversa. Questa funzione è indicata solo per la distribuzione binomiale.
- **Potenza.**  $f(x)=x^\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  è la specifica del numero richiesto e deve essere un numero reale. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni.

## Risposta delle equazioni di stima generalizzate

Figura 7-3  
Equazioni di stima generalizzate: scheda Risposta

In molti casi è possibile specificare semplicemente una variabile dipendente. Tuttavia, le variabili che possono assumere solo due valori e le risposte che registrano gli eventi durante le prove richiedono più attenzione.

- Risposta binaria.** Se la variabile dipendente può assumere solo due valori, è possibile specificare la [categoria di riferimento](#) per la stima dei parametri. Le variabili di risposta binaria possono essere di tipo stringa o numeriche.
- Numero di eventi occorsi in un insieme di prove.** Se la risposta rappresenta il numero di eventi di un insieme di prove, la variabile dipendente contiene il numero di eventi ed è possibile selezionare un'ulteriore variabile che contenga il numero di prove. In alternativa, se il numero di prove è lo stesso per tutti i soggetti, è possibile specificare le prove utilizzando un valore fisso. Il numero di prove deve essere maggiore o uguale al numero di eventi di ciascun caso. Gli eventi devono essere numeri interi non negativi, mentre le prove devono essere numeri interi positivi.

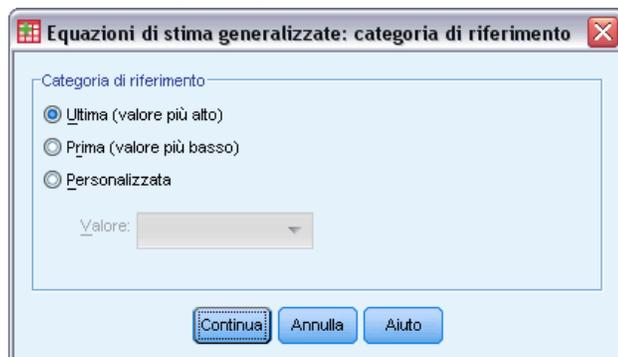
Per i modelli multinomiali ordinali è possibile specificare l'ordine delle categorie della risposta, che può essere: crescente, decrescente o dati. L'ordine dati significa che il primo valore trovato nei dati definisce la prima categoria e l'ultimo valore trovato definisce l'ultima categoria.

**Peso della scala.** Il parametro scala è un parametro del modello stimato correlato alla varianza della risposta. I pesi della scala sono valori "noti" che possono variare a seconda delle osservazioni. Se la variabile del peso della scala è stata specificata, il parametro scala, che è correlato alla varianza della risposta, viene diviso per la suddetta variabile per ciascuna osservazione. Per l'analisi non vengono usati i casi con valori di pesi scala inferiori o uguali a 0 o mancanti.

### ***Categoria di riferimento delle equazioni di stima generalizzate***

Figura 7-4

Finestra di dialogo *Categoria di riferimento per le equazioni di stima generalizzate*

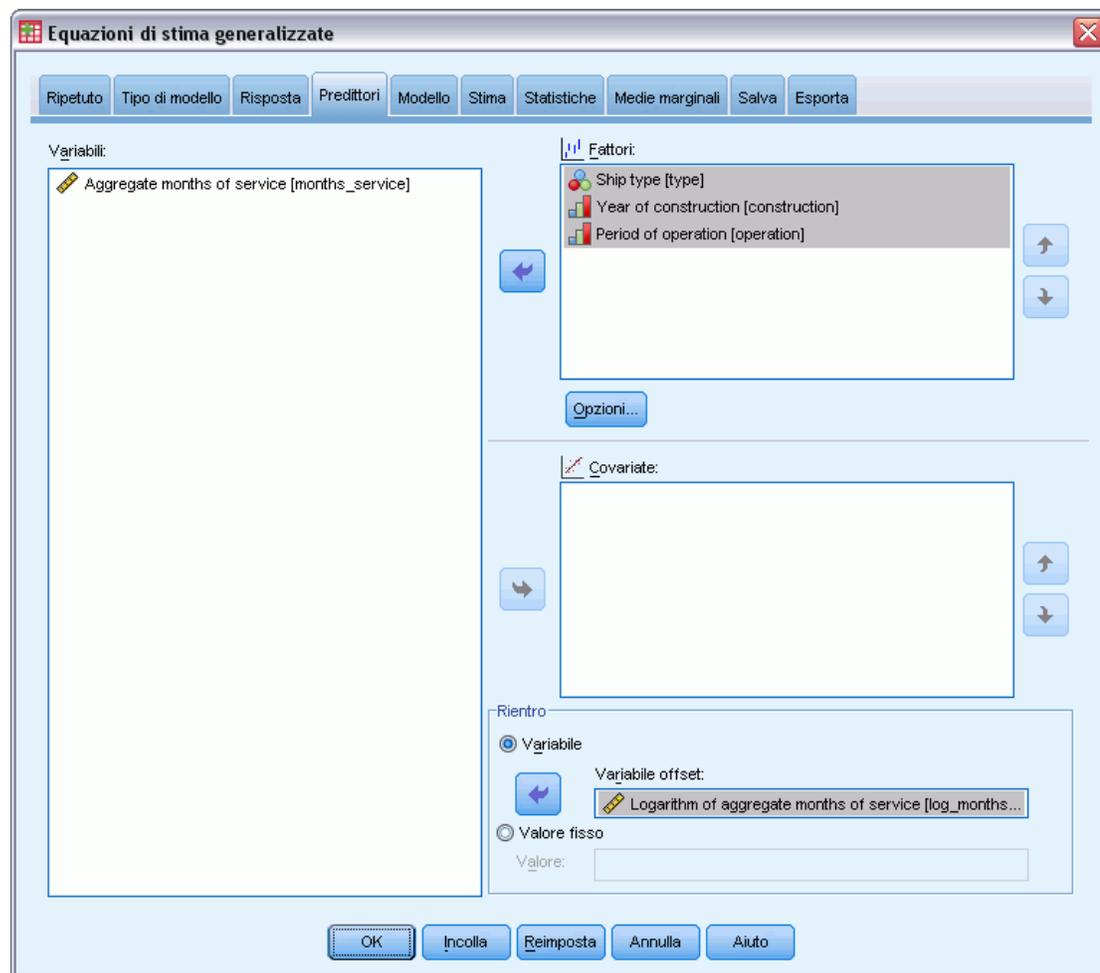


Per la risposta binaria, è possibile scegliere la categoria di riferimento per la variabile dipendente. Questa operazione può influire su un determinato output, ad esempio quello delle stime dei parametri e dei valori salvati, ma non dovrebbe modificare l'adattamento del modello. Ad esempio, se la risposta binaria assume i valori 0 e 1:

- Per impostazione predefinita, la procedura imposta l'ultima categoria (dal valore più alto), o 1, come categoria di riferimento. In questa situazione, le probabilità salvate del modello stimano la probabilità che un determinato caso assuma il valore 0, mentre le stime dei parametri dovrebbero essere interpretate come relative alla verosimiglianza della categoria 0.
- Se si specifica la prima categoria (dal valore più basso), oppure 0, come categoria di riferimento, le probabilità salvate del modello stimano la probabilità che un determinato caso assuma il valore 1.
- Se si specifica la categoria personalizzata e la variabile ha delle etichette definite, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Questo può risultare conveniente quando, durante un'operazione di specifica di un modello, non ci si ricorda esattamente del modo in cui è stata codificata una determinata variabile.

## Predittori delle equazioni di stima generalizzate

Figura 7-5  
Equazioni di stima generalizzate: scheda Predittori



La scheda Predittori consente di specificare i fattori e le covariate usati per definire gli effetti del modello nonché di definire un offset opzionale.

**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.

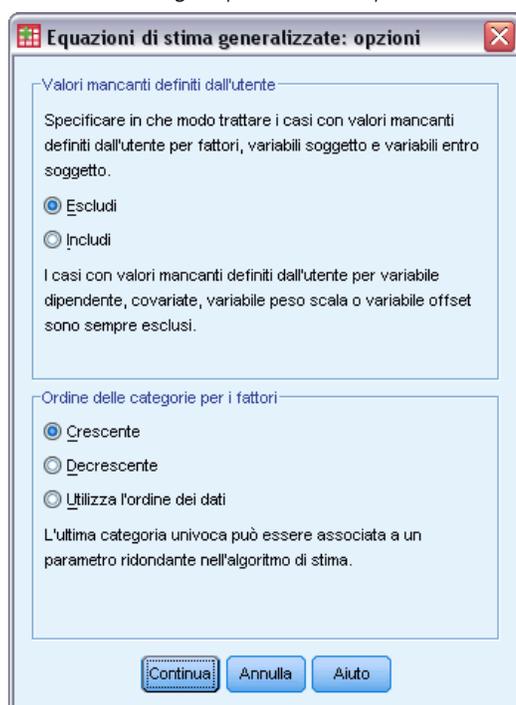
*Nota:* se la risposta è binomiale ed ha un formato binario, la procedura calcola le statistiche di devianza e di bontà d'adattamento chi-quadrato in base alle sottopopolazioni, che si basano sulla classificazione incrociata dei valori osservati per i fattori e le covariate selezionati. È generalmente consigliabile usare lo stesso insieme di predittori nelle varie esecuzioni della procedura per mantenere un numero di sottopopolazioni uniforme.

**Offset.** Questo termine identifica un predittore “strutturale”. Il suo coefficiente non viene stimato utilizzando dal modello ma ritenuto sempre equivalente a 1. Quindi, i valori dell'offset vengono semplicemente aggiunti al predittore lineare della variabile dipendente. Ciò è particolarmente

utile nei modelli di regressione di Poisson, in cui ciascun caso può avere vari livelli di esposizione all'evento desiderato. Ad esempio, quando si creano modelli per calcolare la percentuale di incidenti per i singoli conducenti, esiste una sostanziale differenza tra il conducente che ha avuto un incidente in tre anni di guida e uno che ne ha avuto uno in 25 anni. È possibile creare un modello del numero di incidenti come risposta di Poisson, se l'esperienza del conducente è inclusa come termine di offset.

## Opzioni per le equazioni di stima generalizzate

Figura 7-6  
Finestra di dialogo Opzioni delle equazioni di stima generalizzate



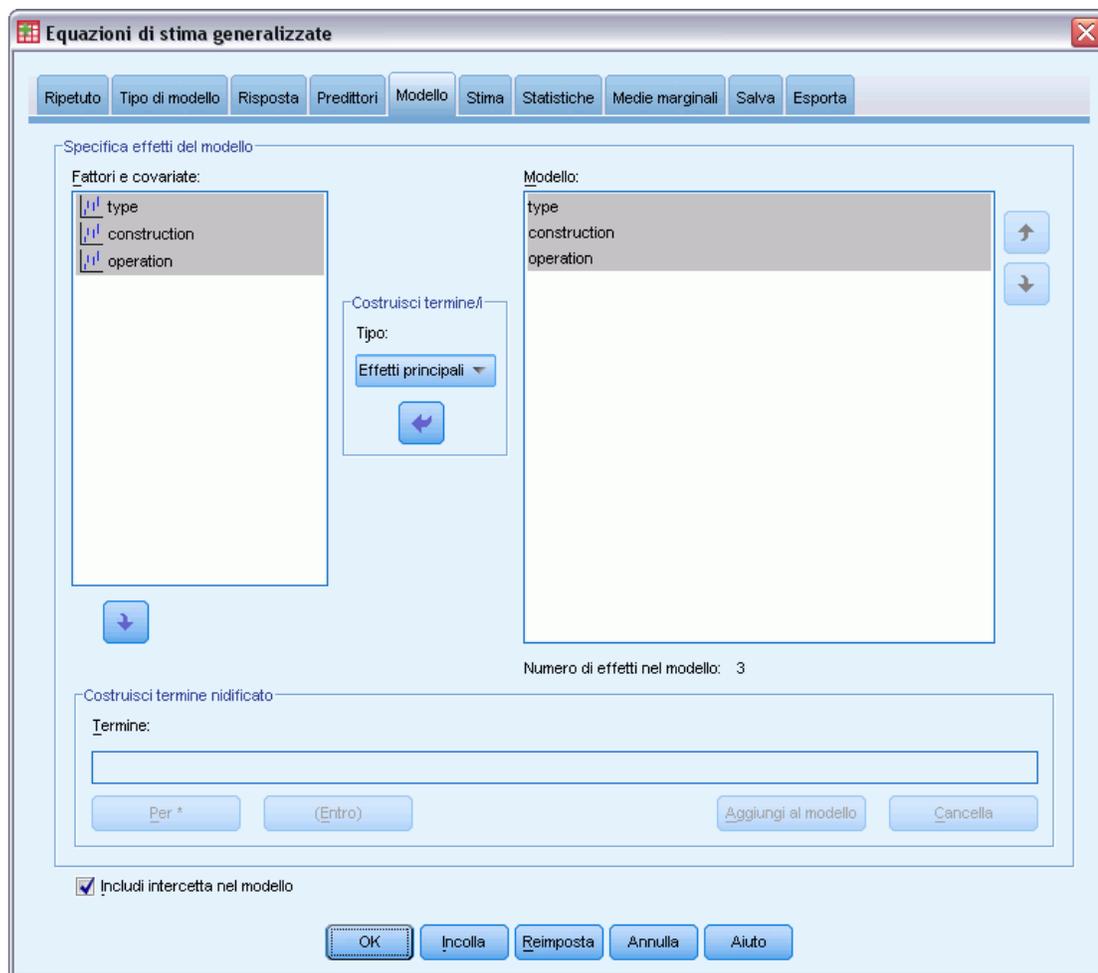
Queste opzioni vengono applicate a tutti i fattori selezionati nella scheda Predittori.

**Valori mancanti definiti dall'utente.** Affinché un caso possa essere incluso nell'analisi, è necessario che i fattori abbiano valori validi. Questi controlli consentono di decidere se i valori mancanti definiti dall'utente devono essere considerati validi nell'ambito delle variabili dei fattori.

**Ordine delle categorie.** Questa opzione è importante per stabilire l'ultimo livello di un fattore che può essere associato a un parametro ridondante nell'algoritmo di stima. La modifica dell'ordine delle categorie influisce sui valori degli effetti a livello di fattore poiché le stime di questi parametri vengono calcolate in funzione dell'“ultimo” livello. I fattori possono essere disposti in ordine crescente (dal valore più basso a quello più alto), in ordine decrescente (dal valore più alto a quello più basso) oppure in base all'ordine dei dati. Ciò significa che il primo valore trovato nei dati definisce la prima categoria, e l'ultimo valore univoco trovato definisce l'ultima categoria.

## Modello per le equazioni di stima generalizzate

Figura 7-7  
Equazioni di stima generalizzate: scheda Modello



**Specifica modello effetti.** Il modello predefinito prevede solo l'uso dell'intercetta, quindi è necessario specificare esplicitamente altri effetti per il modello. In alternativa, è possibile creare termini nidificati o non nidificati.

### **Termini non nidificati**

Per i fattori e le covariate selezionati:

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

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

**Fattoriale.** Consente di creare tutte le possibili interazioni e gli effetti principali delle variabili selezionate.

**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 inoltre possibile includere effetti di interazione o 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'intercetta può essere esclusa.

I modelli con la distribuzione ordinale multinomiale non dispongono di un singolo termine di intercetta; invece, vi sono parametri di soglia che definiscono i punti di transizione tra categorie adiacenti. Le soglie vengono sempre incluse nel modello.

## Stima delle equazioni di stima generalizzate

Figura 7-8  
Equazioni di stima generalizzate: scheda Stima

**Stima dei parametri.** I controlli di questo gruppo consentono di specificare i metodi di stima e di definire i valori iniziali per la stima dei parametri.

- **Metodo.** 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.
- **Metodo basato sul parametro scala.** È possibile selezionare il metodo di stima basato sui parametri scala.

La massima verosimiglianza stima i parametri di scala utilizzando gli effetti del modello, benché questa opzione non sia valida se la risposta ha una distribuzione binomiale negativa, di Poisson o binomiale. Poiché il concetto di verosimiglianza non si applica alle equazioni di

stima generalizzate, questa specifica si applica solo al modello lineare generalizzato iniziale; la stima del parametro scala viene quindi passata alle equazioni di stima generalizzate che aggiornano il parametro scala utilizzando il chi-quadrato di Pearson diviso per i gradi di libertà corrispondenti.

Le opzioni della devianza e del chi-quadrato di Pearson stimano il parametro scala utilizzando il valore delle statistiche del modello lineare generalizzato iniziale; la stima del parametro scala viene quindi passata alle equazioni di stima generalizzate che la considerano come una stima fissa.

In alternativa, è possibile anche specificare un valore fisso per il parametro scala. In questo caso tale valore viene considerato come valore fisso durante la stima del modello lineare generalizzato iniziale e nelle equazioni di stima generalizzate.

- **Valori iniziali.** La procedura calcola automaticamente i valori iniziali dei parametri. In alternativa, è possibile specificare i [valori iniziali](#) per le stime del parametro.

I criteri di iterazione e convergenza specificati in questa scheda si applicano solo al modello lineare generalizzato iniziale. Per informazioni sui criteri di stima usati per l'adattamento delle equazioni di stima generalizzate, vedere la scheda [Ripetuto](#).

#### Iterazioni.

- **Max 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.
- **Verificare la 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. Questa opzione è disponibile per risposte multinomiali e risposte binomiali con formato binario.

#### Criteri di convergenza.

- **Convergenza 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.
- **Convergenza 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.
- **Convergenza hessiana.** In caso di specifica Assoluta, la convergenza viene presunta se la statistica basata sulla convergenza hessiana è inferiore al valore positivo specificato. Per la specifica relativa, la convergenza viene presunta se la statistica è inferiore al prodotto del valore positivo specificato e del valore assoluto del logaritmo di verosimiglianza.

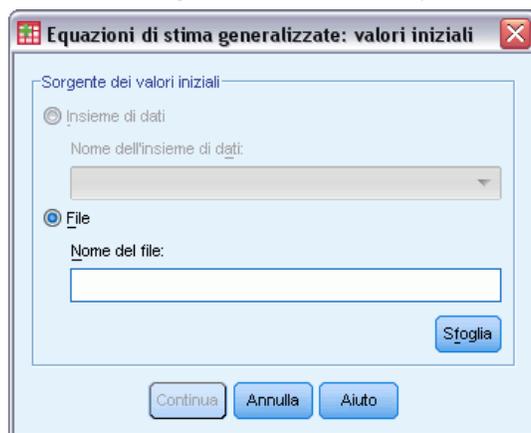
**Tolleranza della singolarità.** Le matrici singole (o non reversibili) hanno colonne linearmente dipendenti che possono creare gravi problemi all'algoritmo di stima. Poiché anche le matrici quasi singolari possono generare risultati inadeguati, la procedura considera le matrici i cui determinanti sono inferiori alla tolleranza come singolari. Specificare un valore positivo.

## Valori iniziali delle equazioni di stima generalizzate

La procedura stima un modello lineare generalizzato iniziale e le stime risultanti da questo modello vengono usate come valori iniziali per le stime dei parametri nella parte del modello lineare delle equazioni di stima generalizzate. La matrice di correlazione di lavoro non ha bisogno dei valori iniziali perché gli elementi della matrice sono basati sulle stime dei parametri. I valori iniziali specificati in questa finestra di dialogo vengono usati come punto di partenza per il modello lineare generalizzato, ma non per le equazioni di stima generalizzate a meno che il numero massimo di iterazioni nella scheda **Stima** non sia impostato su 0.

Figura 7-9

Finestra di dialogo Valori iniziali delle equazioni di stima generalizzate



Se si specificano valori iniziali, è necessario indicarli per tutti i parametri del modello, compresi quelli ridondanti. Nell'insieme dei dati, l'ordine delle variabili da sinistra a destra deve essere: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., dove *RowType\_* e *VarName\_* sono variabili di tipo stringa e *P1*, *P2*, ... sono variabili numeriche che corrispondono a un elenco ordinato dei parametri.

- I valori iniziali vengono forniti in un record con valore *EST* per la variabile *RowType\_*; i valori iniziali effettivi sono specificati nelle variabili *P1*, *P2*, .... La procedura ignora tutti i record per i quali *RowType\_* ha un valore diverso da *EST*, nonché qualsiasi record oltre la prima occorrenza di *RowType\_* uguale a *EST*.
- L'intercetta, se inclusa nel modello, o i parametri di soglia, se la risposta ha una distribuzione multinomiale, devono essere i primi valori iniziali elencati.
- Il parametro di scala *e*, se la risposta ha una distribuzione binomiale negativa, il parametro binomiale negativo, devono essere gli ultimi valori iniziali specificati.
- Se è stata abilitata l'opzione *Dividi*, le variabili devono iniziare con le variabili dei file, suddivise nell'ordine specificato durante la creazione del file suddiviso, seguite da *RowType\_*, *VarName\_*, *P1*, *P2*, ... come indicato in precedenza. Le divisioni nell'insieme di dati specificato devono verificarsi nello stesso ordine in cui si verificano nell'insieme di dati originale.

*Nota:* i nomi delle variabili *P1*, *P2*, ... non sono obbligatori; la procedura accetta qualsiasi nome di variabile valido per i parametri poiché l'associazione delle variabili ai parametri viene effettuato in base alla posizione della variabile, non in base al nome. Tutte le variabili successive all'ultimo parametro vengono ignorate.

La struttura del file per i valori iniziali è la stessa di quella usata per l'esportazione del modello sotto forma di dati; quindi, è possibile usare i valori finali di un'esecuzione precedente della procedura come input per un'esecuzione successiva.

## Statiche delle equazioni di stima generalizzate

Figura 7-10  
Equazioni di stima generalizzate: scheda Statistiche

### Effetti del modello.

- Tipo di analisi.** Specificare il tipo di analisi che si desidera ottenere per verificare gli effetti del modello. Le analisi di tipo I sono generalmente indicate nei casi in cui ci siano ragioni a priori per ordinare i predittori nel modello, mentre le analisi di tipo III sono applicabili a contesti più generali. Le statistiche di Wald o le statistiche punteggio generalizzato vengono calcolate in base alla selezione effettuata nel gruppo Statistiche chi-quadrato.

- **Intervalli di confidenza.** Specificare un livello di confidenza superiore a 50 e inferiore a 100. Gli intervalli di Wald vengono sempre generati indipendentemente dal tipo di statistiche chi-quadrato selezionate e si basano sul presupposto che i parametri abbiano una distribuzione normale asintotica.
- **Funzione della quasi-verosimiglianza log.** Controlla il formato di visualizzazione della funzione della quasi-verosimiglianza log. La funzione completa comprende un ulteriore termine che è costante rispetto alle stime dei parametri; non ha alcun effetto sulla stima dei parametri e viene escluso dalla visualizzazione in alcuni prodotti software.

**Stampa.** L'output disponibile è descritto di seguito.

- **Riepilogo dell'elaborazione dei casi.** Visualizza il numero e la percentuale di casi inclusi ed esclusi nella e dall'analisi e la tabella di riepilogo dei dati correlati.
- **Statistiche descrittive.** Visualizza le statistiche descrittive e le informazioni di riepilogo sulla variabile dipendente, sulle covariate e sui fattori.
- **Informazioni sul modello.** Visualizza il nome dell'insieme dei dati, della variabile dipendente, degli eventi, delle variabili delle prove, della variabile di offset, della variabile dei pesi della scala, della distribuzione di probabilità e della funzione di collegamento.
- **Statistiche di bontà dell'adattamento.** Visualizza le due estensioni del criterio di informazione di Akaike per la selezione del modello: la misura di quasi-simiglianza per il criterio del modello indipendente (QIC) che consente di selezionare la migliore struttura di correlazione e un'altra misura QIC che consente di scegliere il miglior sottoinsieme di predittori.
- **Statistiche riassuntive del modello.** Visualizza i test di adattamento del modello comprese le statistiche del rapporto di verosimiglianza per il test di adattamento del modello Omnibus e le statistiche per i contrasti di tipo I o II per ciascun effetto.
- **Stime dei parametri.** Visualizza le stime dei parametri, le statistiche corrispondenti e gli intervalli di confidenza. È possibile anche visualizzare le stime dei parametri elevati a potenza oltre alle stime dei parametri grezzi.
- **Matrice di covarianza per le stime dei parametri.** Visualizza la matrice di covarianza residua dei parametri stimata.
- **Matrice di correlazione per le stime dei parametri.** Visualizza la matrice di correlazione dei parametri stimata.
- **Matrici dei coefficienti di contrasto (L).** Visualizza i coefficienti di contrasto per gli effetti predefiniti e per le medie marginali stimate se si seleziona l'opzione corrispondente nella scheda Medie marginali.
- **Funzioni funzionali generali.** Visualizza le matrici per la generazione delle matrici dei coefficienti di contrasto (L).
- **Cronologia iterazioni.** Visualizza la cronologia delle iterazioni per le stime dei parametri e la verosimiglianza e stampa l'ultima valutazione del vettore del gradiente e la matrice hessiana. La tabella Cronologia iterazioni visualizza le stime dei parametri per tutte le  $n^{\circ}$  iterazioni a partire dalla 0-esima iterazione (stime iniziali), dove  $n$  è il valore dell'intervallo di stampa. Se la cronologia delle iterazioni è richiesta, l'ultima iterazione viene sempre visualizzata indipendentemente da  $n$ .
- **Matrice di correlazione di lavoro.** Visualizza i valori della matrice che rappresenta le dipendenze entro i soggetti. La sua struttura dipende dalle specifiche presenti nella scheda [Ripetuto](#).

## Medie marginali stimate delle equazioni di stima generalizzate

Figura 7-11  
Equazioni di stima generalizzate: scheda Medie marginali

Questa scheda consente di visualizzare le medie marginali stimate per i livelli e le interazioni tra fattori. È possibile anche visualizzare la media stimata globale. Le medie marginali stimate non sono disponibili per i modelli multinomiali ordinali.

**Fattori e interazioni.** Questo elenco contiene i fattori specificati nella scheda Predittori e le interazioni tra fattori specificate nella scheda Modello. L'elenco non comprende le covariate. I termini possono essere selezionati direttamente dall'elenco oppure combinati in un termine di interazione utilizzando il pulsante Per \*.

**Medie marginali per.** Le medie stimate vengono calcolate per i fattori e le interazioni tra fattori selezionate. Il contrasto determina la modalità di impostazione dei test di ipotesi per il confronto delle medie stimate. Per il contrasto semplice è necessaria una categoria di riferimento o un livello di fattore da usare per il confronto.

- **Paiwise (Analisi dei dati mancanti).** I confronti a coppie vengono calcolati per le combinazioni dei fattori specificati o implicati di tutti i livelli. Questa è l'unico contrasto disponibile per le interazioni tra fattori.
- **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.** Ciascun livello di fattore viene confrontato con la media principale. I contrasti di deviazione non sono ortogonali.
- **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.
- **Ripetuto.** 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.

**Scala.** È possibile calcolare le medie marginali stimate per la risposta o il predittore lineare, utilizzando rispettivamente la scala originale della variabile indipendente e la variabile dipendente trasformata dalla funzione di collegamento.

**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.
- **Bonferroni.** Questo metodo corregge il livello di significatività osservato tenendo conto del fatto che si stanno verificando contrasti multipli.
- **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.
- **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.

## Salvataggio delle equazioni di stima generalizzate

Figura 7-12  
Equazioni di stima generalizzate: scheda Salva

Salva	Elemento da salvare	Nome variabile o nome radice	Categorie da salvare
<input type="checkbox"/>	Valore atteso della media della risposta	MeanPredicted	
<input type="checkbox"/>	Limite inferiore dell'intervallo di confidenza per la media della risposta	CI-MeanPredicted-Lower	
<input type="checkbox"/>	Limite superiore dell'intervallo di confidenza per la media della risposta	CI-MeanPredicted-Upper	
<input checked="" type="checkbox"/>	Categoria prevista	Valoreprevisto	
<input type="checkbox"/>	Valore atteso del predittore lineare	XBPredicted	
<input type="checkbox"/>	Errore standard stimato del valore atteso del predittore lineare	XBStandardError	
<input type="checkbox"/>	Residuo	Residuo	
<input type="checkbox"/>	Residuo di Pearson	PearsonResidual	

Variabile esistente con lo stesso nome

Aggiungi suffisso al nome della nuova variabile (applicabile solo ai nomi predefiniti)

Sostituisci variabile esistente (applicabile ai nomi predefiniti e definiti dall'utente)

Se si forniscono nomi specifici per le variabili, assicurarsi che non siano in conflitto con le variabili esistenti nell'insieme di dati attivo. Selezionare l'opzione Sostituisci variabile esistente per utilizzare i nomi forniti dall'utente per sostituire le variabili esistenti.

OK Incolla Reimposta Annulla Aiuto

Gli elementi selezionati vengono salvati con il nome specificato; è possibile scegliere di sovrascrivere le variabili esistenti con lo stesso nome oppure evitare possibili conflitti aggiungendo suffissi che rendano i nomi delle nuove variabili univoci.

- Valore atteso della media della risposta.** Salva i valori previsti per il modello per ciascun caso presente nella metrica di risposta originale. Se la distribuzione della risposta è binomiale e la variabile dipendente è binaria, la procedura salva le probabilità previste. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Probabilità prevista cumulativa e la procedura salva la probabilità prevista cumulativa per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- Lower bound of confidence interval for mean of response.** Salva il limite inferiore dell'intervallo di confidenza per la media della risposta. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Correzione inferiore dell'intervallo di confidenza per la probabilità prevista cumulativa e la procedura salva il limite inferiore per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.

- **Upper bound of confidence interval for mean of response.** Salva il limite superiore dell'intervallo di confidenza per la media della risposta. Se la distribuzione della risposta è multinomiale, l'etichetta dell'elemento diventa Correzione superiore dell'intervallo di confidenza per la probabilità prevista cumulativa e la procedura salva il limite superiore per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- **Categoria prevista.** Per i modelli con la distribuzione binomiale e la variabile dipendente binaria oppure con la distribuzione multinomiale, salva la categoria della risposta prevista per ogni caso. Questa opzione non è disponibile per altre distribuzioni della risposta.
- **Valore atteso del predittore lineare.** Salva i valori previsti per il modello per ciascun caso nella metrica nel predittore lineare (risposta trasformata tramite la funzione di collegamento specifica). Se la distribuzione della risposta è multinomiale, la procedura salva il valore previsto per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.
- **Errore standard stimato del valore atteso del predittore lineare.** Se la distribuzione della risposta è multinomiale, la procedura salva l'errore standard stimato per ciascuna categoria della risposta, ad eccezione dell'ultima, fino al numero delle categorie specificate da salvare.

I seguenti elementi non sono disponibili quando la distribuzione della risposta è multinomiale.

- **residuo grezzo.** La differenza tra un valore osservato e il valore stimato dal modello.
- **Residuo di Pearson.** Radice quadrata del contributo di un caso alla statistica chi-quadrato di Pearson con il segno del residuo grezzo.

## Esportazione delle equazioni di stima generalizzate

Figura 7-13  
Equazioni di stima generalizzate: scheda Esporta

**Esporta modello come dati.** 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.

- **Variabili di distinzione.** Se utilizzata, qualsiasi variabile definisce le divisioni.
- **RowType\_** Assume valori (ed etichette di 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 stimati (ad eccezione dei parametri di scala o dei parametri binomiali negativi), per i tipi di riga *COV* or *CORR*, con etichette dei valori corrispondenti alle stringhe

dei parametri mostrate nella tabella 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 (inclusi i parametri di scala e i parametri binomiali negativi, se necessario), con etichette delle variabili corrispondenti alle stringhe dei parametri mostrate nella tabella 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.

Per il parametro di scala, le covarianze, le correlazioni, il livello di significatività e i gradi di libertà sono impostati sul valore mancante di sistema. Se il parametro di scala viene stimato tramite la massima verosimiglianza, viene generato l'errore standard; altrimenti, tale parametro viene impostato sul valore mancante di sistema.

Per il parametro binomiale negativo, le covarianze, le correlazioni, il livello di significatività e i gradi di libertà sono impostati sul valore mancante di sistema. Se il parametro binomiale negativo viene stimato tramite la massima verosimiglianza, viene generato l'errore standard; altrimenti, tale parametro viene impostato sul valore mancante di sistema.

Se vi sono divisioni, l'elenco di parametri deve essere accumulato in tutte le divisioni. In una determinata divisione, alcuni parametri possono risultare irrilevanti; cosa ben diversa dai parametri ridondanti. Per i parametri irrilevanti, tutte le covarianze o le correlazioni, le stime dei parametri, gli errori standard, i livelli di significatività e i gradi di libertà sono impostati sul valore mancante di sistema.

Questo file matrice può essere usato per definire i valori iniziali e stimare altri modelli. Si noti che questo file non può essere immediatamente usato 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. Nonostante ciò, è necessario accertarsi che tutti i parametri presenti in questo file di matrice abbiano lo stesso significato per la procedura che esegue la lettura del file.

**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.

## ***Funzioni aggiuntive del comando GENLIN***

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i valori iniziali per le stime dei parametri come elenco di numeri (utilizzando il sottocomando `CRITERIA`).
- Specificare una matrice di correlazione di lavoro fissa (utilizzando il sottocomando `REPEATED`).
- Fissare le covariate sui valori diversi dalle relative medie durante il calcolo delle medie marginali stimate (utilizzando il sottocomando `EMMEANS`).

- Specificare contrasti polinomiali personalizzati per le medie marginali stimate (utilizzando il sottocomando `EMMEANS`).
- Specificare un sottoinsieme dei fattori che consente di visualizzare le medie marginali stimate e di confrontarle utilizzando il tipo di contrasto specificato (utilizzando le parole chiave `TABLES` e `COMPARE` del sottocomando `EMMEANS`).

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

# Modelli misti lineari generalizzati

I modelli misti lineari generalizzati estendono il modello lineare in modo che:

- L'obiettivo venga linearmente correlato ai fattori e alle covariate tramite una funzione di collegamento specifica
- L'obiettivo possa avere una distribuzione non normale
- Le osservazioni possano essere correlate.

I modelli misti lineari generalizzati includono un'ampia varietà di modelli, dai modelli di regressione lineare semplice ai modelli complessi a più livelli per dati longitudinali non normali.

Figura 8-1  
Scheda *Struttura dati*

La scheda Struttura dati consente di specificare le relazioni strutturali tra i record nell'insieme di dati quando le osservazioni sono correlate. Se i record nell'insieme di dati rappresentano osservazioni indipendenti, non è necessario specificare alcuna opzione in questa scheda.

**Soggetti.** La combinazione di valori dei campi categoriali specificati deve definire in modo univoco i soggetti nell'insieme di dati. Ad esempio, il campo singolo *ID paziente* è generalmente sufficiente per definire i pazienti di un ospedale. Tuttavia, se i codici di identificazione dei pazienti non sono univoci tra gli ospedali, può essere necessario usare la combinazione di variabili *ID ospedale* e *ID paziente*. Se sono previste più misurazioni, vengono registrate più osservazioni per ciascun soggetto, quindi è possibile che ciascun soggetto occupi più record nell'ambito dell'insieme dei dati.

Un **soggetto** è un'unità di osservazione che può essere considerata indipendente rispetto ad altri soggetti. Ad esempio, i risultati della misurazione della pressione sanguigna di un paziente in uno studio medico sono da considerarsi indipendenti rispetto alle misurazioni effettuate sugli altri pazienti. La definizione dei soggetti è particolarmente importante nel caso in cui esistano misurazioni ripetute per ciascun soggetto e si desidera determinare una correlazione tra queste osservazioni. Ad esempio, è possibile stabilire se esiste una correlazione tra le misurazioni della pressione sanguigna di un paziente rilevate nel corso di più visite mediche.

Tutti i campi specificati come Soggetti nella scheda Struttura dati vengono usati per definire i soggetti della struttura della covarianza residua e forniscono l'elenco dei campi possibili per la definizione dei soggetti per le strutture di covarianza a effetti casuali nel [Blocco effetti casuali](#).

**Misure ripetute.** I campi specificati in questo punto vengono utilizzati per identificare le osservazioni ripetute. Ad esempio, la variabile singola *Settimana* potrà identificare le 10 settimane di osservazione in uno studio medico, mentre le variabili *Mese* e *Giorno* utilizzate contemporaneamente potranno identificare le osservazioni giornaliere nel corso di un anno.

**Definisci gruppi covarianza per.** I campi specificati in questo punto definiscono insiemi indipendenti di parametri di covarianza a effetti ripetuti, uno per ogni categoria definita dalla classificazione incrociata dei campi di raggruppamento. Tutti i soggetti presentano lo stesso tipo di covarianza; i soggetti all'interno dello stesso gruppo di covarianza presentano gli stessi valori per i parametri.

**Tipo di covarianza ripetuta.** Questo tipo di covarianza specifica la struttura della covarianza per i residui. Le strutture disponibili sono:

- Autoregressivo di primo ordine (AR1)
- Autoregressivo a media mobile (1,1) (ARMA11)
- Simmetria composta
- Diagonale
- Identità scalata
- Toeplitz
- Non strutturato
- Componenti della varianza

Per ulteriori informazioni, vedere l'argomento [Strutture della covarianza](#) in l'appendice B a pag. 161.

## Per ottenere un modello misto lineare generalizzato

Questa funzione richiede il modulo Advanced Statistics.

Dai menu, scegliere:

Analizza > Modelli misti > Lineare generalizzato...

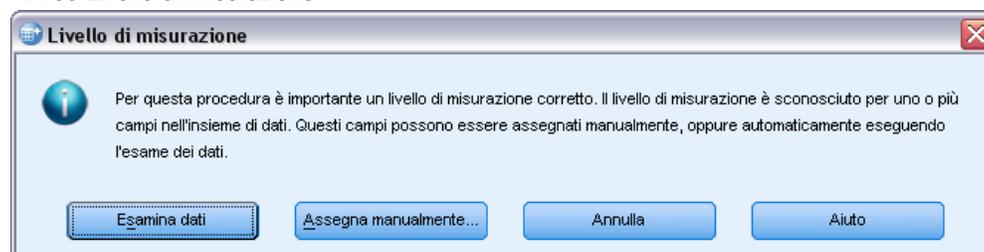
- ▶ Definire la struttura dei soggetti dell'insieme di dati nella scheda Struttura dati.
- ▶ Nella scheda Campi ed effetti deve essere presente un solo obiettivo, che può avere qualsiasi livello di misurazione, oppure una specifica eventi/prove; in tal caso, le specifiche di eventi e prove devono essere continue. Se lo si desidera, specificarne le funzioni di distribuzione e di collegamento, gli effetti fissi, eventuali blocchi di effetti casuali, le variabili offset o i pesi di analisi.
- ▶ Fare clic su Opzioni di costruzione per specificare le impostazioni facoltative per la costruzione.
- ▶ Fare clic su Opzioni modello per salvare i punteggi negli insiemi di dati attivi ed esportare il modello in un file esterno.
- ▶ Fare clic su Esegui per eseguire la procedura e creare gli oggetti Modello.

### Campi con livello di misurazione sconosciuto

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

Figura 8-2

Avviso Livello di misurazione



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

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

## Obiettivo

Figura 8-3  
Impostazioni obiettivo

Struttura dati | **Campi ed effetti** | Opzioni di costruzione | Opzioni modello

Selezione un elemento:

- Obiettivo
- Effetti fissi
- Effetti casuali
- Peso e offset

**Obiettivo**

Usa obiettivo predefinito     Usa obiettivo personalizzato

Obiettivo\*:

Altro

Usa numero di prove come denominatore     Personalizza categoria di riferimento

Usa campo:     Valore di riferimento:

Usa valore:

**Relazione e distribuzione obiettivo (Collegamento) con modello lineare**

Modello lineare     Regressione logistica multinomiale

Regressione gamma     Regressione logistica binomiale

Loglineare     Probit binaria

Regressione binomiale negativa     Sopravvivenza di censura per intervallo

Personalizzato

Distribuzione:     Funzione di collegamento:

Potenza:

**Descrizione**

Loglineare utilizza una distribuzione di Poisson con un collegamento log da utilizzare quando l'obiettivo rappresenta un conteggio di occorrenze in un determinato periodo di tempo.

Queste impostazioni definiscono l'obiettivo, la sua distribuzione e la sua relazione con i predittori attraverso la funzione di collegamento.

**Obiettivo.** L'obiettivo è obbligatorio. Può avere qualsiasi livello di misurazione e il livello di misurazione dell'obiettivo vincola le distribuzioni e le funzioni di collegamento appropriate.

- Usa numero di prove come denominatore.** Se la risposta obiettivo rappresenta il numero di eventi di un insieme di prove, il campo obiettivo contiene il numero di eventi ed è possibile selezionare un ulteriore campo che contenga il numero di prove. Ad esempio, quando si testa un nuovo pesticida è possibile esporre dei campioni di formiche a diverse concentrazioni di pesticida e quindi registrare il numero di formiche uccise e il numero di formiche esposte in ogni campione. In questo caso, il campo che registra il numero di formiche uccise deve essere specificato come campo obiettivo (eventi) e il campo che registra il numero di formiche in ogni campione deve essere specificato come campo prove. Se il numero di formiche è lo stesso per ogni campione, il numero di prove deve essere specificato utilizzando un valore fisso.

Il numero di prove deve essere maggiore o uguale al numero di eventi di ciascun record. Gli eventi devono essere numeri interi non negativi, mentre le prove devono essere numeri interi positivi.

- **Personalizza categoria di riferimento.** Per un obiettivo categoriale, è possibile scegliere la categoria di riferimento. Questa operazione può influire su un determinato output, ad esempio quello delle stime dei parametri, ma non dovrebbe modificare l'adattamento del modello. Ad esempio, se l'obiettivo assume i valori 0, 1 e 2, per impostazione predefinita, la procedura imposta l'ultima categoria (dal valore più alto), o 2, come categoria di riferimento. In questa situazione, le stime dei parametri devono essere interpretate come relative alla verosimiglianza della categoria 0 o 1 *relativa* alla verosimiglianza della categoria 2. Se si specifica una categoria personalizzata e l'obiettivo ha delle etichette definite, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Questo può risultare conveniente quando, durante un'operazione di specifica di un modello, non ci si ricorda esattamente del modo in cui è stato codificato un determinato campo.

**Relazione e distribuzione obiettivo (Collegamento) con modello lineare.** Dati i valori dei predittori, il modello prevede che la distribuzione dei valori dell'obiettivo segua la forma specificata e che i valori obiettivo siano linearmente correlati ai predittori attraverso la funzione di collegamento specificata. Sono disponibili dei collegamenti per vari modelli comuni, oppure è possibile scegliere un'impostazione Personalizzata se esiste una particolare combinazione delle funzioni di distribuzione e di collegamento che si desidera adattare e che non si trova nell'elenco breve.

- **Modello lineare.** Specifica una distribuzione normale con un collegamento identità, utile quando l'obiettivo può essere previsto con un modello di regressione lineare o ANOVA.
- **Regressione gamma.** Specifica una distribuzione gamma con collegamento log da utilizzare quando l'obiettivo contiene tutti valori positivi e propende verso valori maggiori.
- **Loglineare.** Specifica una distribuzione di Poisson con collegamento log da utilizzare quando l'obiettivo rappresenta un conteggio di occorrenze in un determinato periodo di tempo.
- **Regressione binomiale negativa.** Specifica una distribuzione binomiale negativa con collegamento log da utilizzare quando l'obiettivo e il denominatore rappresentano il numero di prove necessarie per osservare i successi  $k$ .
- **Regressione logistica multinomiale.** Specifica una distribuzione multinomiale con collegamento logit generalizzato da utilizzare quando l'obiettivo è una risposta multi-categoria.
- **Regressione logistica binaria.** Specifica una distribuzione binomiale con collegamento logit da utilizzare quando l'obiettivo è una risposta binaria prevista da un modello di regressione logistica.
- **Probit binaria.** Specifica una distribuzione binomiale con collegamento probit da utilizzare quando l'obiettivo è una risposta binaria con una distribuzione normale sottostante.
- **Sopravvivenza di censura per intervallo.** Specifica una distribuzione binomiale con collegamento log-log complementare, utile nell'analisi di sopravvivenza quando alcune osservazioni non includono un evento di terminazione.

### **Distribuzione**

Questa selezione specifica la distribuzione dell'obiettivo. La possibilità di specificare una distribuzione non normale e una funzione di collegamento senza identità è uno dei principali vantaggi offerti dal modello misto lineare generalizzato rispetto a quello misto lineare. Dal momento che è possibile combinare più distribuzioni e funzioni di collegamento e che molte di queste sono adatte a qualsiasi combinazione di dati, è generalmente consigliabile fare una valutazione teorica a priori oppure selezionare la combinazione che si ritiene possa essere più adatta.

- **Binomiale.** Questa distribuzione è indicata solo per un obiettivo che rappresenta una risposta binaria o un numero di eventi.
- **Gamma.** Questa distribuzione è indicata per un obiettivo con valori di scala positivi che propendono verso valori positivi maggiori. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Gaussiana inversa.** Questa distribuzione è indicata per un obiettivo con valori di scala positivi che propendono verso valori positivi maggiori. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Multinomiale.** Questa distribuzione è indicata per un obiettivo che rappresenta una risposta multi-categoria.
- **Binomiale negativo.** La regressione binomiale negativa utilizza una distribuzione binomiale con collegamento log da utilizzare quando l'obiettivo rappresenta un conteggio di occorrenze con varianza elevata.
- **Normale.** Questa distribuzione è indicata per un obiettivo continuo i cui valori presentano una distribuzione simmetrica a forma di campana intorno al valore centrale (medio).
- **Poisson.** Questa distribuzione può essere considerata equivalente al numero di occorrenze di un evento desiderato in un intervallo di tempo fisso ed è indicata per le variabili con valori interi non negativi. Se il valore dei dati è un numero non intero, inferiore a 0 o mancante, il caso corrispondente non viene usato nell'analisi.

### **Funzioni di collegamento**

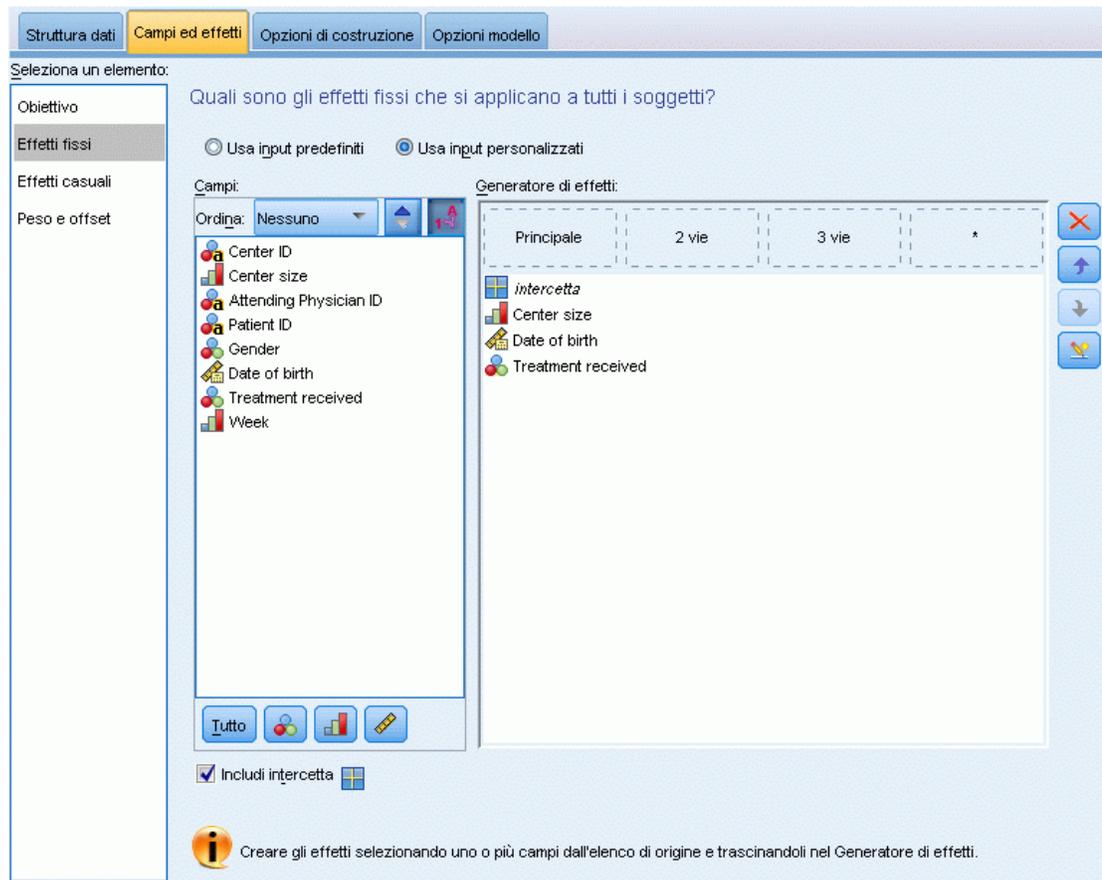
La funzione di collegamento è la trasformazione dell'obiettivo che permette di stimare il modello. Sono disponibili le seguenti funzioni:

- **Identità.**  $f(x)=x$ . L'obiettivo non viene trasformato. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni, tranne che per quella multinomiale.
- **Log-log complementare.**  $f(x)=\log(-\log(1-x))$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Log.**  $f(x)=\log(x)$ . Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni, tranne che per quella multinomiale.
- **Log complementare.**  $f(x)=\log(1-x)$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Logit.**  $f(x)=\log(x / (1-x))$ . Questa funzione è adatta solo con le distribuzioni binomiale o multinomiale ed è l'unica funzione di collegamento valida per la distribuzione multinomiale.
- **Log-log negativa.**  $f(x)=-\log(-\log(x))$ . Questa funzione è indicata solo per la distribuzione binomiale.

- **Probit.**  $f(x)=\Phi^{-1}(x)$ , dove  $\Phi^{-1}$  è la funzione di distribuzione cumulata normale standard inversa. Questa funzione è indicata solo per la distribuzione binomiale.
- **Potenza.**  $f(x)=x^\alpha$ , se  $\alpha \neq 0$ .  $f(x)=\log(x)$ , se  $\alpha=0$ .  $\alpha$  è la specifica del numero richiesto e deve essere un numero reale. Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni, tranne che per quella multinomiale.

## Effetti fissi

Figura 8-4  
Impostazioni degli effetti fissi



I fattori a effetti fissi vengono generalmente considerati come campi i cui valori di interesse sono integralmente rappresentati nell'insieme di dati e possono essere utilizzati per il calcolo del punteggio. Per impostazione predefinita, i campi con il ruolo di input predefinito che non sono specificati in altri punti della finestra di dialogo vengono immessi nella sezione degli effetti fissi del modello. I campi categoriali (nominali e ordinali) vengono utilizzati come fattori nel modello e i campi continui vengono utilizzati come covariate.

Inserire gli effetti nel modello selezionando uno o più campi nell'elenco origine e trascinandoli nell'elenco degli effetti. Il tipo di effetto creato dipende dall'area sensibile nella quale si rilascia la selezione.

- **Principale.** I campi rilasciati vengono visualizzati come effetti principali separati in fondo all'elenco degli effetti.
- **2 vie.** Tutte le possibili coppie dei campi rilasciati vengono visualizzate come interazioni a 2 vie in fondo all'elenco degli effetti.
- **3 vie.** Tutti i possibili gruppi di tre dei campi rilasciati vengono visualizzati come interazioni a 3 vie in fondo all'elenco degli effetti.
- **\***. La combinazione di tutti i campi rilasciati viene visualizzata come un'interazione singola in fondo all'elenco degli effetti.

I pulsanti a destra del generatore di effetti consentono di:



Eliminare termini dal modello a effetti fissi selezionando i termini da eliminare e facendo clic sul pulsante di eliminazione.



Riordinare i termini nel modello a effetti fissi selezionando i termini da riordinare e facendo clic sul pulsante freccia su o giù.

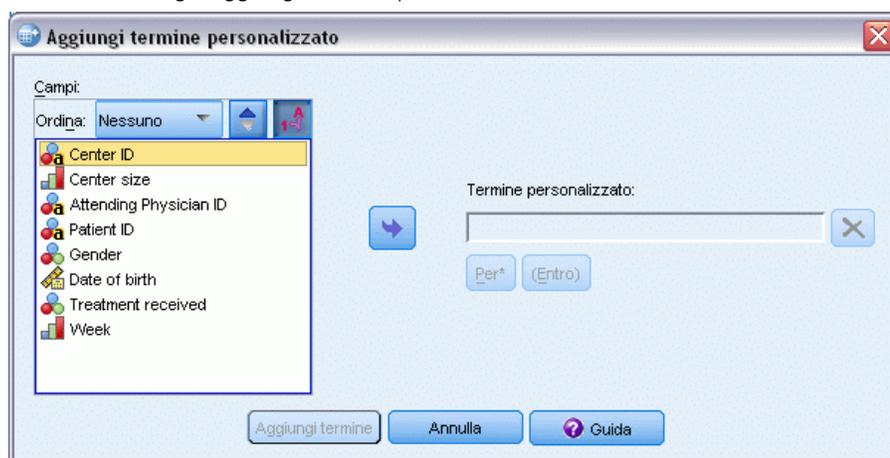


Aggiungere termini nidificati al modello utilizzando la finestra di dialogo [Aggiungi termine personalizzato](#), facendo clic sul pulsante **Aggiungi termine personalizzato**.

**Includi intercetta.** L'intercetta viene in genere inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

## Aggiungi termine personalizzato

Figura 8-5  
Finestra di dialogo *Aggiungi termine personalizzato*



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  $A(X)$ .

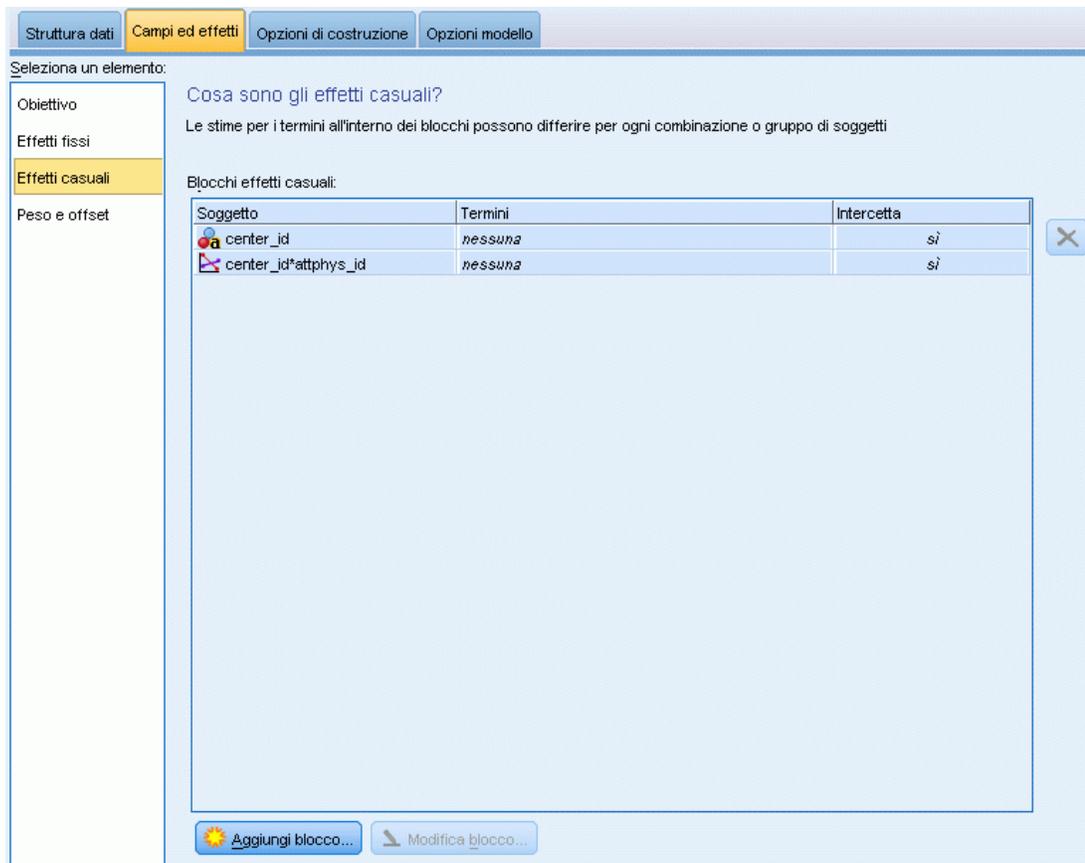
#### ***Costruzione di un termine nidificato***

- ▶ Selezionare un fattore o una covariata nidificati all'interno di un altro fattore, quindi fare clic sul pulsante freccia.
- ▶ Fare clic su (Entro).
- ▶ Selezionare il fattore entro il quale sono nidificati il fattore o la covariata precedenti, quindi fare clic sul pulsante freccia.
- ▶ Fare clic su Aggiungi termine.

È anche possibile includere effetti di interazione o aggiungere più livelli di nidificazione al termine nidificato.

## Effetti casuali

Figura 8-6  
Impostazioni degli effetti casuali



I fattori a effetti casuali sono campi i cui valori, contenuti nel file dei dati, possono essere considerati un campione casuale di una popolazione di valori più ampia. Questi fattori sono utili per spiegare l'eccessiva variabilità dell'obiettivo. Per impostazione predefinita, se è stato selezionato più di un soggetto nella scheda Struttura dati viene creato un Blocco effetti casuali per ogni soggetto oltre quello interno. Ad esempio, se si è selezionato Scuola, Classe e Studente come soggetti nella scheda Struttura dati, vengono creati automaticamente i seguenti blocchi effetti casuali:

- Effetto casuale 1: il soggetto è scuola (senza effetti, solo intercetta)
- Effetto casuale 2: il soggetto è scuola \* classe (senza effetti, solo intercetta)

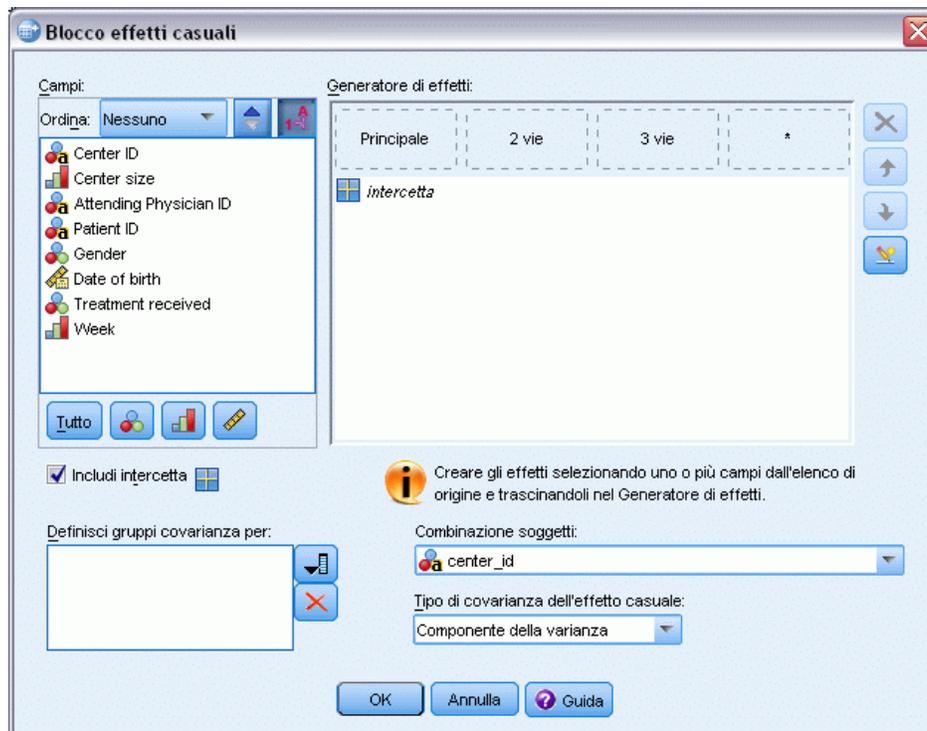
È possibile utilizzare i blocchi effetti casuali nei seguenti modi:

- ▶ Per aggiungere un nuovo blocco, fare clic su **Aggiungi blocco...** Si apre la finestra di dialogo **Blocco effetti casuali**.
- ▶ Per modificare un blocco esistente, selezionare il blocco desiderato e fare clic su **Modifica blocco...** Si apre la finestra di dialogo **Blocco effetti casuali**.

- Per eliminare uno o più blocchi, selezionare i blocchi desiderati e fare clic sul pulsante di eliminazione.

## Blocco effetti casuali

Figura 8-7  
Finestra di dialogo Blocco effetti casuali



Inserire gli effetti nel modello selezionando uno o più campi nell'elenco origine e trascinandoli nell'elenco degli effetti. Il tipo di effetto creato dipende dall'area sensibile nella quale si rilascia la selezione. I campi categoriali (nominali e ordinali) vengono utilizzati come fattori nel modello e i campi continui vengono utilizzati come covariate.

- **Principale.** I campi rilasciati vengono visualizzati come effetti principali separati in fondo all'elenco degli effetti.
- **2 vie.** Tutte le possibili coppie dei campi rilasciati vengono visualizzate come interazioni a 2 vie in fondo all'elenco degli effetti.
- **3 vie.** Tutti i possibili gruppi di tre dei campi rilasciati vengono visualizzati come interazioni a 3 vie in fondo all'elenco degli effetti.
- **\***. La combinazione di tutti i campi rilasciati viene visualizzata come un'interazione singola in fondo all'elenco degli effetti.

I pulsanti a destra del generatore di effetti consentono di:



Eliminare termini dal modello a effetti fissi selezionando i termini da eliminare e facendo clic sul pulsante di eliminazione,



Riordinare termini nel modello a effetti fissi selezionando i termini da riordinare e facendo clic sul pulsante freccia su o giù,



Aggiungere termini nidificati al modello utilizzando la finestra di dialogo [Aggiungi termine personalizzato](#), facendo clic sul pulsante Aggiungi termine personalizzato.

**Includi intercetta.** L'intercetta non viene inclusa nel modello effetti casuali per impostazione predefinita. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

**Definisci gruppi covarianza per.** I campi specificati in questo punto definiscono insiemi indipendenti di parametri di covarianza a effetti casuali, uno per ogni categoria definita dalla classificazione incrociata dei campi di raggruppamento. Per ogni blocco di effetti casuali è possibile specificare un insieme diverso di campi di raggruppamento. Tutti i soggetti presentano lo stesso tipo di covarianza; i soggetti all'interno dello stesso gruppo di covarianza presentano gli stessi valori per i parametri.

**Combinazione soggetti.** Consente di specificare i soggetti a effetti casuali da combinazioni predefinite di soggetti nella scheda Struttura dati. Ad esempio, se nella scheda Struttura dati sono definiti i soggetti *Scuola*, *Classe* e *Studente*, in questo ordine, l'elenco a discesa Combinazione soggetti presenterà le opzioni Nessuno, Scuola, Classe \* scuola e Studente \* classe \* scuola.

**Tipo di covarianza dell'effetto casuale.** Questo tipo di covarianza specifica la struttura della covarianza per i residui. Le strutture disponibili sono:

- Autoregressivo di primo ordine (AR1)
- Autoregressivo a media mobile (1,1) (ARMA11)
- Simmetria composta
- Diagonale
- Identità scalata
- Toeplitz
- Non strutturato
- Componenti della varianza

Per ulteriori informazioni, vedere l'argomento [Strutture della covarianza](#) in l'appendice B a pag. 161.

## Peso e offset

Figura 8-8  
Impostazioni di peso e offset

**Peso analisi.** Il parametro scala è un parametro del modello stimato correlato alla varianza della risposta. I pesi dell’analisi sono valori “noti” che possono variare a seconda delle osservazioni. Se il campo del peso dell’analisi è stato specificato, il parametro scala, che è correlato alla varianza della risposta, viene diviso per i valori del peso dell’analisi per ciascuna osservazione. Per l’analisi non vengono usati i record con valori di pesi analisi inferiori o uguali a 0 o mancanti.

**Offset.** Questo termine identifica un predittore “strutturale”. Il suo coefficiente non viene stimato dal modello ma ritenuto sempre equivalente a 1. Quindi, i valori dell’offset vengono semplicemente aggiunti al predittore lineare dell’obiettivo. Ciò è particolarmente utile nei modelli di regressione di Poisson, in cui ciascun caso può avere vari livelli di esposizione all’evento desiderato. Ad esempio, quando si creano modelli per calcolare la percentuale di incidenti per i singoli conducenti, esiste una sostanziale differenza tra il conducente che ha avuto un incidente in tre anni di guida e uno che ne ha avuto uno in 25 anni. È possibile creare un modello del numero di incidenti come risposta di Poisson, se l’esperienza del conducente è inclusa come termine di offset.

## Opzioni di costruzione

Figura 8-9  
Impostazioni delle opzioni di costruzione

Queste selezioni specificano alcuni criteri avanzati utilizzati per la creazione del modello.

**Ordinamento.** Questi controlli determinano l'ordinamento delle categorie di obiettivo e fattori (input categoriali) ai fini della determinazione dell'ultima categoria. L'impostazione dell'ordinamento viene ignorata se l'obiettivo non è categoriale oppure se nelle impostazioni **Obiettivo** è specificata una categoria di riferimento personalizzata.

**Regole di interruzione.** È possibile specificare il numero massimo di iterazioni che verranno eseguite nell'algoritmo. Specifica un intero non negativo. Il valore predefinito è 100.

**Impostazioni post-stima.** Queste impostazioni specificano come alcuni output del modello vengono calcolati per la visualizzazione.

- **Livello di confidenza.** Si tratta del livello di confidenza utilizzato per calcolare stime di intervallo per i coefficienti del modello. Specificare un valore maggiore di 0 e minore di 100. L'impostazione di default è 95.
- **Gradi di libertà.** Specifica come vengono calcolati i gradi di libertà per i test della significatività. Scegliere Fissi per tutti i test (metodo dei residui) se le dimensioni del campione sono sufficientemente ampie, oppure se i dati sono bilanciati, o ancora se il modello utilizza

un tipo di covarianza più semplice, ad esempio l'identità scalata o diagonale. È l'impostazione di default. Scegliere Diversi nei singoli test (approssimazione di Satterthwaite) se le dimensioni del campione sono ridotte, oppure se i dati non sono bilanciati, o ancora se il modello utilizza un tipo di covarianza complesso, ad esempio la covarianza non strutturata.

- **Test degli effetti fissi e dei coefficienti.** Questo metodo consente di calcolare la matrice di covarianza delle stime dei parametri. Scegliere la stima robusta se si teme che le ipotesi del modello vengano violate.

## Medie stimate

Figura 8-10  
Impostazioni delle medie stimate

Struttura dati | Campi ed effetti | Opzioni di costruzione | **Opzioni modello**

Seleziona un elemento:

Medie stimate  
Salva campi

Esegui una stima dell'obiettivo?  
Specificare medie stimate e contrasti personalizzati.

Termini:

Termini su base categoriale:	Stima medie	Tipo contrasto	Campo contrasto
center_size	<input checked="" type="checkbox"/>	Nessuna	center_size
Gender	<input type="checkbox"/>	Nessuna	Gender
treatment	<input type="checkbox"/>	Nessuna	treatment
Week	<input type="checkbox"/>	Nessuna	Week

I campi continui verranno mantenuti costanti durante la stima dell'obiettivo.

Campi:

Campi continui	Costante	Valore
dob	Media	

Visualizza medie stimate in termini di:  
 Scala obiettivo originale  
 Trasformazione funzione collegamento

Adatta per confronti multipli utilizzando:  
 Differenza meno significativa

Questa scheda consente di visualizzare le medie marginali stimate per i livelli e le interazioni tra fattori. Le medie marginali stimate non sono disponibili per i modelli multinomiali.

**Termini.** In questo punto vengono elencati i termini del modelli negli Effetti fissi interamente composti da campi categoriali. Selezionare ogni termine per il quale si desidera che il modello generi medie marginali stimate.

- **Tipo contrasto.** Specifica il tipo di contrasto da utilizzare per i livelli del campo contrasto. Se è selezionato Nessuno, non viene generato alcun contrasto. A coppie genera confronti a coppie per le combinazioni di tutti i livelli dei fattori specificati. Questa è l'unico contrasto disponibile per le interazioni tra fattori. I contrasti di tipo Deviazione confrontano ciascun livello del fattore con la media principale. I contrasti di tipo Semplice confrontano ciascun livello del fattore, a eccezione dell'ultimo, con l'ultimo livello. L'ultimo livello viene determinato dall'ordinamento dei fattori specificato nelle Opzioni di costruzione. Tenere presente che tutti questi tipi di contrasto non sono ortogonali.
- **Campo contrasto.** Specifica un fattore i cui livelli vengono confrontati utilizzando il tipo di contrasto selezionato. Se è selezionato Nessuno come tipo di contrasto, non è possibile (o necessario) selezionare alcun campo contrasto.

**Campi continui.** I campi continui elencati vengono estratti dai termini negli Effetti fissi che utilizzano campi continui. Durante il calcolo delle medie marginali stimate, le covariate sono fisse ai valori specificati. Selezionare la media oppure specificare un valore personalizzato.

**Visualizza medie stimate in termini di.** Specifica se calcolare le medie marginali stimate in base alla scala originale dell'obiettivo o in base alla trasformazione della funzione di collegamento. Scala obiettivo originale calcola le medie marginali stimate per l'obiettivo. Tenere presente che quando l'obiettivo viene specificato utilizzando l'opzione eventi/prove, fornisce le medie marginali stimate per la proporzione di eventi/prove anziché per il numero di eventi. Trasformazione funzione di collegamento calcola le medie marginali stimate per il predittore lineare.

**Adatta per confronti multipli utilizzando.** 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 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.
- **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 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.

Il metodo della differenza meno significativa è meno conservativo rispetto al metodo Sidak sequenziale, che a sua volta è meno conservativo rispetto al metodo Bonferroni sequenziale; cioè, la differenza meno significativa rifiuta almeno tante ipotesi individuali quante il metodo Sidak sequenziale, che a sua volta rifiuta almeno tante ipotesi individuali quante il metodo Bonferroni sequenziale.

## Salva

Figura 8-11  
Impostazioni di salvataggio

Struttura dati | Campi ed effetti | Opzioni di costruzione | **Opzioni modello**

Selezione un elemento:

Medie stimate  
Salva campi

Salva nell'insieme di dati

Valori attesi  
Nome campo: PredictedValue

Probabilità prevista per obiettivi categoriali  
Nome radice: PredictedProbability  
Numero massimo di categorie da salvare: 25

Intervalli di confidenza  
Nome radice: CI

Residuo di Pearson  
Nome campo: PearsonResidual

Confidenze  
Nome campo: Confidence

Basata su:  
 Probabilità del valore atteso  
 Aumento della probabilità a partire dal successivo valore più probabile

Esporta modello  
Nome file:

Il modello esportato è un insieme di uno o più file .xml.

Gli elementi selezionati vengono salvati con il nome specificato; non sono possibili conflitti con nomi di campo esistenti.

**Valori attesi.** Consente di salvare i valori attesi dell'obiettivo. Il nome predefinito del campo è *PredictedValue*.

**Probabilità prevista per obiettivi categoriali.** Se l'obiettivo è categoriale, questa parola chiave salva le probabilità previste delle prime  $n$  categorie, fino al valore specificato come Numero massimo di categorie da salvare. Il nome radice predefinito è *PredictedProbability*. Per salvare la probabilità prevista della categoria prevista, salvare la confidenza (vedere di seguito).

**Intervalli di confidenza.** Salva i limiti superiore e inferiore dell'intervallo di confidenza per il valore atteso o la probabilità prevista. Per tutte le distribuzioni tranne che per quella multinomiale, questa opzione crea due variabili e il nome radice predefinito è *CI*, con *\_Inferiore* e *\_Superiore* come suffissi.

Per la distribuzione multinomiale, viene creato un campo per ogni categoria di variabili dipendenti a eccezione dell'ultima ([Per ulteriori informazioni, vedere l'argomento Opzioni di costruzione a pag. 111.](#)). Vengono pertanto salvati il limite superiore e inferiore della probabilità prevista cumulativa per le prime  $n$  categorie, fino all'ultima esclusa, e fino al valore specificato come Numero massimo di categorie da salvare. Il nome radice predefinito è *CI* e i nomi di campo predefiniti sono *CI\_Inferiore\_1*, *CI\_Superiore\_1*, *CI\_Inferiore\_2*, *CI\_Superiore\_2*, e così via, in base all'ordine delle categorie obiettivo.

**Residuo di Pearson.** Consente di salvare il residuo di Pearson per ogni record, per poterlo utilizzare nella diagnosi di post-stima dell'adattamento del modello. Il nome predefinito del campo è *PearsonResidual*.

**Confidenze.** Consente di salvare la confidenza nel valore atteso per l'obiettivo categoriale. La confidenza calcolata può essere basata sulla probabilità del valore atteso (la probabilità prevista più alta) o sulla differenza tra la probabilità prevista più alta e la seconda probabilità prevista più alta. Il nome predefinito del campo è *Confidence*.

**Esporta modello.** Consente di scrivere il modello in un file *.zip* esterno. È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Specificare un nome file valido e univoco. Se la definizione del file fa riferimento a un file esistente, il file verrà sovrascritto.

## Vista del modello

Questa procedura crea un oggetto Modello nel Viewer. Attivando l'oggetto con un doppio clic, si accede a una vista interattiva del modello

Per impostazione predefinita, viene mostrata la visualizzazione Riepilogo modello. Per un'altra visualizzazione del modello, selezionarla dalle anteprime delle visualizzazioni.

## Riepilogo modello

Questa visualizzazione è un'istantanea, un riepilogo immediato del modello e del suo adattamento.

**Tabella.** La tabella identifica l'obiettivo, la distribuzione della probabilità e la funzione di collegamento specificati nelle [Impostazioni obiettivo](#). Se l'obiettivo è definito da eventi e prove, la cella è divisa in modo da visualizzare il campo eventi e il campo prove o il numero fisso di prove. Vengono inoltre visualizzati il criterio di informazione di Akaike corretto per il campione finito (AICC) e il criterio di informazione bayesiano (BIC).

- **Akaike corretto.** Una misura per selezionare e confrontare modelli misti basata sulla  $-2 \log$  verosimiglianza (ristretta). I valori più bassi indicano i modelli migliori. AICC corregge AIC in presenza di campioni piccoli. Mano a mano che aumenta la dimensione del campione, il criterio AICC converge nel criterio AIC.
- **Bayesiano.** Una misura per selezionare e confrontare modelli in base alla  $-2 \log$ -verosimiglianza. I valori più bassi indicano i modelli migliori. Anche BIC penalizza i modelli sovrapparametrizzati, ma in modo più rigoroso rispetto ad AIC.

**Grafico.** Se l'obiettivo è categoriale, un grafico visualizza la precisione del modello finale, che è la percentuale di classificazioni corrette.

### **Struttura dati**

Questa visualizzazione fornisce un riepilogo della struttura dati specificata e consente di verificare che i soggetti e le misure ripetute siano stati specificati correttamente. Per ogni campo soggetto e per ogni campo misure ripetute vengono visualizzate le informazioni osservate per il primo soggetto e l'obiettivo. Viene inoltre visualizzato il numero di livelli per ogni campo soggetto e per ogni campo misure ripetute.

### **Previsioni e osservazioni**

Per gli obiettivi continui, inclusi gli obiettivi specificati come eventi/prove, visualizza un grafico a dispersione in bin dei valori attesi sull'asse verticale in base ai valori osservati sull'asse orizzontale. Idealmente, i punti devono trovarsi su una linea a 45 gradi; questa visualizzazione segnala se vi sono record che presentano particolari problemi di previsione da parte del modello.

### **Classificazione**

Per gli obiettivi categoriali, visualizza la classificazione incrociata dei valori osservati dell'obiettivo rispetto a quelli attesi, in una mappa termica, oltre alla percentuale complessiva di correttezza.

**Stili tabella.** Esistono diversi stili di visualizzazione, disponibili nell'elenco a discesa Stile.

- **Percentuale riga.** Visualizza le percentuali delle righe (il conteggio delle celle espresso in percentuale sul totale delle righe) nelle celle. È l'impostazione di default.
- **Conteggi di cella.** Visualizza i conteggi di cella nelle celle. L'ombreggiatura della mappa termica è basata sulle percentuali di righe.
- **Mappa termica.** Non visualizza valori nelle celle ma solo l'ombreggiatura.
- **Compresso.** Non visualizza intestazioni di righe e di colonne né valori nelle celle. Può essere utile quando l'obiettivo ha molte categorie.

**Mancante/i.** Se un record presenta valori mancanti sull'obiettivo, viene visualizzato alla riga (Mancante/i) sotto a tutte le righe valide. I record con valori mancanti non contribuiscono alla percentuale complessiva di correttezza.

**Obiettivi multipli.** Se sono presenti più obiettivi categoriali, ogni obiettivo viene visualizzato in una tabella separata ed è disponibile un elenco a discesa Obiettivo che controlla gli obiettivi da visualizzare.

**Tablette di grandi dimensioni.** Se l'obiettivo visualizzato ha oltre 100 categorie, non viene visualizzata alcuna tabella.

## Effetti fissi

Questa visualizzazione mostra le dimensioni di ogni effetto fisso nel modello.

**Stili.** Esistono diversi stili di visualizzazione, disponibili nell'elenco a discesa Stile.

- **Diagramma.** Si tratta di un grafico in cui gli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi. Le linee di collegamento del diagramma vengono pesate in base alla significatività degli effetti, con la maggiore ampiezza della linea corrispondente agli effetti più significativi (minori valori  $p$ ). È l'impostazione di default.
- **Tabella.** Tabella ANOVA per gli effetti generali e specifici del modello. I singoli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi.

**Significatività.** Il dispositivo di scorrimento Significatività consente di determinare quali effetti mostrare nella visualizzazione. Gli effetti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. L'azione del dispositivo di scorrimento non modifica il modello, consente semplicemente di evidenziare gli effetti più importanti. Il valore predefinito è 1,00, ovvero gli effetti non vengono filtrati in base alla significatività.

## Coefficienti fissi

Questa visualizzazione mostra il valore di ogni coefficiente fisso nel modello. Si noti che i fattori (predittori categoriali) sono codificati mediante un indicatore nel modello, in modo tale che agli **effetti** contenenti fattori possano essere associati più **coefficienti**, uno per ogni categoria esclusa la categoria corrispondente al coefficiente ridondante.

**Stili.** Esistono diversi stili di visualizzazione, disponibili nell'elenco a discesa Stile.

- **Diagramma.** Si tratta di un grafico che visualizza prima l'intercetta e successivamente ordina gli effetti dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi. Negli effetti che contengono fattori i coefficienti vengono ordinati in ordine crescente in base ai valori dei dati. Le linee di collegamento del diagramma sono colorate e pesate in base alla significatività dei coefficienti, con la maggiore ampiezza della linea corrispondente ai coefficienti più significativi (minori valori  $p$ ). Questo è lo stile di default.
- **Tabella.** Indica i valori, i test di significatività e gli intervalli di confidenza per i singoli coefficienti del modello. Dopo l'intercetta, gli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi. Negli effetti che contengono fattori i coefficienti vengono ordinati in ordine crescente in base ai valori dei dati.

**Multinomiale.** Se è attiva la distribuzione multinomiale, l'elenco a discesa Multinomiale determina quale categoria obiettivo visualizzare. Il criterio di ordinamento dei valori nell'elenco è determinato dalle opzioni specificate nelle impostazioni Opzioni di costruzione.

**Esponenziale.** Visualizza le stime dei coefficienti esponenziali e gli intervalli di confidenza per alcuni tipi di modello, inclusi la regressione logistica binaria (distribuzione binomiale e collegamento logit), la regressione logistica nominale (distribuzione multinomiale e collegamento logit), la regressione binomiale negativa (distribuzione binomiale negativa e collegamento log) e il modello log-lineare (distribuzione di Poisson e collegamento log).

**Significatività.** Il dispositivo di scorrimento Significatività consente di determinare quali coefficienti mostrare nella visualizzazione. I coefficienti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. L'azione del dispositivo di scorrimento non modifica il modello, consente semplicemente di evidenziare i coefficienti più importanti. Il valore di default è 1,00, ovvero i coefficienti non vengono filtrati in base alla significatività.

## ***Covarianze di effetti casuali***

Questa visualizzazione mostra la matrice di covarianza degli effetti casuali (**G**).

**Stili.** Esistono diversi stili di visualizzazione, disponibili nell'elenco a discesa Stile.

- **Valori di covarianza.** Si tratta di una mappa termica della matrice di covarianza in cui gli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi. I colori nel diagramma di correlazione corrispondono ai valori delle celle mostrati nella chiave. È l'impostazione di default.
- **Diagramma di correlazione.** Si tratta di una mappa termica della matrice di covarianza.
- **Compresso.** Si tratta di una mappa termica della matrice di covarianza senza le intestazioni delle righe e delle colonne.

**Blocchi.** Se sono presenti più blocchi effetti casuali, è disponibile un elenco a discesa Blocco che consente di selezionare il blocco da visualizzare.

**Gruppi.** Se un blocco effetti casuali ha una specifica di gruppo, è disponibile un elenco a discesa Gruppo che consente di selezionare il livello del gruppo da visualizzare.

**Multinomiale.** Se è attiva la distribuzione multinomiale, l'elenco a discesa Multinomiale determina quale categoria obiettivo visualizzare. Il criterio di ordinamento dei valori nell'elenco è determinato dalle opzioni specificate nelle impostazioni Opzioni di costruzione.

## ***Parametri di covarianza***

Questa visualizzazione mostra le stime dei parametri di covarianza e le statistiche associate agli effetti casuali e residui. Si tratta di risultati avanzati ma fondamentali che forniscono informazioni sull'idoneità della struttura di covarianza.

**Tabella riassuntiva.** Si tratta di un riferimento rapido per il numero di parametri nelle matrici di covarianza degli effetti residui (**R**) e casuali (**G**), la classifica (numero di colonne) nelle matrici di disegno degli effetti fissi (**X**) e degli effetti casuali (**Z**) e il numero di soggetti stabiliti dai campi soggetto che definiscono la struttura dati.

**Tabella Parametri di covarianza.** Visualizza la stima, l'errore standard e l'intervallo di confidenza per ogni parametro di covarianza dell'effetto selezionato. Il numero di parametri visualizzato dipende dalla struttura di covarianza dell'effetto e, per i blocchi di effetti casuali, dal numero di effetti nel blocco. Se si ritiene che i parametri esterni alla diagonale non siano significativi, è possibile utilizzare una struttura di covarianza più semplice.

**Effetti.** Se sono presenti più blocchi effetti casuali, è disponibile un elenco a discesa Effetto che consente di selezionare il blocco di effetti residui o casuali da visualizzare. L'effetto residuo è sempre disponibile.

**Gruppi.** Se un blocco di effetti residui o casuali ha una specifica di gruppo, è disponibile un elenco a discesa Gruppo che consente di selezionare il livello del gruppo da visualizzare.

**Multinomiale.** Se è attiva la distribuzione multinomiale, l'elenco a discesa Multinomiale determina quale categoria obiettivo visualizzare. Il criterio di ordinamento dei valori nell'elenco è determinato dalle opzioni specificate nelle impostazioni Opzioni di costruzione.

### **Medie stimate: effetti significativi**

Si tratta dei grafici visualizzati per i 10 fattori a effetti completamente fissi “più significativi”, iniziando dalle interazioni a tre vie, seguite dalle interazioni a due vie e quindi dagli effetti principali. Il grafico visualizza il valore del modello stimato dell'obiettivo sull'asse verticale per ogni valore dell'effetto principale (o dell'effetto elencato per primo in un'interazione) sull'asse orizzontale; viene inoltre generata una linea distinta per ogni valore del secondo effetto elencato in un'interazione a due vie; per ogni valore nel terzo effetto elencato in un'interazione a tre vie viene generato un grafico distinto; tutti gli altri predittori rimangono costanti. È una visualizzazione utile degli effetti dei coefficienti di ciascun predittore sull'obiettivo. Tenere presente che se non vi sono predittori significativi, non vengono generate medie stimate.

**Confidenza.** Visualizza i limiti di confidenza superiore e inferiore per le medie marginali, utilizzando il livello di confidenza specificato nelle Opzioni di costruzione.

### **Medie stimate: effetti personalizzati**

Si tratta di tabelle e grafici per i fattori a effetti completamente fissi richiesti dall'utente.

**Stili.** Esistono diversi stili di visualizzazione, disponibili nell'elenco a discesa Stile.

- **Diagramma.** Questo stile visualizza un grafico lineare del valore del modello stimato dell'obiettivo sull'asse verticale per ogni valore dell'effetto principale (o dell'effetto elencato per primo in un'interazione) sull'asse orizzontale; viene inoltre generata una linea distinta per ogni valore del secondo effetto elencato in un'interazione a due vie; per ogni valore nel terzo effetto elencato in un'interazione a tre vie viene generato un grafico distinto; tutti gli altri predittori rimangono costanti.

Se sono stati richiesti dei contrasti, viene visualizzato un altro grafico che consente di confrontare i livelli del campo contrasto; per le interazioni, viene visualizzato un grafico per ogni combinazione di livello degli effetti diversi dal campo contrasto. Per i contrasti **a coppie**, si tratta di un grafico di rete delle distanze, cioè una rappresentazione grafica della tabella dei confronti in cui le distanze fra i nodi della rete corrispondono a differenze tra i campioni. Le linee gialle corrispondono alle differenze statisticamente significative; le linee nere corrispondono alle differenze non significative. Se si passa il mouse sopra una linea della rete viene visualizzata una descrizione con la significatività corretta della differenza tra i nodi collegati dalla linea.

Per i contrasti di **deviazione**, viene visualizzato un grafico a barre con il valore del modello stimato dell'obiettivo sull'asse verticale e i valori del campo contrasto sull'asse orizzontale; per le interazioni, viene visualizzato un grafico per ogni combinazione di livello degli effetti diversi dal campo contrasto. Le barre mostrano la differenza tra ogni livello del campo contrasto e la media globale, rappresentata da una linea orizzontale nera.

Per i contrasti **semplici**, viene visualizzato un grafico a barre con il valore del modello stimato dell'obiettivo sull'asse verticale e i valori del campo contrasto sull'asse orizzontale; per le interazioni, viene visualizzato un grafico per ogni combinazione di livello degli effetti diversi dal campo contrasto. Le barre mostrano la differenza tra ogni livello del campo contrasto (a eccezione dell'ultimo) e l'ultimo livello, rappresentato da una linea orizzontale nera.

- **Tabella.** Questo stile visualizza una tabella del valore del modello stimato dell'obiettivo, dei suoi errori standard e dell'intervallo di confidenza per ogni combinazione di livello dei campi nell'effetto; tutti gli altri predittori rimangono costanti.

Se sono stati richiesti i contrasti, viene visualizzata un'altra tabella con la stima, l'errore standard, il test di significatività e l'intervallo di confidenza per ogni contrasto; per le interazioni, è presente un insieme separato di righe per ogni combinazione di livello degli effetti diversi dal campo contrasto. Viene inoltre visualizzata una tabella con i risultati dei test globali; per le interazioni, è disponibile un test globale distinto per ogni combinazione di livello degli effetti diversi dal campo contrasto.

**Confidenza.** Attiva o disattiva la visualizzazione dei limiti di confidenza superiore e inferiore per le medie marginali, utilizzando il livello di confidenza specificato nelle Opzioni di costruzione.

**Layout.** Attiva o disattiva la visualizzazione del diagramma dei contrasti a coppie. Il layout circolare mette meno in evidenza i contrasti rispetto al layout a rete ma evita la sovrapposizione delle linee.

## *Selezione Analisi Loglineare*

La procedura Selezione Analisi Loglineare consente di analizzare tavole di contingenza a più vie e di adattare modelli loglineari gerarchici a tavole di contingenza multidimensionali tramite un algoritmo iterativo di adattamento proporzionale. Questa procedura agevola inoltre l'individuazione delle variabili categoriali associate. Per la creazione di modelli, sono disponibili metodi di inserimento forzato ed eliminazione all'indietro. Per i modelli saturi è possibile richiedere stime dei parametri e test di associazione parziale. Con un modello saturo viene aggiunto 0,5 a tutte le celle.

**Esempio.** In uno studio relativo alla preferenza dei consumatori tra due detersivi per lavatrice, i ricercatori hanno contato il numero di soggetti in ciascun gruppo, combinando le varie categorie delle variabili durezza dell'acqua (alta, media o bassa), l'uso precedente di una delle due marche di detersivo e la temperatura del lavaggio (fredda o calda). Dallo studio è stata identificata la relazione tra temperatura e durezza dell'acqua e tra temperatura e marca di detersivo preferita.

**Statistiche.** Frequenze, residui, stime dei parametri, errori standard, intervalli di confidenza, test di associazione parziale. Per modelli personalizzati, grafici di residui e grafici di probabilità normale.

**Dati.** Le variabili fattore sono categoriali. Tutte le variabili da analizzare devono essere numeriche. Prima di avviare l'analisi, è possibile ricodificare le variabili stringa categoriali in forma di variabili numeriche.

Evitare di definire molte variabili con molti livelli. In questo modo si rischia di arrivare a una situazione in cui molte celle hanno un numero limitato di osservazioni e i valori della statistica chi-quadrato potrebbero diventare inutili.

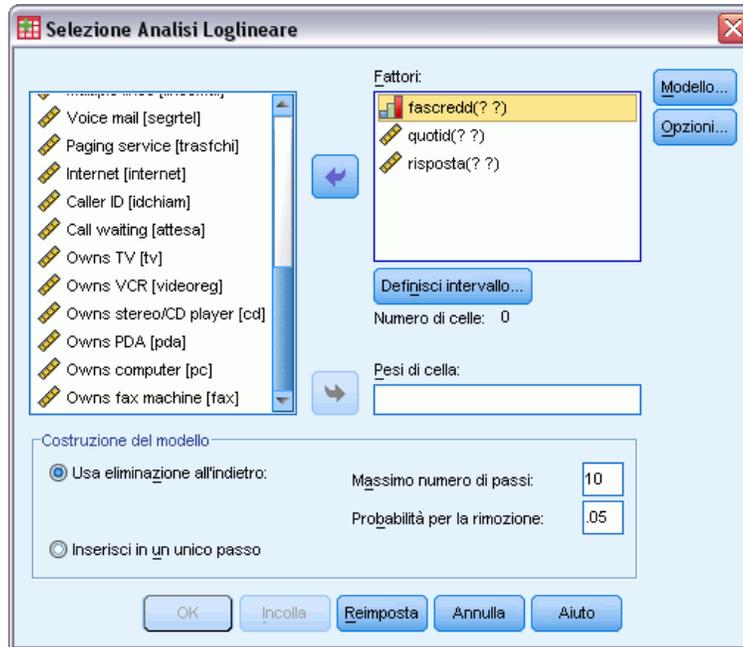
**Procedure correlate.** La procedura Selezione modello può agevolare l'identificazione dei termini richiesti nel modello. È quindi possibile continuare la valutazione del modello tramite la procedura Analisi loglineare generale o Analisi loglineare Logit. Le variabili stringa possono essere ricodificate tramite la ricodifica automatica. Se a una variabile numerica sono associate categorie vuote, è necessario usare il comando Ricodifica in modo da creare valori interi consecutivi.

### *Per ottenere un'analisi loglineare con selezione del modello*

Dai menu, scegliere:

Analizza > Loglineare > Selezione modello...

Figura 9-1  
Finestra di dialogo Selezione Analisi Loglineare

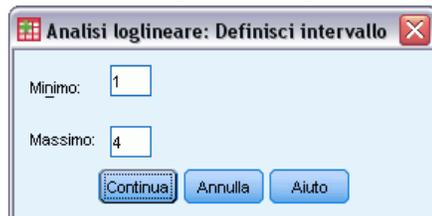


- ▶ Selezionare due o più fattori categoriali numerici.
- ▶ Selezionare una o più variabili fattore dall'elenco Fattori, quindi fare clic su Definisci intervallo.
- ▶ Definire l'intervallo di valori per ciascuna variabile fattore.
- ▶ Selezionare un'opzione del gruppo Costruzione del modello.

Se necessario, è possibile selezionare una variabile di ponderazione di cella per la definizione di zero strutturali.

## **Analisi loglineare: Definisci intervallo**

Figura 9-2  
Finestra di dialogo Analisi loglineare: Definisci intervallo



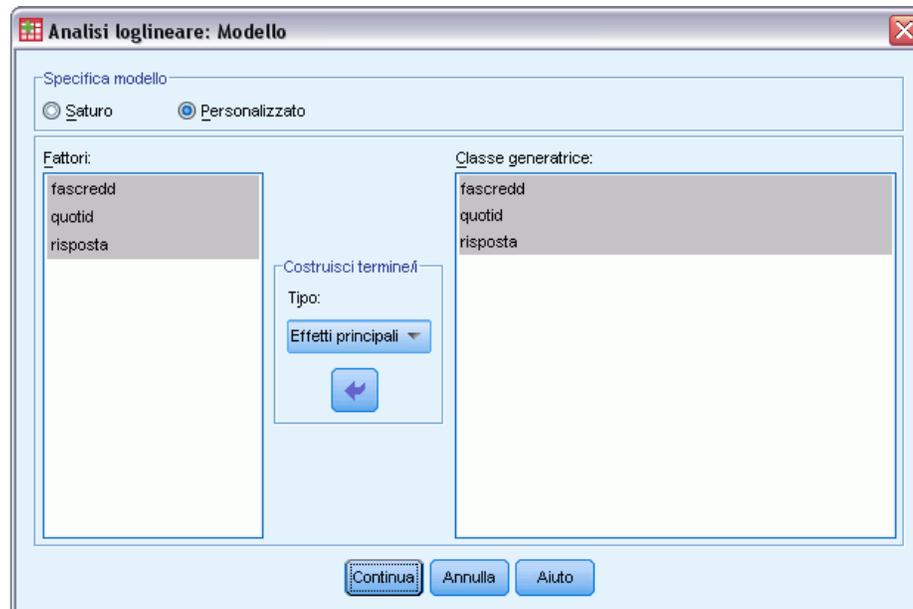
È necessario indicare l'intervallo di categorie per ciascuna variabile fattore. Il valore massimo e il valore minimo corrispondono rispettivamente alla categoria inferiore e a quella superiore della variabile fattore. Entrambi i valori devono essere numeri interi e il valore minimo deve essere minore di quello massimo. Sono esclusi i casi con valori al di fuori dei limiti. Se, ad esempio, si

specifica un valore minimo di 1 e un valore massimo di 3, verranno utilizzati solo i valori 1, 2 e 3. Ripetere l'operazione per ciascuna variabile fattore.

## Analisi loglineare: Modello

Figura 9-3

Finestra di dialogo Analisi loglineare: Modello



**Specifica modello.** Un modello saturo include tutti gli effetti principali dei fattori e tutte le interazioni tra fattori. Selezionare Personalizzato per impostare una classe generatrice per un modello non saturo.

**Classe generatrice.** Una classe generatrice consiste in un elenco dei termini di ordine più elevato in cui sono inclusi i fattori. Un modello gerarchico contiene i termini che definiscono la classe generatrice e tutti gli elementi di ordine inferiore. Si supponga di selezionare le variabili *A*, *B* e *C* dall'elenco Fattori e quindi Interazione dall'elenco Costruisci termini. Il modello risultante includerà l'interazione a 3 vie specificata  $A*B*C$ , le interazioni a 2 vie  $A*B$ ,  $A*C$  e  $B*C$  e gli effetti principali per *A*, *B* e *C*. Non specificare gli elementi di ordine inferiore della classe generatrice.

## Costruisci termini

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

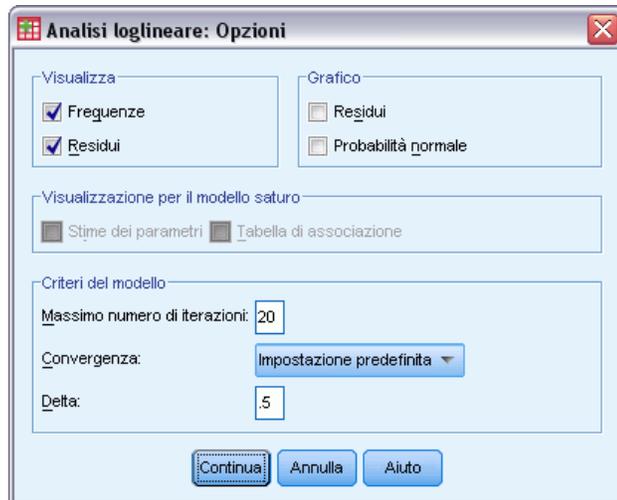
**Tutti 4-vie.** Consente di creare tutte le possibili interazioni a quattro vie delle variabili selezionate.

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

## Selezione Analisi Loglineare: Opzioni

Figura 9-4

Finestra di dialogo Analisi loglineare: Opzioni



**Visualizzazione.** È possibile scegliere Frequenze o Residui oppure entrambe le opzioni. In un modello saturo le frequenze osservate e attese sono uguali e i residui sono uguali a 0.

**Grafico.** Per i modelli personalizzati, è possibile scegliere uno dei tipi di grafico Residui e Probabilità normale oppure entrambi. Tali modelli consentono di stabilire il grado di adeguatezza del modello rispetto ai dati.

**Visualizzazione per il modello saturo.** Per i modelli saturi è possibile scegliere Stime dei parametri. Le stime dei parametri possono agevolare l'identificazione dei termini che è possibile eliminare dal modello. È inoltre disponibile una tabella associata in cui sono elencati i test di associazione parziale. Con tabelle contenenti molti fattori questa opzione richiede operazioni di calcolo complesse.

**Criteri del modello.** Le stime dei parametri vengono ottenute tramite un algoritmo iterativo di adattamento proporzionale. È possibile ignorare uno o più criteri di stima specificando Massimo numero di iterazioni, Convergenza o Delta (un valore aggiunto a tutte le frequenze di cella dei modelli saturi).

## Opzioni aggiuntive del comando HILOGLINEAR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i pesi di cella in forma di matrice (tramite il sottocomando CWEIGHT).
- Generare analisi di diversi modelli con un solo comando (tramite il sottocomando DESIGN).

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

## ***Analisi loglineare generale***

La procedura Analisi loglineare generale consente di analizzare i conteggi di frequenza delle osservazioni che rientrano in ciascuna categoria di classificazione incrociata di una tavola di contingenza. Ciascuna classificazione incrociata inclusa nella tavola corrisponde a una cella, mentre ciascuna variabile categoriale è un fattore. La variabile dipendente corrisponde al numero di casi (frequenza) in una cella della tavola di contingenza, mentre le variabili esplicative corrispondono a fattori e covariate. Questa procedura consente di stimare i parametri di massima verosimiglianza dei modelli loglineari gerarchici e non gerarchici in cui viene usato il metodo di Newton-Raphson. È possibile analizzare una distribuzione di Poisson o multinomiale.

Per definire le celle di una tavola è possibile selezionare fino a 10 fattori. Una variabile di struttura di cella consente di definire zero strutturali per tavole incomplete, prevede un termine di rientro nel modello, si adatta a un modello a tasso logaritmico o consente di implementare il metodo di correzione delle tavole marginali. Le variabili di contrasto consentono di calcolare i rapporti logaritmici generalizzati (GLOR).

Le informazioni e le statistiche sulla bontà dell'adattamento relative al modello vengono visualizzate automaticamente. È inoltre possibile visualizzare una serie di statistiche e grafici diversi, nonché salvare valori residui e attesi nel file dati attivo.

**Esempio.** I dati di un rapporto relativi agli incidenti stradali in Lombardia vengono usati per stabilire la relazione tra l'uso delle cinture di sicurezza e il livello di gravità, fatale o meno, delle lesioni subite. Il rapporto odd indica l'evidenza significativa di una relazione.

**Statistiche.** Frequenze osservate e attese, residui semplici, corretti e di devianza, matrice di disegno, stime dei parametri, rapporto odd, rapporto logaritmico, GLOR, statistica di Wald e intervalli di confidenza. Grafici: residui corretti, residui di devianza e probabilità normale.

**Dati.** I fattori sono categoriali, mentre le covariate di cella sono continue. Quando il modello include una covariata, il valore di covariata medio dei casi di una cella viene applicato automaticamente a tale cella. Le variabili di contrasto sono continue. Consentono di calcolare i rapporti logaritmici generalizzati. I valori della variabile di contrasto corrispondono ai coefficienti della combinazione lineare dei logaritmi dei conteggi di cella attesi.

Una variabile di struttura di cella consente di assegnare pesi. Se, ad esempio, alcune celle sono zeri strutturali, il valore della variabile di struttura di cella sarà 0 o 1. Non usare una variabile di struttura di cella per assegnare pesi a dati aggregati. Per eseguire questa operazione, scegliere invece Pesa casi dal menu Dati.

**Assunzioni.** Nell'analisi loglineare generale sono disponibili i due tipi di distribuzione di Poisson e multinomiale.

L'ipotesi della distribuzione di Poisson prevede quanto segue:

- Prima di iniziare lo studio, le dimensioni totali del campione non sono fisse oppure l'analisi non dipende dalle dimensioni totali del campione.
- La presenza di un'osservazione in una cella è statisticamente indipendente dai conteggi di cella di altre celle.

L'ipotesi della distribuzione multinomiale prevede quanto segue:

- Le dimensioni totali del campione sono fisse oppure l'analisi dipende dalle dimensioni totali del campione.
- I conteggi di cella non sono statisticamente indipendenti.

**Procedure correlate.** Per esaminare le tavole di contingenza, usare la procedura Tavole di contingenza. Usare invece la procedura Analisi loglineare Logit quando è normale considerare una o più variabili categoriali come variabili di risposta e le altre come variabili esplicative.

### **Per ottenere un'analisi loglineare generale**

- ▶ Dai menu, scegliere:  
Analizza > Loglineare > Generale...

Figura 10-1  
Finestra di dialogo Analisi loglineare generale



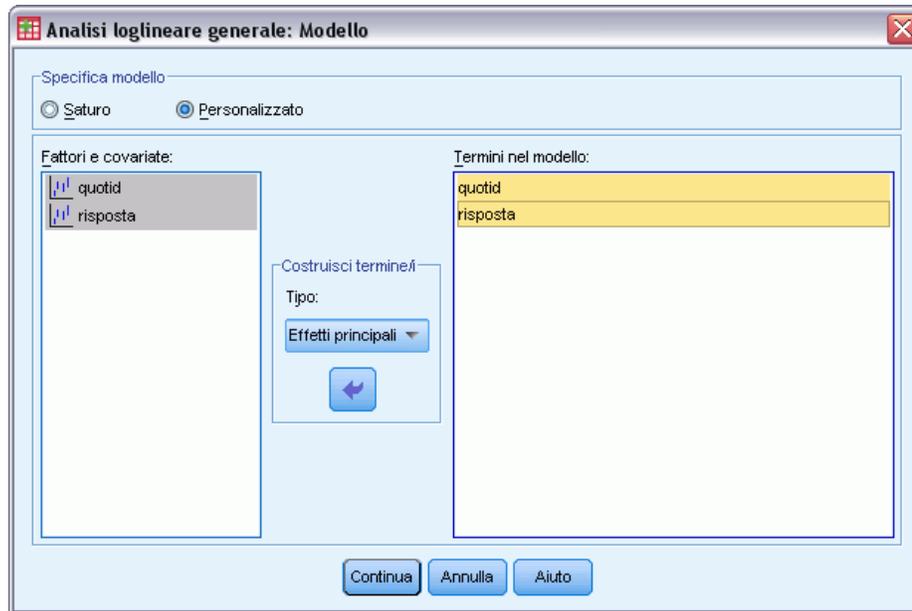
- ▶ Nella finestra di dialogo Analisi loglineare generale selezionare fino a 10 variabili fattore.

Se lo si desidera, è possibile:

- Selezionare covariate di cella.
- Selezionare una variabile di struttura di cella per definire zero strutturali o includere un termine di rientro.
- Selezionare una variabile di contrasto.

## Analisi loglineare generale: Modello

Figura 10-2  
Finestra di dialogo Analisi loglineare generale: Modello



**Specifica modello.** Un modello saturo include tutti gli effetti principali e le interazioni che implicano l'uso di variabili fattore. Non contiene termini di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate.

**Fattori e covariate.** I fattori e le covariate sono elencati.

**Termini nel modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire. È necessario indicare tutti i termini da includere nel modello.

### Costruisci termini

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

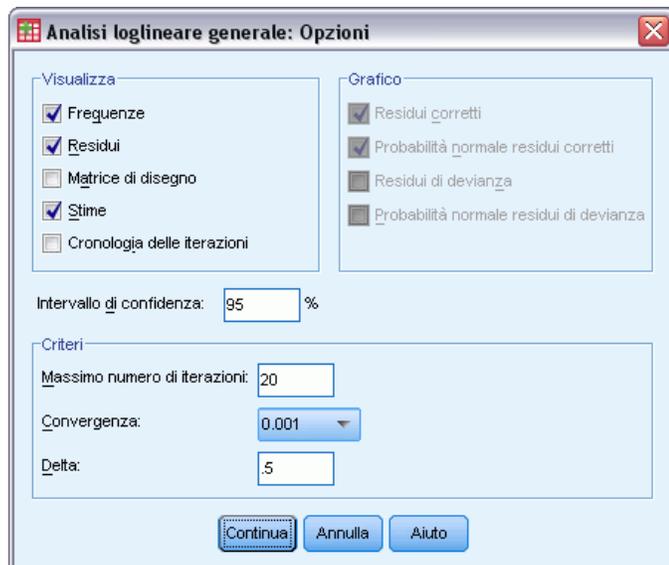
**Tutti 3-vie.** Consente di creare tutte le possibili interazioni a tre vie delle variabili selezionate.

**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.

## Analisi loglineare generale: Opzioni

Figura 10-3  
Finestra di dialogo Analisi loglineare generale: Opzioni



La procedura Analisi loglineare generale consente di visualizzare informazioni e statistiche sulla bontà dell'adattamento relative al modello. È inoltre possibile selezionare una o più delle seguenti opzioni:

**Visualizzazione.** Per la visualizzazione sono disponibili numerose statistiche, ovvero frequenze di cella osservate e attese, residui semplici, corretti e di devianza, una matrice di disegno e stime dei parametri per il modello.

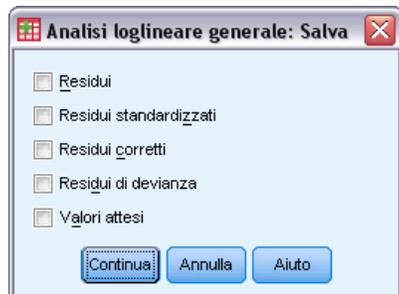
**Grafico.** I grafici sono disponibili solo con i modelli personalizzati e includono due matrici per grafico a dispersione (residui corretti o di devianza confrontati con conteggi di cella osservati e attesi). È inoltre possibile visualizzare la probabilità normale e grafici normali detrendizzati dei residui corretti o della devianza.

**Intervallo di confidenza.** L'intervallo di confidenza delle stime dei parametri può essere corretto.

**Criteri.** Il metodo di Newton-Raphson consente di ottenere le stime dei parametri della massima verosimiglianza. È possibile inserire nuovi valori per il numero massimo di iterazioni, per il criterio di convergenza e per il delta (una costante aggiunta in tutte le celle per approssimazioni iniziali). Nei modelli saturi il delta viene mantenuto nelle celle.

## Analisi loglineare generale: Salva

Figura 10-4  
Finestra di dialogo Analisi loglineare generale: Salva



Selezionare i valori che si desidera salvare come nuove variabili nel file dati attivo. Il suffisso  $n$  nel nome delle nuove variabili viene automaticamente incrementato di un'unità in modo da creare un nome univoco per ciascuna variabile salvata.

I valori salvati fanno riferimento ai dati aggregati, ovvero alle celle della tavola di contingenza, anche se tali dati sono registrati in osservazioni singole nell'Editor dei dati. Un valore residuo o atteso per dati non aggregati che viene salvato per una cella della tavola di contingenza viene inserito nell'Editor dei dati per ogni singolo caso incluso in tale cella. Per poter usare in modo utile i valori salvati, è consigliabile aggregare i dati in modo da ottenere i conteggi di cella.

È possibile salvare quattro tipi di residui: semplici, standardizzati, corretti e di devianza. È inoltre possibile salvare valori attesi.

- **Residui.** Chiamato anche residuo semplice, è la differenza tra il conteggio di cella osservato e il relativo conteggio atteso.
- **Residui standardizzati.** Il residuo diviso per una stima del suo errore standard. I residui standardizzati vengono anche detti residui di Pearson.
- **Residui corretti.** Il residuo standardizzato diviso per il proprio errore standard stimato. I residui corretti seguono asintoticamente una normale standardizzata quando il modello è appropriato. Per questo motivo sono preferibili ai residui standardizzati per verificare l'ipotesi di normalità.
- **Residui di devianza.** Il residuo di devianza è la radice quadrata con segno del contributo individuale alla statistica del Chi-quadrato del rapporto di verosimiglianza ( $G$  quadrato). Il segno è lo stesso del residuo (frequenza osservata meno frequenza attesa). La distribuzione dei residui di devianza è una normale asintotica standard.

## Opzioni aggiuntive del comando GENLOG

Il linguaggio della sintassi dei comandi consente inoltre di:

- Calcolare combinazioni lineari di frequenze di celle osservate e attese e calcolare residui di stampa, residui standardizzati e residui corretti di tali combinazioni (tramite il sottocomando GERESID).
- Modificare il valore soglia predefinito per la verifica della ridondanza (tramite il sottocomando CRITERIA).
- Visualizzare i residui standardizzati (tramite il sottocomando PRINT).

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

## ***Analisi loglineare Logit***

La procedura Analisi loglineare Logit consente di analizzare la relazione tra variabili dipendenti, o di risposta, e variabili indipendenti, o esplicative. Le variabili dipendenti sono sempre categoriali, mentre quelle indipendenti possono essere anche categoriali (fattori). Altre variabili indipendenti, ovvero le covariate di cella, possono essere continue, ma non vengono applicate in modo specifico per ogni singolo caso. La media covariata ponderata per una cella viene applicata a tale cella. Il logaritmo della probabilità a favore delle variabili dipendenti viene espresso come combinazione lineare di parametri. Viene automaticamente assunta la distribuzione multinomiale. Questi modelli sono a volte definiti modelli logit multinomiali. Con questa procedura vengono stimati i parametri di modelli loglineari logit tramite l'algoritmo di Newton-Raphson.

È possibile selezionare da 1 a 10 variabili dipendenti e variabili fattore in combinazione. Una variabile di struttura di cella consente di definire zero strutturali per tavole incomplete, prevede un termine di rientro nel modello, si adatta a un modello a tasso logaritmico o consente di implementare il metodo di correzione delle tavole marginali. Le variabili di contrasto consentono di calcolare i rapporti logaritmici generalizzati (GLOR). I valori della variabile di contrasto corrispondono ai coefficienti della combinazione lineare dei logaritmi dei conteggi di cella attesi.

Le informazioni e le statistiche sulla bontà dell'adattamento relative al modello vengono visualizzate automaticamente. È inoltre possibile visualizzare una serie di statistiche e grafici diversi, nonché salvare valori residui e attesi nel file dati attivo.

**Esempio.** In uno studio condotto in Florida sono stati analizzati 219 alligatori. Come varia il tipo di cibo consumato dagli alligatori in relazione alle dimensioni degli alligatori e al lago in cui vivono tra i quattro laghi presi in considerazione? Dallo studio è risultato che la probabilità a favore di un alligatore di piccole dimensioni che preferisce i rettili ai pesci è 0,70 volte minore rispetto a quella di alligatori più grandi. Inoltre, la probabilità a favore della selezione di rettili anziché di pesci è risultata maggiore nel terzo lago.

**Statistiche.** Frequenze osservate e attese, residui semplici, corretti e di devianza, matrice di disegno, stime dei parametri, rapporto logaritmico generalizzato, statistica di Wald e intervalli di confidenza. Grafici: residui corretti, residui di devianza e probabilità normale.

**Dati.** Le variabili dipendenti sono categoriali. I fattori sono categoriali. Le covariate di cella possono essere continue, ma quando una covariata è inclusa nel modello, il valore della covariata media per i casi inclusi in una cella viene applicato a tale cella. Le variabili di contrasto sono continue. Consentono di calcolare i rapporti logaritmici generalizzati (GLOR). I valori della variabile di contrasto corrispondono ai coefficienti della combinazione lineare dei logaritmi dei conteggi di cella attesi.

Una variabile di struttura di cella consente di assegnare pesi. Se, ad esempio, alcune celle sono zero strutturali, il valore della variabile di struttura di cella sarà 0 o 1. Non usare una variabile di struttura di cella per assegnare pesi a dati aggregati. Per eseguire questa operazione, scegliere invece Pesa casi dal menu Dati.

**Assunzioni.** Si presume che nei conteggi all'interno di ogni combinazione di categorie di variabili esplicative venga usata una distribuzione multinomiale. L'ipotesi della distribuzione multinomiale prevede quanto segue:

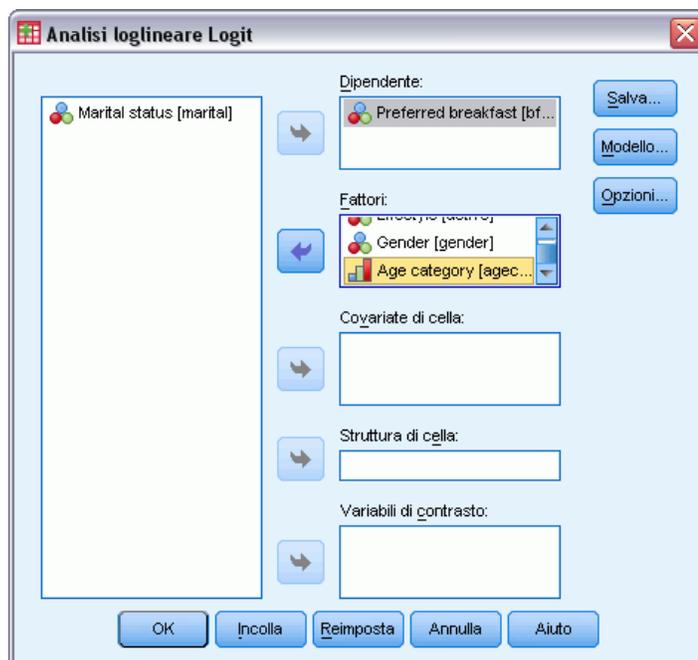
- Le dimensioni totali del campione sono fisse oppure l'analisi dipende dalle dimensioni totali del campione.
- I conteggi di cella non sono statisticamente indipendenti.

**Procedure correlate.** Per visualizzare le tavole di contingenza, usare la procedura Tavole di contingenza. Usare invece la procedura Analisi loglineare generale quando si desidera analizzare la relazione tra un conteggio osservato e un insieme di variabili esplicative.

### Per ottenere un'analisi loglineare Logit

- Dai menu, scegliere:  
Analizza > Loglineare > Logit...

Figura 11-1  
Finestra di dialogo Analisi loglineare Logit



- Nella finestra di dialogo Analisi loglineare Logit selezionare una o più variabili dipendenti.
- Selezionare una o più variabili fattore.

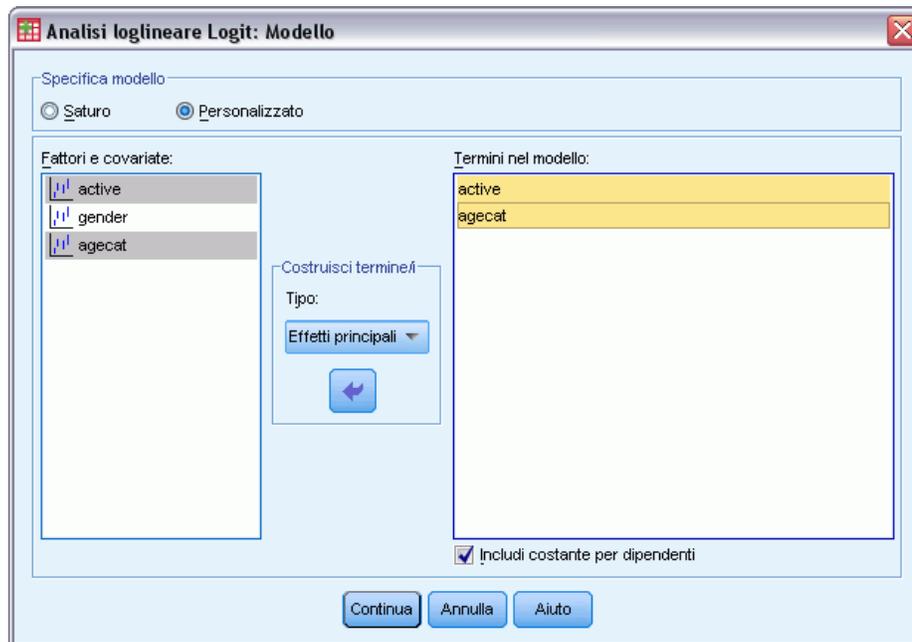
Il numero totale di variabili di fattore e dipendenti non può essere maggiore di 10.

Se lo si desidera, è possibile:

- Selezionare covariate di cella.
- Selezionare una variabile di struttura di cella per definire zero strutturali o includere un termine di rientro.
- Selezionare una o più variabili di contrasto.

## Analisi loglineare Logit: Modello

Figura 11-2  
Finestra di dialogo Analisi loglineare Logit: Modello



**Specifica modello.** Un modello saturo include tutti gli effetti principali e le interazioni che implicano l'uso di variabili fattore. Non contiene termini di covariate. Selezionare Personalizzato per specificare un solo sottoinsieme di interazioni o interazioni tra fattori e covariate.

**Fattori e covariate.** I fattori e le covariate sono elencati.

**Termini nel modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato Personalizzato, è possibile selezionare gli effetti principali e le interazioni desiderate per l'analisi da eseguire. È necessario indicare tutti i termini da includere nel modello.

I termini aggiunti nel disegno vengono creati prendendo in considerazione tutte le possibili combinazioni dei termini dipendenti e abbinando ogni combinazione a ciascun termine incluso nell'elenco del modello. Se è selezionata l'opzione Includi costante per la dipendente, nell'elenco verrà aggiunto anche un termine di unità (1).

Si supponga, ad esempio, che le variabili  $D1$  e  $D2$  siano le variabili dipendenti. Con la procedura Analisi loglineare Logit verrà creato un elenco dei termini dipendenti ( $D1$ ,  $D2$ ,  $D1*D2$ ). Se l'elenco Termini nel modello include  $M1$  e  $M2$  e si specifica una costante, l'elenco includerà

1,  $M1$  e  $M2$ . Il disegno finale includerà le combinazioni di ciascun termine del modello con ciascun termine dipendente:

$D1$ ,  $D2$ ,  $D1*D2$

$M1*D1$ ,  $M1*D2$ ,  $M1*D1*D2$

$M2*D1$ ,  $M2*D2$ ,  $M2*D1*D2$

**Includi costante per la dipendente.** Consente di includere una costante per la variabile dipendente in un modello personalizzato.

### Costruisci termini

Per i fattori e le covariate selezionati:

**Interazione.** Consente di creare il termine di interazione di livello maggiore rispetto a tutte le variabili selezionate. È l'impostazione predefinita.

**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.

### Analisi loglineare Logit: Opzioni

Figura 11-3  
Finestra di dialogo Analisi loglineare Logit: Opzioni

**Analisi loglineare Logit: Opzioni**

**Visualizza**

- Frequenze
- Residui
- Matrice di disegno
- Stime
- Cronologia delle iterazioni

**Grafico**

- Residui corretti
- Probabilità normale residui corretti
- Residui di devianza
- Probabilità normale residui di devianza

Intervallo di confidenza: 95 %

**Criteri**

Massimo numero di iterazioni: 20

Convergenza: 0.001

Delta: .5

Continua Annulla Aiuto

La procedura Analisi loglineare Logit consente di visualizzare informazioni e statistiche sulla bontà dell'adattamento relative al modello. È inoltre possibile selezionare una o più delle seguenti opzioni:

**Visualizzazione.** Possono essere visualizzate numerose statistiche: frequenze di cella osservate e attese, residui semplici, corretti e di devianza, una matrice di disegno e stime dei parametri per il modello.

**Grafico.** I grafici, disponibili solo con i modelli personalizzati, includono due matrici per grafico a dispersione (residui corretti o di devianza confrontati con conteggi di cella osservati e attesi). È inoltre possibile visualizzare la probabilità normale e grafici normali detrendizzati dei residui corretti o della devianza.

**Intervallo di confidenza.** L'intervallo di confidenza delle stime dei parametri può essere corretto.

**Criteri.** Il metodo di Newton-Raphson consente di ottenere le stime dei parametri della massima verosimiglianza. È possibile inserire nuovi valori per il numero massimo di iterazioni, per il criterio di convergenza e per il delta (una costante aggiunta in tutte le celle per approssimazioni iniziali). Nei modelli saturi il delta viene mantenuto nelle celle.

## Analisi loglineare Logit: Salva

Figura 11-4  
Finestra di dialogo Analisi loglineare Logit: Salva



Selezionare i valori che si desidera salvare come nuove variabili nel file dati attivo. Il suffisso  $n$  nel nome delle nuove variabili viene automaticamente incrementato di un'unità in modo da creare un nome univoco per ciascuna variabile salvata.

I valori salvati fanno riferimento ai dati aggregati, ovvero alle celle della tavola di contingenza, anche se tali dati sono registrati in osservazioni singole nell'Editor dei dati. Un valore residuo o atteso per dati non aggregati che viene salvato per una cella della tavola di contingenza viene inserito nell'Editor dei dati per ogni singolo caso incluso in tale cella. Per poter usare in modo utile i valori salvati, è consigliabile aggregare i dati in modo da ottenere i conteggi di cella.

È possibile salvare quattro tipi di residui: semplici, standardizzati, corretti e di devianza. È inoltre possibile salvare valori attesi.

- **Residui.** Chiamato anche residuo semplice, è la differenza tra il conteggio di cella osservato e il relativo conteggio atteso.
- **Residui standardizzati.** Il residuo diviso per una stima del suo errore standard. I residui standardizzati vengono anche detti residui di Pearson.

- **Residui corretti.** Il residuo standardizzato diviso per il proprio errore standard stimato. I residui corretti seguono asintoticamente una normale standardizzata quando il modello è appropriato. Per questo motivo sono preferibili ai residui standardizzati per verificare l'ipotesi di normalità.
- **Residui di devianza.** Il residuo di devianza è la radice quadrata con segno del contributo individuale alla statistica del Chi-quadrato del rapporto di verosimiglianza (G quadrato). Il segno è lo stesso del residuo (frequenza osservata meno frequenza attesa). La distribuzione dei residui di devianza è una normale asintotica standard.

## ***Opzioni aggiuntive del comando GENLOG***

Il linguaggio della sintassi dei comandi consente inoltre di:

- Calcolare combinazioni lineari di frequenze di celle osservate e attese e calcolare residui di stampa, residui standardizzati e residui corretti di tali combinazioni (tramite il sottocomando GERESID).
- Modificare il valore soglia predefinito per la verifica della ridondanza (tramite il sottocomando CRITERIA).
- Visualizzare i residui standardizzati (tramite il sottocomando PRINT).

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

## ***Tavole di sopravvivenza***

In molte situazioni può risultare utile esaminare la distribuzione dei tempi tra due eventi, ad esempio la durata dell'impiego, ovvero il periodo di tempo intercorso tra la data di assunzione e la data di interruzione dell'impiego. Questo tipo di dati tuttavia include in genere alcuni casi per i quali il secondo evento non viene registrato, ad esempio la presenza di persone ancora assunte dopo la fine dello studio. I motivi possono essere diversi. In alcuni casi l'evento semplicemente non si verifica prima della fine dello studio, in altri viene persa traccia del relativo stato prima della fine dello studio, mentre altri casi vengono interrotti per motivi non correlati allo studio, ad esempio quando un dipendente si ammala e rimane assente dal lavoro. In tutte queste situazioni si tratta di **casi troncati**, i quali rendono questo tipo di studio non adeguato per l'applicazione di tecniche tradizionali, quali i test  $t$  o la regressione lineare.

Una tecnica statistica che risulta utile con questo tipo di dati è definita **tavola di sopravvivenza** follow-up. Il concetto alla base della tavola di sopravvivenza prevede la suddivisione del periodo di osservazione in intervalli di tempo più brevi. Per ciascun intervallo, tutti i soggetti che sono stati sotto osservazione per almeno tale periodo di tempo vengono usati per il calcolo della probabilità di un evento terminale che si è verificato in tale intervallo di tempo. Le probabilità stimate in base a ciascun intervallo vengono quindi usate per la stima della probabilità globale dell'evento che si verifica in punti temporali diversi.

**Esempio.** Si intende valutare se una nuova terapia con cerotti di nicotina sia più efficace della terapia tradizionale per fare smettere di fumare le persone. È possibile condurre uno studio su due gruppi di fumatori curando il primo con la terapia tradizionale e il secondo con la terapia sperimentale. La creazione di tavole di sopravvivenza in base ai dati consente di confrontare il tasso di astinenza globale dei due gruppi in modo da stabilire se la cura sperimentale rappresenta o meno un miglioramento rispetto a quella tradizionale. È inoltre possibile tracciare le funzioni di sopravvivenza o di rischio e confrontare i grafici ottenuti in modo da ricavare informazioni più dettagliate.

**Statistiche.** Inserimento di numeri, esclusione di numeri, numero esposti a rischio, numero di eventi terminali, proporzione di terminazione, proporzione di sopravvivenza, proporzione di sopravvivenza cumulata (ed errore standard), densità di probabilità (ed errore standard), tasso di rischio (ed errore standard) per ogni intervallo di tempo di ciascun gruppo, tempo di sopravvivenza medio per ciascun gruppo e test di Wilcoxon (Gehan) per il confronto delle distribuzioni di sopravvivenza nei gruppi. Grafici: grafici di funzione per sopravvivenza, sopravvivenza logaritmica, densità e tasso di rischio e uno meno sopravvivenza.

**Dati.** La variabile tempo deve essere quantitativa. La variabile di stato deve essere dicotomica o categoriale, codificata in forma di interi e con gli eventi codificati in forma di valore singolo o intervallo di valori consecutivi. Le variabili fattore devono essere categoriali e codificate in forma di interi.

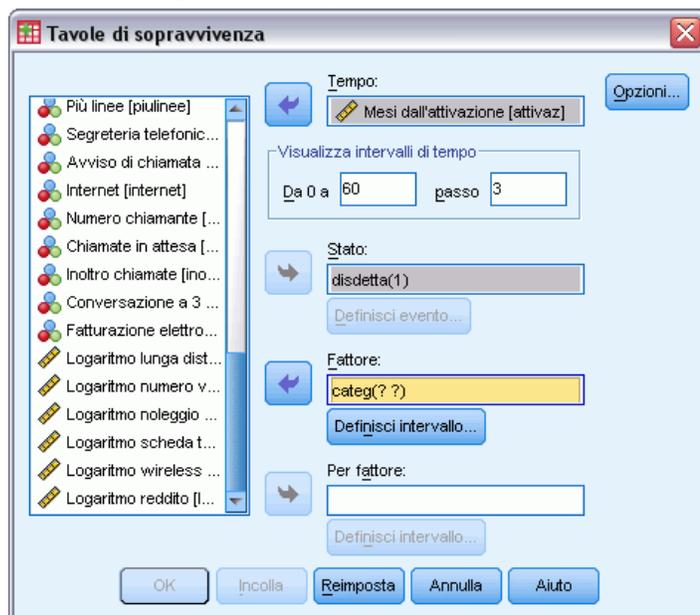
**Assunzioni.** Le probabilità per l'evento desiderato devono dipendere solo dal tempo successivo all'evento iniziale, ovvero si presume che siano stabili in relazione al tempo assoluto. Ciò significa che i casi aggiunti allo studio in momenti diversi, ad esempio i pazienti che iniziano la cura in un momento successivo, devono produrre risultati simili. Non devono inoltre esistere differenze sistematiche tra i casi in cui non si è verificato l'evento e i casi in cui l'evento si è invece manifestato. Se, ad esempio, molti dei casi in cui non si è verificato l'evento corrispondono a pazienti più gravi, i risultati potrebbero essere non imparziali.

**Procedure correlate.** Nella procedura Tavole di sopravvivenza viene adottato un approccio attuariale a questo tipo di analisi, definita in genere Analisi di sopravvivenza. Nella procedura Analisi di sopravvivenza Kaplan-Meier viene adottato un metodo di calcolo delle tavole di sopravvivenza leggermente diverso, non basato sulla suddivisione del periodo di osservazione in intervalli di tempo più brevi. È consigliabile adottare questo metodo quando il numero di osservazioni è ridotto, ad esempio con un numero ridotto di osservazioni in ciascun intervallo di tempo di sopravvivenza. Se esistono variabili che si suppone possano essere correlate al periodo alle variabili di sopravvivenza da controllare (covariate), usare la procedura Regressione di Cox. Se alle covariate in uso possono essere associati valori diversi in punti temporali diversi dello stesso caso, usare la regressione di Cox con covariate dipendenti dal tempo.

### Per creare una tavola di sopravvivenza

- Dai menu, scegliere:  
Analizza > Sopravvivenza > Tavole di sopravvivenza...

Figura 12-1  
Finestra di dialogo Tavole di sopravvivenza



- Selezionare una variabile di sopravvivenza di tipo **numerico**.
- Specificare gli intervalli di tempo da esaminare.

- ▶ Selezionare una variabile di stato per la definizione dei casi per i quali l'evento terminale si è verificato.
- ▶ Fare clic su Definisci evento per specificare il valore della variabile di stato che indica che si è verificato un evento.

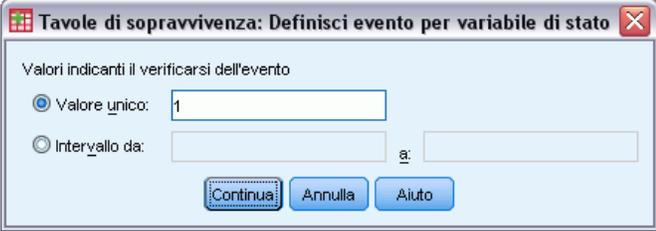
È inoltre possibile selezionare i casi utilizzando una variabile fattore di ordine 1. Le tavole attuariali per la variabile di sopravvivenza vengono generate per ciascuna categoria della variabile fattore.

È inoltre possibile selezionare una variabile *per fattore* di ordine 2. Le tabelle attuariali per la variabile di sopravvivenza vengono generate per ciascuna combinazione delle variabili fattore di ordine 1 e 2.

## Tavole di sopravvivenza: Definisci evento per variabile di stato

Figura 12-2

Finestra di dialogo Tavole di sopravvivenza: Definisci evento per variabile di stato

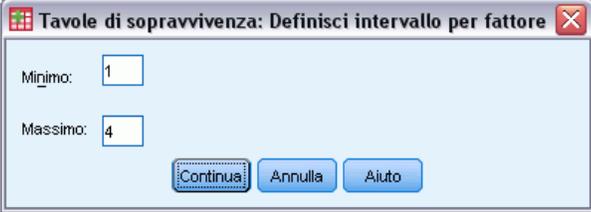


Le occorrenze del valore o dei valori selezionati per la variabile di stato indicano che per tali casi l'evento terminale si è verificato. Tutti gli altri casi sono considerati casi in cui l'evento non si è verificato. Inserire un singolo valore o un intervallo di valori che identifichi l'evento desiderato.

## Tavole di sopravvivenza: Definisci intervallo

Figura 12-3

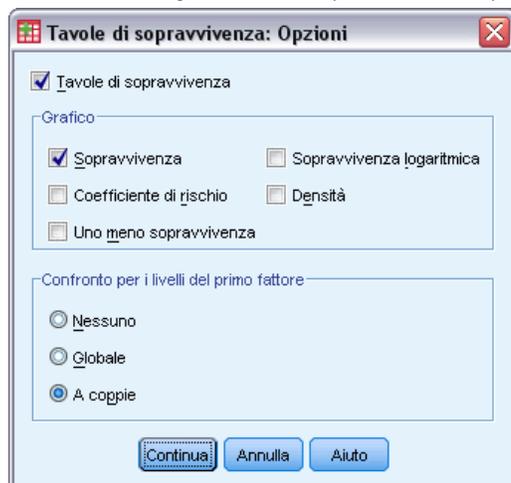
Finestra di dialogo Tavole di sopravvivenza: Definisci intervallo



I casi a cui è associato un valore per la variabile fattore compreso nell'intervallo specificato verranno inclusi nell'analisi e per ogni valore compreso nell'intervallo verranno generate tavole distinte e, se richiesto, grafici.

## Tavole di sopravvivenza: Opzioni

Figura 12-4  
Finestra di dialogo Tavole di sopravvivenza: Opzioni



È possibile controllare diversi aspetti dell'analisi Tavole di sopravvivenza.

**Tavole di sopravvivenza.** Per disattivare la visualizzazione delle tavole di sopravvivenza nell'output, deselezionare l'opzione Tavole di sopravvivenza.

**Grafico.** Consente di richiedere grafici delle funzioni di sopravvivenza. Se sono state definite variabili fattore, verranno generati grafici per ciascun sottogruppo definito da tali variabili. Sono disponibili i grafici di sopravvivenza, sopravvivenza logaritmica, rischio, densità e uno meno sopravvivenza.

- **Sopravvivenza.** Visualizza la funzione di sopravvivenza cumulata in una scala lineare.
- **Sopravvivenza logaritmica.** Visualizza la funzione di sopravvivenza cumulata in scala logaritmica.
- **Coefficiente di rischio (Kaplan-Meier).** Visualizza la funzione di rischio cumulato in scala lineare.
- **Densità.** Visualizza la funzione di densità.
- **1-Sopravvivenza.** Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.

**Confronto per i livelli del primo fattore.** Se è disponibile una variabile di controllo di ordine 1, è possibile selezionare una delle opzioni di questo gruppo per eseguire il test di Wilcoxon (Gehan), con cui viene confrontata la sopravvivenza dei sottogruppi. I test vengono eseguiti sul fattore di ordine 1. Se è stato definito un fattore di ordine 2, i test verranno eseguiti per ciascun livello della variabile di ordine 2.

## Opzioni aggiuntive del comando SURVIVAL

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare più di una variabile dipendente.

- Specificare intervalli equamente distanziati.
- Specificare più di una variabile di stato.
- Specificare confronti che non includono tutte le variabili fattore e di controllo.
- Calcolare confronti approssimati, anziché esatti.

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

# ***Analisi di sopravvivenza di Kaplan-Meier***

In molte situazioni può risultare utile esaminare la distribuzione dei tempi tra due eventi, ad esempio la durata dell'impiego, ovvero il periodo di tempo intercorso tra la data di assunzione e la data di interruzione dell'impiego. Questo tipo di dati tuttavia include in genere alcuni casi troncati. Si tratta di casi per i quali il secondo evento non viene registrato, ad esempio la presenza di persone ancora assunte alla fine dello studio. La procedura Kaplan-Meier rappresenta un metodo per la stima di modelli con relazione tempo-eventi in presenza di casi in cui non si è verificato l'evento. Il modello Kaplan-Meier si basa sulla stima delle probabilità condizionate in ciascun punto temporale in cui si verifica un evento e sull'uso del limite di tali probabilità per la stima del tasso di sopravvivenza in ciascun punto.

**Esempio.** Una nuova cura per l'AIDS ha effetti terapeutici benefici sul prolungamento della durata della vita? È possibile condurre uno studio su due gruppi di pazienti affetti da AIDS curando il primo gruppo in base alla terapia tradizionale e il secondo in base alla cura sperimentale. La creazione di un modello di Kaplan-Meier in base ai dati consente di confrontare il tasso di sopravvivenza globale dei due gruppi in modo da stabilire se la cura sperimentale rappresenti o meno un miglioramento rispetto a quella tradizionale. È inoltre possibile tracciare le funzioni di sopravvivenza o di rischio e confrontare i grafici ottenuti in modo da ricavare informazioni più dettagliate.

**Statistiche.** La tavola di sopravvivenza che include tempo, stato, sopravvivenza cumulativa, errore standard, eventi cumulati e rimanenti, il tempo di sopravvivenza media e mediana con errore standard e intervallo di confidenza del 95%. Grafici: sopravvivenza, rischio, sopravvivenza logaritmica, uno meno sopravvivenza.

**Dati.** La variabile tempo deve essere continua, la variabile di stato può essere categoriale o continua, mentre le variabili fattore e di stratificazione devono essere categoriali.

**Assunzioni.** Le probabilità per l'evento desiderato devono dipendere solo dal tempo successivo all'evento iniziale, ovvero si presume che siano stabili in relazione al tempo assoluto. Ciò significa che i casi aggiunti allo studio in momenti diversi, ad esempio i pazienti che iniziano la cura in un momento successivo, devono produrre risultati simili. Non devono inoltre esistere differenze sistematiche tra i casi in cui non si è verificato l'evento e i casi in cui l'evento si è invece manifestato. Se, ad esempio, molti dei casi in cui non si è verificato l'evento corrispondono a pazienti più gravi, i risultati potrebbero essere non imparziali.

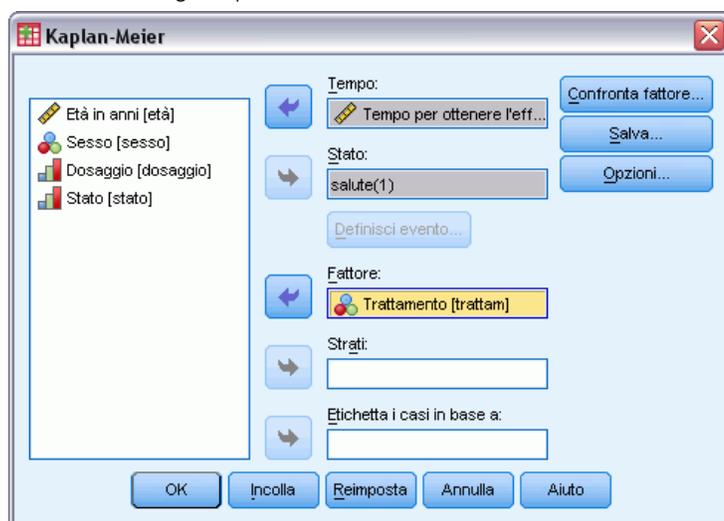
**Procedure correlate.** Nella procedura Kaplan-Meier viene usato un metodo di calcolo delle tavole di sopravvivenza con cui viene eseguita la stima della funzione di sopravvivenza o di rischio al verificarsi di ciascun evento. Nella procedura Tavole di sopravvivenza viene adottato un approccio attuariale in relazione all'analisi di sopravvivenza, basato sulla suddivisione del periodo di

osservazione in intervalli di tempo più brevi e che può risultare utile per la gestione di campioni di grandi dimensioni. Se esistono variabili che si suppone possano essere correlate al periodo o alle variabili di sopravvivenza da controllare (covariate), usare la procedura Regressione di Cox. Se alle covariate in uso possono essere associati valori diversi in punti temporali diversi dello stesso caso, usare la regressione di Cox con covariate dipendenti dal tempo.

### **Per ottenere un'analisi di sopravvivenza Kaplan-Meier**

- Dai menu, scegliere:  
Analizza > Sopravvivenza > Kaplan-Meier...

Figura 13-1  
Finestra di dialogo Kaplan-Meier



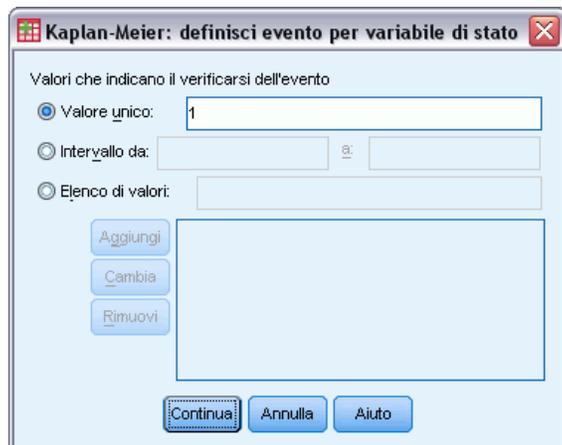
- Selezionare una variabile tempo.
- Selezionare una variabile di stato per l'identificazione dei casi per i quali si è verificato l'evento terminale. Questa variabile può essere di tipo numerico oppure una **stringa corta**. Fare quindi clic su Definisci evento.

Se necessario, è possibile selezionare una variabile fattore per esaminare le differenze di gruppo. È inoltre possibile selezionare una variabile di stratificazione con cui si ottengono analisi distinte per ciascun livello o strato della variabile.

## Kaplan-Meier: Definisci evento per variabile di stato

Figura 13-2

Finestra di dialogo Kaplan-Meier: Definisci evento per variabile di stato

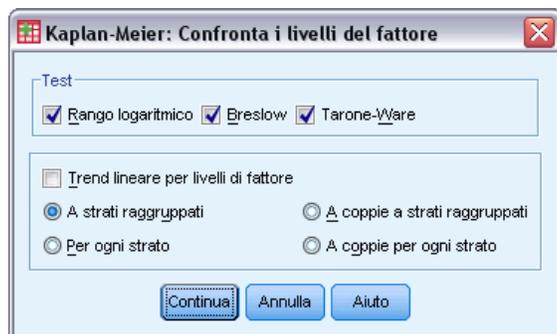


Inserire il valore o i valori per indicare che l'evento terminale si è verificato. È possibile inserire un solo valore, oppure un intervallo o un elenco di valori. L'opzione Intervallo da a è disponibile solo se la variabile di stato in uso è di tipo numerico.

## Kaplan-Meier: Confronta i livelli del fattore

Figura 13-3

Finestra di dialogo Kaplan-Meier: Confronta i livelli del fattore



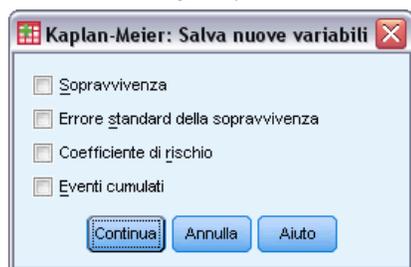
È possibile richiedere statistiche per verificare l'uguaglianza delle distribuzioni di sopravvivenza tra i vari livelli del fattore. Le statistiche disponibili sono Rango logaritmico, Breslow e Tarone-Ware. Selezionare una delle alternative per specificare i confronti da eseguire: a strati raggruppati, per ogni strato, per coppie a strati raggruppati e per coppie per ogni strato.

- **Rango logaritmico.** Un test per confrontare l'uguaglianza di distribuzioni di sopravvivenza. In questo test tutti i punti temporali vengono pesati ugualmente.
- **Breslow.** Un test per confrontare l'uguaglianza di distribuzioni di sopravvivenza. I punti temporali vengono ponderati dal numero di casi a rischio in ogni punto.

- **Tarone-Ware.** Un test per confrontare l'uguaglianza delle distribuzioni di sopravvivenza. I punti temporali sono ponderati per la radice quadrata del numero dei casi a rischio in ogni punto.
  - **A strati raggruppati.** Confronta tutti i livelli simultaneamente per verificare l'uguaglianza delle curve di sopravvivenza.
  - **A coppie a strati raggruppati.** Confronta i livelli del fattore a coppie. Non disponibile per i test del trend a coppie.
  - **Per ogni strato.** Esegue un test separato di uguaglianza di tutti i livelli di fattore per ogni strato. In mancanza della variabile di stratificazione i test non verranno eseguiti.
  - **A coppie per ogni strato.** Confronta i livelli del fattore a coppie, ripetendo il test per ogni categoria della variabile Strati definita nella finestra principale. Non disponibile per i test del trend a coppie. In mancanza della variabile di stratificazione i test non verranno eseguiti.
- Trend lineare per livelli di fattore.** Consente di verificare un trend lineare tra i vari livelli del fattore. Questa opzione è disponibile solo per confronti globali, anziché per coppie, dei livelli del fattore.

## Kaplan-Meier: Salva nuove variabili

Figura 13-4  
Finestra di dialogo Kaplan-Meier: Salva nuove variabili

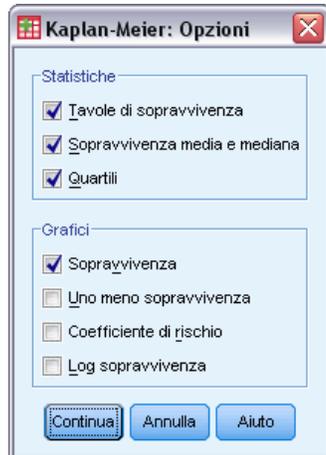


Le informazioni della tavola Kaplan-Meier possono essere salvate come nuove variabili, che è quindi possibile usare in analisi successive per la verifica delle ipotesi e il controllo delle assunzioni. È possibile salvare come nuove variabili eventi di sopravvivenza, di errore di sopravvivenza standard, di rischio e cumulativi.

- **Sopravvivenza.** La stima di probabilità di sopravvivenza cumulata. Il nome predefinito delle variabili è costituito da un prefisso `sur_` e da un numero sequenziale. Se, ad esempio, esiste già la variabile `sur_1`, verrà assegnato il nome `sur_2`.
- **Errore standard della sopravvivenza.** L'errore standard della stima di sopravvivenza cumulata. Il nome delle variabili predefinito è composto da un prefisso `cum_` e da un numero progressivo. Se, ad esempio, esiste già la variabile `se_1`, verrà assegnato il nome `se_2`.
- **Coefficiente di rischio (Kaplan-Meier).** Stima della funzione di rischio cumulato. Le variabili hanno il nome composto da un prefisso `haz_` e da un numero progressivo. Se, ad esempio, esiste già la variabile `haz_1`, verrà assegnato il nome `haz_2`.
- **Eventi cumulati.** Frequenza cumulata degli eventi quando i casi sono ordinati per tempo di sopravvivenza e stato. Le variabili hanno il nome composto da un prefisso `cum_` e da un numero progressivo. Se, ad esempio, esiste già la variabile `cum_1`, verrà assegnato il nome `cum_2`.

## Kaplan-Meier: Opzioni

Figura 13-5  
Finestra di dialogo Kaplan-Meier: Opzioni



Quando si esegue l'analisi Kaplan-Meier, è possibile richiedere vari tipi di output.

**Statistiche.** È possibile selezionare le statistiche visualizzate per le funzioni di sopravvivenza calcolate, incluse le tavole di sopravvivenza, la sopravvivenza media e mediana e i quartili. Se sono state incluse variabili fattore, per ciascun gruppo verranno generate statistiche distinte.

**Grafici.** I grafici consentono di esaminare visivamente le funzioni di sopravvivenza, uno meno sopravvivenza e di sopravvivenza logaritmica. Se sono state incluse variabili fattore, le funzioni verranno tracciate per ciascun gruppo.

- **Sopravvivenza.** Visualizza la funzione di sopravvivenza cumulata in una scala lineare.
- **1-Sopravvivenza.** Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.
- **Coefficiente di rischio (Kaplan-Meier).** Visualizza la funzione di rischio cumulato in scala lineare.
- **Sopravvivenza logaritmica.** Visualizza la funzione di sopravvivenza cumulata in scala logaritmica.

## Opzioni aggiuntive del comando KM

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere le tabelle di frequenza in cui vengono considerati i casi persi da seguire come categoria distinta rispetto ai casi troncati.
- Specificare una distanza disuguale per la verifica del trend lineare.
- Ottenere percentili diversi dai quartili per la variabile del tempo della sopravvivenza.

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

# ***Analisi della regressione di Cox***

La Regressione di Cox crea un modello predittivo per i dati della relazione tempo-evento. Il modello crea una funzione di sopravvivenza che prevede la probabilità che l'evento di interesse si sia verificato in un dato periodo  $t$  per valori specifici delle variabili predittore. La forma della funzione di sopravvivenza e i coefficienti di regressione dei predittori vengono stimati da soggetti osservati; il modello può essere quindi applicato a nuovi casi che hanno misurazioni per le variabili predittore. Tenere presente che le informazioni dei soggetti troncati, ovvero quelli che non vivono l'evento di interesse durante il periodo di osservazione, contribuiscono efficacemente alla stima del modello.

**Esempio.** Per uomini e donne esiste una diversa percentuale di rischio di cancro ai polmoni causato dal fumo? Creando un modello di regressione di Cox che prevede come covariate il numero di sigarette fumate al giorno e il sesso, è possibile verificare ipotesi diverse relative all'incidenza di sesso e numero di sigarette sulla relazione tempo-sintomi iniziali per il cancro ai polmoni.

**Statistiche.** Per ciascun modello:  $-2LL$ , la statistica del rapporto di verosimiglianza e il chi-quadrato complessivo. Per le variabili del modello: stime dei parametri, errori standard e statistiche di Wald. Per variabili non incluse nel modello: punteggi e chi-quadrato residuo.

**Dati.** La variabile del tempo deve essere quantitativa, ma quella dello stato può essere categoriale o continua. Le variabili indipendenti (covariate) possono essere continue o categoriali. Se sono categoriali, devono essere fittizie o codificate in forma di indicatori. Nella procedura è disponibile un'opzione per la ricodifica automatica delle variabili categoriali. Le variabili di stratificazione devono essere categoriali e codificate in forma di interi o stringhe brevi.

**Assunzioni.** Le osservazioni devono essere indipendenti e il tasso di rischio deve essere costante nel tempo. Ciò significa che la proporzionalità tra i rischi di un caso e quelli di un altro caso non deve variare nel corso del tempo. L'ultima assunzione è definita **ipotesi di proporzionalità del rischio**.

**Procedure correlate.** Se l'ipotesi di proporzionalità del rischio non è valida (vedi sopra), potrebbe essere necessario usare la procedura di Cox con covariate dipendenti dal tempo. Se non è disponibile alcuna covariata oppure è disponibile una sola covariata categorica, è possibile usare la procedura Tavole di sopravvivenza o Kaplan-Meier per esaminare le funzioni di sopravvivenza o di rischio dei campioni studiati. Se il campione non include dati troncati, ovvero in tutti i casi si è manifestato l'evento terminale, è possibile usare la procedura Regressione lineare per definire la relazione tra stimatori e tempo-evento.

## ***Per ottenere un'analisi Regressione di Cox***

- Dai menu, scegliere:  
Analizza > Sopravvivenza > Regressione di Cox...

Figura 14-1  
Finestra di dialogo Regressione di Cox

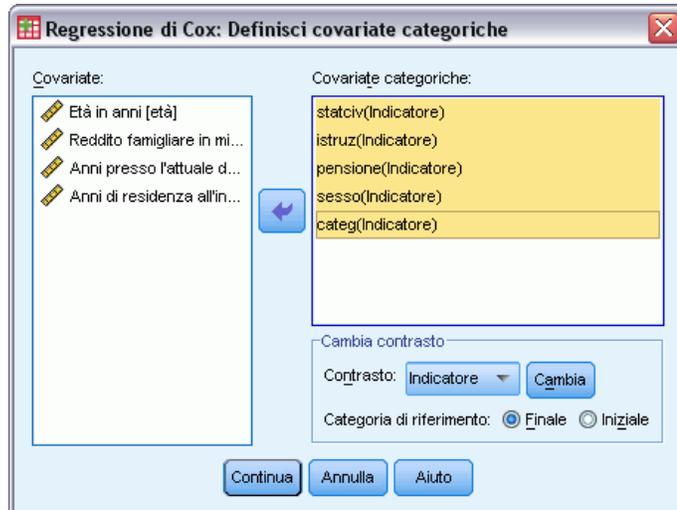


- ▶ Selezionare una variabile tempo. I casi con valori dell'ora negativi non vengono analizzati.
- ▶ Selezionare una variabile di stato, quindi fare clic su Definisci evento.
- ▶ Selezionare una o più covariate. Per includere i termini di interazione, selezionare tutte le variabili coinvolte nell'interazione e quindi fare clic su >a\*b>.

Se necessario, è possibile calcolare modelli distinti per gruppi diversi tramite la definizione di una variabile di stratificazione.

## Regressione di Cox: Definisci variabili categoriali

Figura 14-2  
Finestra di dialogo Regressione di Cox: Definisci covariate categoriche



È possibile specificare la modalità con cui le variabili categoriali vengono gestite dalla procedura Regressione di Cox.

**Covariate.** Contiene un elenco di tutte le covariate specificate in qualsiasi strato della finestra di dialogo principale, da sole o come parte di un'interazione. Se alcune di queste variabili sono stringa o categoriali, è possibile utilizzarle solo come covariate categoriche.

**Covariate categoriche.** Elenca le variabili identificate come categoriali. Ogni variabile include una notazione tra parentesi che indica la codificazione di contrasto da utilizzare. Variabili stringa (codificate con il simbolo < dopo i relativi nomi) sono già presenti nell'elenco Covariate categoriche. Selezionare altre covariate categoriche dall'elenco di covariate e spostarle nell'elenco di covariate categoriche.

**Cambia.** Consente di modificare il metodo di contrasto. I metodi di contrasto disponibili sono:

- **Indicatore.** I contrasti indicano l'appartenenza o la non appartenenza alla categoria. La categoria di riferimento è rappresentata nella matrice di contrasto come una riga di zero.
- **Semplice.** Tutte le categorie della variabile stimatore (eccetto la categoria di riferimento) vengono confrontate con la categoria di riferimento.
- **Differenza.** Tutte le categorie della variabile stimatore (eccetto la prima categoria) vengono confrontate con l'effetto medio delle categorie precedenti. Sono noti anche come contrasti inversi di Helmert.
- **Helmert.** Tutte le categorie della variabile stimatore (eccetto l'ultima categoria) vengono confrontate con l'effetto medio delle categorie successive.
- **Ripetuto.** Tutte le categorie della variabile stimatore (eccetto la prima categoria) vengono confrontate con la categoria che le precede.

- **Polinomiale.** Contrasti polinomiali ortogonali. Si presume che le categorie siano equamente distanziate. I contrasti polinomiali sono disponibili solo per le variabili numeriche.
- **standard.** Tutte le categorie della variabile stimatore (eccetto la categoria di riferimento) vengono confrontate con l'effetto globale.

Se si seleziona Deviazione, Semplice o Indicatore, selezionare Iniziale o Finale come categoria di riferimento. Si noti che il metodo non viene effettivamente modificato finché non si fa clic su Cambia.

Le covariate di tipo stringa devono essere covariate categoriche. Per rimuovere una variabile di tipo stringa dall'elenco Covariate categoriche, è necessario rimuovere tutti i termini contenenti la variabile dall'elenco di covariate nella finestra di dialogo principale.

## Regressione di Cox: Grafici

Figura 14-3  
Finestra di dialogo Regressione di Cox: Grafici



I grafici agevolano la valutazione del modello stimato e l'interpretazione dei risultati. È possibile tracciare le funzioni di sopravvivenza, rischio e trasformazione  $\ln(-\ln)$ .

- **Sopravvivenza.** Visualizza la funzione di sopravvivenza cumulata in una scala lineare.
- **Coefficiente di rischio.** Visualizza la funzione di rischio cumulato in scala lineare.
- **Trasformazione  $\ln(-\ln)$ .** La stima di sopravvivenza cumulata dopo che la trasformazione  $\ln(-\ln)$  è stata applicata alla stima.
- **Uno meno sopravvivenza.** Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.

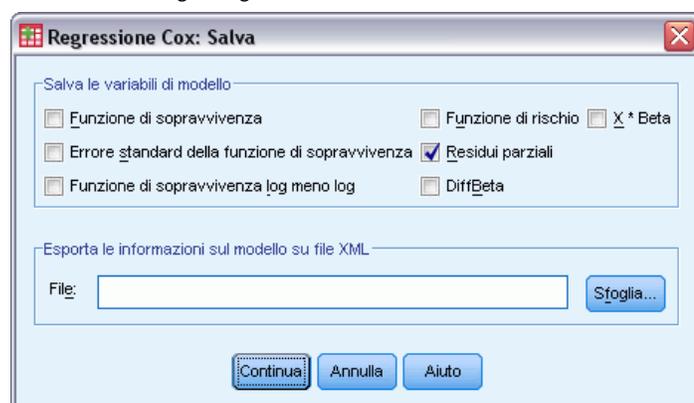
Dato che queste funzioni dipendono dai valori delle covariate, per tracciare le funzioni in rapporto al tempo è necessario usare valori di costanti per le covariate. Per impostazione predefinita viene usata come valore costante la media di ciascuna covariata. È tuttavia possibile inserire i valori desiderati per il grafico usando le opzioni del gruppo di controllo Cambia valore.

È possibile tracciare una linea distinta per ciascun valore di una covariata categorica spostando la covariata nella casella di testo Linee separate per. Questa opzione è disponibile solo per le covariate categoriche, contraddistinte da (Cat) che segue il corrispondente nome nell'elenco Valori della covariata tracciati a.

## Regressione di Cox: Salva nuove variabili

Figura 14-4

Finestra di dialogo Regressione di Cox: Salva nuove variabili



È possibile salvare i vari risultati dell'analisi come nuove variabili. Le variabili possono quindi essere usate in analisi successive per valutare le ipotesi o controllare i presupposti.

**Salva le variabili di modello.** Consente di salvare la funzione di sopravvivenza e il relativo errore standard, le stime  $\ln(-\ln)$ , la funzione di rischio, i residui parziali, i DiffBeta per la regressione e il predittore lineare  $X \cdot \text{Beta}$  come nuove variabili.

- **Funzione di sopravvivenza.** Il valore della funzione di sopravvivenza cumulata per un determinato periodo. Equivale alla probabilità di sopravvivenza in quel periodo di tempo.
- **Funzione di sopravvivenza log meno log.** La stima di sopravvivenza cumulata dopo che la trasformazione  $\ln(-\ln)$  è stata applicata alla stima.
- **Funzione di rischio.** Salva la stima della funzione di rischio cumulato (nota anche come residuo di Cox-Snell)
- **Residui parziali.** I residui parziali possono essere rappresentati in un grafico con i tempi di sopravvivenza sull'altro asse, per verificare l'ipotesi di proporzionalità del rischio. Viene salvata una variabile per ogni covariata. Questi residui sono disponibili solo per i modelli che contengono almeno una covariata.

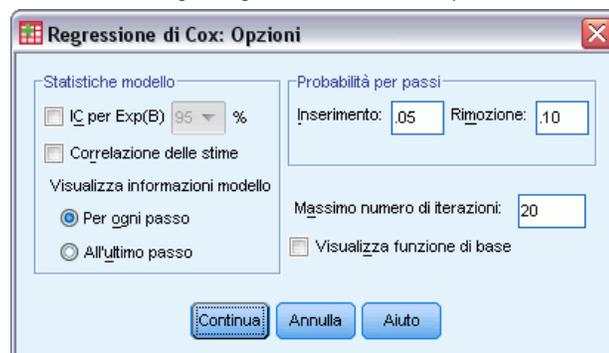
- **DiffBeta.** La stima del cambiamento di un coefficiente se un caso viene rimosso. Viene salvata una variabile per ogni covariata. Queste stime sono disponibili solo per i modelli che contengono almeno una covariata.
- **X\*Beta.** Punteggio predittore lineare. La somma del prodotto dei valori della covariata centrati sulla media e le corrispondenti stime di parametro per ogni caso.

Se si esegue la regressione di Cox con una covariata dipendente dal tempo, le uniche variabili che è possibile salvare saranno DiffBeta e il predittore lineare X\*Beta.

**Esporta le informazioni sul modello su file XML.** Le stime dei parametri vengono esportate nel file specificato in formato XML. È possibile utilizzare questo file di modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

## Regressione di Cox: Opzioni

Figura 14-5  
Finestra di dialogo Regressione di Cox: Opzioni



È possibile controllare diversi aspetti dell'analisi e dell'output.

**Statistiche modello.** È possibile ottenere statistiche per i parametri del modello, compresi gli intervalli di confidenza per  $\exp(B)$  e la correlazione di stime. È possibile richiedere queste statistiche in corrispondenza di ogni passo oppure solo dell'ultimo.

**Probabilità per Stepwise.** Se è stato selezionato un metodo stepwise, è possibile specificare la probabilità per ciascun inserimento o rimozione eseguiti nel modello. Una variabile viene inserita quando il livello di significatività del corrispondente valore  $F$  di inserimento è minore del valore Inserimento, mentre viene rimossa se il livello di significatività è maggiore del valore Rimozione. Il valore Inserimento deve essere minore del valore Rimozione.

**Massimo numero di iterazioni.** Consente di specificare il numero massimo di iterazioni per il modello, in modo da controllare la durata della ricerca di una soluzione durante l'esecuzione della procedura.

**Visualizza funzione di base.** Consente di visualizzare la funzione di rischio di base e la sopravvivenza cumulativa in corrispondenza della media delle covariate. Tale visualizzazione non è disponibile se sono state specificate covariate dipendenti dal tempo.

## ***Regressione di Cox: Definisci evento per la variabile di stato***

Inserire il valore o i valori per indicare che l'evento terminale si è verificato. È possibile inserire un solo valore, oppure un intervallo o un elenco di valori. L'opzione Intervallo da a è disponibile solo se la variabile di stato in uso è di tipo numerico.

### ***Opzioni aggiuntive del comando COXREG***

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere le tabelle di frequenza in cui vengono considerati i casi persi da seguire come categoria distinta rispetto ai casi troncati.
- Selezionare una categoria di riferimento, diversa dalla prima e dall'ultima, per i metodi di contrasto Deviazione, Semplice e Indicatore.
- Specificare una spaziatura disuguale delle categorie per il metodo di contrasto Polinomiale.
- Specificare criteri di iterazione aggiuntivi.
- Controllare la gestione dei valori mancanti.
- Specificare il nome delle variabili salvate.
- Scrivere output in un file di dati di IBM® SPSS® Statistics esterno.
- Mantenere i dati di ogni singolo gruppo in un file temporaneo esterno durante l'elaborazione. Ciò consente di risparmiare risorse di memoria durante l'esecuzione di analisi che implicano l'uso di insiemi di dati di grandi dimensioni. Questa opzione non è disponibile con le covariate dipendenti dal tempo.

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

## ***Calcolo di covariate dipendenti dal tempo***

Esistono situazioni in cui si desidera calcolare un modello di regressione di Cox ma in cui l'ipotesi del rischio proporzionale non è valida. Ovvero, i rapporti di rischio variano nel tempo e i valori di una (o più) covariate sono diversi in momenti diversi. In questi casi, è necessario usare un modello di regressione di Cox esteso che consenta di specificare le **covariate dipendenti dal tempo**.

Per poter analizzare un modello di questo tipo, è necessario definire innanzitutto la covariata dipendente dal tempo (con l'appropriata sintassi del comando, è anche possibile specificare più covariate dipendenti dal tempo). Per semplificare questa operazione, è disponibile una variabile di sistema per il tempo, denominata  $T_$ . La variabile può essere usata in due modi diversi:

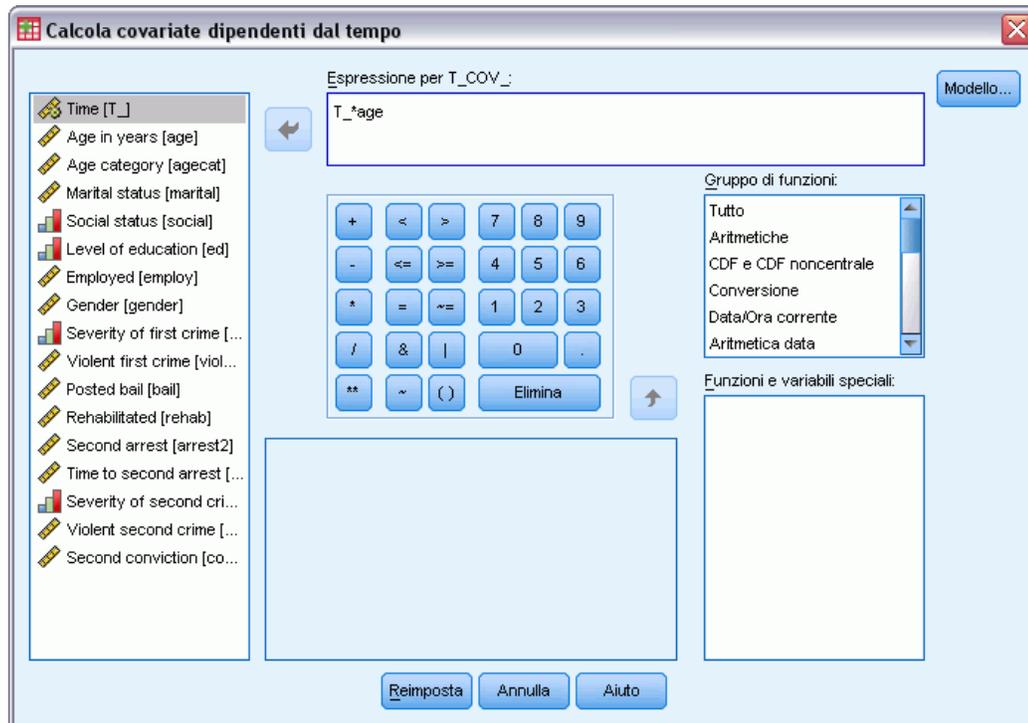
- Se si desidera verificare l'ipotesi della proporzionalità del rischio in relazione a una covariata specifica oppure stimare un modello di regressione di Cox esteso che consenta la presenza di rischi non proporzionali, è possibile definire la covariata dipendente dal tempo come funzione della variabile di tempo  $T_$  e della covariata specificata. Un esempio comune è il prodotto semplice della variabile di tempo e della covariata. È tuttavia possibile specificare anche funzioni più complesse. La verifica della significatività del coefficiente della covariata dipendente dal tempo consente di stabilire se l'ipotesi della proporzionalità del rischio è accettabile.
- 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 una **covariata dipendente dal tempo** 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 una covariata 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  $BP4$ ), la covariata dipendente dal tempo può essere definita 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 un caso specifico, uno solo dei termini tra parentesi sarà uguale a uno, mentre tutti gli altri saranno uguali a 0. Questa funzione può essere interpretata nel modo seguente: se il periodo di tempo è inferiore a una settimana, usare  $BP1$ , se è superiore a una settimana ma inferiore a due, usare  $BP2$  e così via.

Nella finestra di dialogo Calcola covariate dipendenti dal tempo è possibile usare i controlli per la costruzione della funzione per creare l'espressione per la covariata dipendente dal tempo, oppure è possibile inserirla direttamente nel riquadro di testo Espressione per  $T\_COV_$ . 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 definita  $T\_COV_$  e deve essere inclusa nel modello di regressione di Cox come covariata.

## Per calcolare una covariata dipendente dal tempo

- Dai menu, scegliere:  
Analizza > Sopravvivenza > Cox con covariate dipendenti dal tempo...

Figura 15-1  
Finestra di dialogo Calcola covariate dipendenti dal tempo



- Inserire un'espressione per la covariata dipendente dal tempo.
- Fare clic su Modello per procedere con la regressione di Cox.

*Nota:* Verificare che la nuova variabile  $T\_COV\_$  sia inclusa nel modello di regressione di Cox come covariata.

Per ulteriori informazioni, vedere l'argomento [Analisi della regressione di Cox in il capitolo 14 a pag. 146](#).

## Opzioni aggiuntive della regressione di Cox con covariate dipendenti dal tempo

Il linguaggio della sintassi dei comandi consente inoltre di specificare più covariate dipendenti dal tempo. Sono disponibili funzioni aggiuntive della sintassi di comando per la regressione di Cox con o senza covariate dipendenti dal tempo.

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

# ***Schemi di codifica delle variabili categoriali***

In molte procedure, è possibile richiedere la sostituzione automatica di una variabile categoriale indipendente con un insieme di variabili di contrasto, che verranno poi inserite o rimosse da un'equazione come blocco. È possibile specificare la modalità di codifica delle variabili di contrasto, solitamente nel sottocomando `CONTRAST`. In questa appendice viene spiegata e illustrata la diversità di funzionamento dei vari tipi di contrasto richiesti nel sottocomando `CONTRAST`.

## ***Deviazione***

**Deviazione dalla media principale.** In termini di matrici, questi contrasti assumono la forma di:

media	(	$1/k$	$1/k$	...	$1/k$	$1/k$	)
df(1)	(	$1-1/k$	$-1/k$	...	$-1/k$	$-1/k$	)
df(2)	(	$-1/k$	$1-1/k$	...	$-1/k$	$-1/k$	)
.			.			.	
.			.			.	
df(k-1)	(	$-1/k$	$-1/k$	...	$1-1/k$	$-1/k$	)

dove  $k$  è il numero di categorie per la variabile indipendente e l'ultima categoria viene omessa per impostazione predefinita. Ad esempio, i contrasti di deviazione per una variabile indipendente con tre categorie sono i seguenti:

(	$1/3$	$1/3$	$1/3$	)
(	$2/3$	$-1/3$	$-1/3$	)
(	$-1/3$	$2/3$	$-1/3$	)

Per omettere una categoria diversa dall'ultima, specificare tra parentesi il numero della categoria omessa dopo la parola chiave `DEVIATION`. Ad esempio, il seguente sottocomando consente di recuperare le deviazioni per la prima e la terza categoria, omettendo la seconda:

`/CONTRAST (FACTOR) =DEVIATION (2)`

Si supponga che il *fattore* abbia tre categorie. La matrice di contrasto risultante sarà

$$\begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{pmatrix}$$

## Semplice

**Contrasti semplici.** Confronta ogni livello di un fattore con l'ultimo. La forma generale della matrice è

$$\begin{array}{l} \text{media} \\ \text{df}(1) \\ \text{df}(2) \\ \cdot \\ \cdot \\ \text{df}(k-1) \end{array} \begin{pmatrix} 1/k & 1/k & \dots & 1/k & 1/k \\ 1 & 0 & \dots & 0 & -1 \\ 0 & 1 & \dots & 0 & -1 \\ \cdot & \cdot & & & \\ \cdot & \cdot & & & \\ 0 & 0 & \dots & 1 & -1 \end{pmatrix}$$

dove  $k$  è il numero di categorie per la variabile indipendente. Ad esempio, i contrasti semplici per una variabile indipendente con quattro categorie sono i seguenti:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Per utilizzare un'altra categoria invece dell'ultima come categoria di riferimento, specificare tra parentesi, dopo la parola chiave `SIMPLE`, il numero di sequenza della categoria di riferimento, che non corrisponderà necessariamente al valore associato alla categoria. Ad esempio, il seguente sottocomando `CONTRAST` ottiene una matrice di contrasto che omette la seconda categoria:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Si supponga che il *fattore* abbia quattro categorie. La matrice di contrasto risultante sarà

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix}$$

## Helmert

**Contrasti di Helmert.** Confronta le categorie di una variabile indipendente con la media delle categorie successive. La forma generale della matrice è

media	( 1/k	1/k	...	1/k	1/k )
df(1)	( 1	-1/(k-1)	...	-1/(k-1)	-1/(k-1) )
df(2)	( 0	1	...	-1/(k-2)	-1/(k-2) )
.	.	.	.	.	.
df(k-2)	( 0	0	1	-1/2	-1/2 )
df(k-1)	( 0	0	...	1	-1 )

dove  $k$  è il numero di categorie per la variabile indipendente. Ad esempio, una variabile indipendente con quattro categorie avrà una matrice di contrasto di Helmert composta nel seguente modo:

( 1/4	1/4	1/4	1/4 )
( 1	-1/3	-1/3	-1/3 )
( 0	1	-1/2	-1/2 )
( 0	0	1	-1 )

## Differenza

**Differenza o contrasti inversi di Helmert.** Confronta le categorie di una variabile indipendente con la media delle categorie precedenti della variabile. La forma generale della matrice è

media	( 1/k	1/k	1/k	...	1/k )
df(1)	( -1	1	0	...	0 )
df(2)	( -1/2	-1/2	1	...	0 )
.	.	.	.	.	.
df(k-1)	( -1/(k-1)	-1/(k-1)	-1/(k-1)	...	1 )

dove  $k$  è il numero di categorie per la variabile indipendente. Ad esempio, i contrasti differenza per una variabile indipendente con quattro categorie sono i seguenti:

( 1/4	1/4	1/4	1/4 )
( -1	1	0	0 )
( -1/2	-1/2	1	0 )
( -1/3	-1/3	-1/3	1 )

## Polinomiale

**Contrasti polinomiali ortogonali.** Il primo grado di libertà include l'effetto lineare di tutte le categorie, il secondo grado di libertà l'effetto quadratico, il terzo grado di libertà l'effetto cubico e così via per gli effetti di ordine più alto.

È possibile specificare la distanza tra i diversi livelli del trattamento misurata in base alla variabile categoriale. La distanza uguale, per impostazione predefinita se si omette la metrica, può essere specificata con numeri interi consecutivi da 1 a  $k$ , dove  $k$  è il numero di categorie. Se la variabile relativa al *medicinale* ha tre categorie, il sottocomando

```
/CONTRAST (DRUG) = POLYNOMIAL
```

è uguale a

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 2, 3)
```

Non è tuttavia sempre necessaria una distanza uguale. Ad esempio, si supponga che *DRUG* rappresenti diversi dosaggi di medicinale somministrati a tre gruppi. Se il dosaggio somministrato al secondo gruppo è pari al doppio del primo gruppo e il dosaggio somministrato al terzo gruppo è il triplo di quello del primo gruppo, le categorie del trattamento presentano la stessa distanza. La metrica appropriata per questa situazione è costituita da numeri interi consecutivi:

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 2, 3)
```

Se, tuttavia, il dosaggio somministrato al secondo gruppo è quattro volte più grande rispetto a quello del primo gruppo e il dosaggio somministrato al terzo gruppo è sette volte quello del primo gruppo, la metrica appropriata sarà

```
/CONTRAST (DRUG) = POLYNOMIAL (1, 4, 7)
```

In ambedue i casi, il risultato della specifica del contrasto è che il primo grado di libertà per la variabile *drug* include l'effetto lineare dei livelli di dosaggio e il secondo grado di libertà include l'effetto quadratico.

I contrasti polinomiali si rivelano particolarmente utili nei test dei trend e nelle indagini sulla natura delle superfici di risposta. È inoltre possibile utilizzare i contrasti polinomiali per eseguire l'adattamento della curva non lineare, ad esempio nella regressione curvilineare.

## Ripetuto

**Confronta i livelli adiacenti di una variabile indipendente.** La forma generale della matrice è

media	( 1/k	1/k	1/k	...	1/k	1/k )
df(1)	( 1	-1	0	...	0	0 )
df(2)	( 0	1	-1	...	0	0 )
.	.	.	.	.	.	.
.	.	.	.	.	.	.
df(k-1)	( 0	0	0	...	1	-1 )

dove  $k$  è il numero di categorie per la variabile indipendente. Ad esempio, i contrasti ripetuti per una variabile indipendente con quattro categorie sono i seguenti:

$$\begin{pmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$

Questi contrasti risultano utili per l'analisi del profilo e in qualsiasi caso siano necessari punteggi di differenza.

## Contrasto speciale

**Un contrasto definito dall'utente.** Consente l'immissione di contrasti speciali come matrici quadrate con un numero di righe e di colonne pari al numero delle categorie della variabile indipendente. Per MANOVA e LOGLINEAR, la prima riga inserita corrisponde sempre all'effetto medio, o costante, e rappresenta l'insieme di pesi che determina la media delle altre variabili in base alla variabile specificata. In genere, questo contrasto è un vettore.

Le altre righe della matrice contengono i contrasti speciali che indicano i confronti desiderati tra le categorie della variabile. Solitamente, risultano più utili i contrasti ortogonali. I contrasti ortogonali sono statisticamente indipendenti e non ridondanti. I contrasti sono ortogonali se:

- Per ogni riga, la somma dei coefficienti di contrasto è pari a 0.
- La somma dei prodotti dei coefficienti corrispondenti relativi a tutte le coppie di righe non congiunte corrisponde a 0.

Si supponga, ad esempio, che il trattamento presenti quattro livelli e che si desideri confrontarli tra loro. Un contrasto speciale appropriato è

$( 1 \quad 1 \quad 1 \quad 1 )$	pesi per il calcolo della media
$( 3 \quad -1 \quad -1 \quad -1 )$	confronto tra il livello 1 e i livelli da 2 a 4
$( 0 \quad 2 \quad -1 \quad -1 )$	confronto tra il livello 2 e i livelli da 3 a 4
$( 0 \quad 0 \quad 1 \quad -1 )$	confronto tra il livello 3 e il livello 4

Il confronto viene specificato tramite il seguente sottocomando CONTRAST per MANOVA, LOGISTIC REGRESSION e COXREG:

```
/CONTRAST (TREATMNT)=SPECIAL ( 1 1 1 1
                                3 -1 -1 -1
                                0 2 -1 -1
                                0 0 1 -1 )
```

Per LOGLINEAR, è necessario specificare:

```
/CONTRAST (TREATMNT)=BASIS SPECIAL ( 1 1 1 1
                                       3 -1 -1 -1
                                       0 2 -1 -1
                                       0 0 1 -1 )
```

La somma di ogni riga ad eccezione della riga relativa alla media è pari a 0. Anche la somma dei prodotti di ogni coppia di righe non congiunte è pari a 0:

$$\text{Righe 2 e 3:} \quad (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0$$

$$\text{Righe 2 e 4:} \quad (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0$$

$$\text{Righe 3 e 4:} \quad (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0$$

Non è necessario che i contrasti speciali siano ortogonali. Tuttavia, non possono essere combinazioni lineari gli uni degli altri. In caso lo siano, la procedura segnala la dipendenza lineare e l'elaborazione viene interrotta. I contrasti di Helmert, di differenza e polinomiali sono contrasti ortogonali.

## ***Indicatore***

**Codifica della variabile dell'indicatore.** Denominata anche codifica fittizia, non è disponibile in LOGLINEAR o MANOVA. Il numero delle nuove variabili codificate è  $k-1$ . La codifica dei casi nella categoria di riferimento è pari a 0 per tutte le variabili  $k-1$ . La codifica di un caso nella categoria  $i$  è 0 per tutte le variabili dell'indicatore, ad eccezione della variabile  $i$  per la quale è 1.

## Strutture della covarianza

In questa sezione sono disponibili informazioni aggiuntive sulle strutture della covarianza.

**Ante dipendenza: primo ordine.** Questa struttura della covarianza ha varianze e correlazioni tra elementi adiacenti eterogenee. La correlazione tra due elementi non adiacenti è rappresentata dal prodotto delle correlazioni tra gli elementi compresi tra gli elementi desiderati.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_1 & \sigma_3\sigma_1\rho_1\rho_2 & \sigma_4\sigma_1\rho_1\rho_2\rho_3 \\ \sigma_2\sigma_1\rho_1 & \sigma_2^2 & \sigma_3\sigma_2\rho_2 & \sigma_4\sigma_2\rho_2\rho_3 \\ \sigma_3\sigma_1\rho_1\rho_2 & \sigma_3\sigma_2\rho_2 & \sigma_3^2 & \sigma_4\sigma_3\rho_3 \\ \sigma_4\sigma_1\rho_1\rho_2\rho_3 & \sigma_4\sigma_2\rho_2\rho_3 & \sigma_4\sigma_3\rho_3 & \sigma_4^2 \end{bmatrix}$$

**AR(1).** Si tratta di una struttura autoregressiva del primo ordine con varianze omogenee. La correlazione tra due elementi qualsiasi è uguale a  $\rho$  per elementi adiacenti,  $\rho^2$  per elementi separati da un terzo elemento e così via. A  $\rho$  viene applicato il vincolo  $-1 < \rho < 1$ .

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

**AR(1): eterogenea.** Si tratta di una struttura autoregressiva del primo ordine con varianze eterogenee. La correlazione tra due elementi qualsiasi è uguale a  $\rho$  per elementi adiacenti,  $\rho^2$  per due elementi separati da un terzo elemento e così via.  $\rho$  deve essere compreso tra  $-1$  e  $1$ .

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho^2 & \sigma_4\sigma_1\rho^3 \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho^2 \\ \sigma_3\sigma_1\rho^2 & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho^3 & \sigma_4\sigma_2\rho^2 & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

**ARMA(1,1).** Si tratta di una struttura a media mobile autoregressiva del primo ordine, che ha varianze omogenee. La correlazione tra due elementi è uguale a  $\phi^*\rho$  per gli elementi adiacenti, ad  $\phi^*(\rho^2)$  per gli elementi separati da un terzo elemento, e così via.  $\rho$  e  $\phi$  sono parametri della media autoregressiva e mobile, quindi i loro valori devono essere compresi tra  $-1$  e  $1$  compreso.

$$\sigma^2 \begin{bmatrix} 1 & \phi\rho & \phi\rho^2 & \phi\rho^3 \\ \phi\rho & 1 & \phi\rho & \phi\rho^2 \\ \phi\rho^2 & \phi\rho & 1 & \phi\rho \\ \phi\rho^3 & \phi\rho^2 & \phi\rho & 1 \end{bmatrix}$$

**Simmetria composta.** Questa struttura ha una varianza e una covarianza costanti.

$$\begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

**Simmetria composta: metrica di correlazione.** Questa struttura della covarianza ha varianze e correlazioni tra gli elementi omogenee.

$$\sigma^2 \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

**Simmetria composta: eterogenea.** Questa struttura della covarianza ha varianze eterogenee e correlazione tra elementi costante.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho & \sigma_3\sigma_1\rho & \sigma_4\sigma_1\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 & \sigma_3\sigma_2\rho & \sigma_4\sigma_2\rho \\ \sigma_3\sigma_1\rho & \sigma_3\sigma_2\rho & \sigma_3^2 & \sigma_4\sigma_3\rho \\ \sigma_4\sigma_1\rho & \sigma_4\sigma_2\rho & \sigma_4\sigma_3\rho & \sigma_4^2 \end{bmatrix}$$

**Diagonale.** Questa struttura della covarianza ha varianze eterogenee e nessuna correlazione tra elementi.

$$\begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{bmatrix}$$

**Analitica fattoriale: primo ordine.** Questa struttura della covarianza ha varianze omogenee composte da un termine eterogeneo e da un termine omogeneo negli elementi. La covarianza tra due elementi qualsiasi è la radice quadrata del prodotto dei relativi termini della varianza eterogenei.

$$\begin{bmatrix} \lambda_1^2 + d & \lambda_2\lambda_1 & \lambda_3\lambda_1 & \lambda_4\lambda_1 \\ \lambda_2\lambda_1 & \lambda_2^2 + d & \lambda_3\lambda_2 & \lambda_4\lambda_2 \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + d & \lambda_4\lambda_3 \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + d \end{bmatrix}$$

**Analitica fattoriale: primo ordine, eterogenea.** Questa struttura della covarianza ha varianze eterogenee composte da due termini eterogenei negli elementi. La covarianza tra due elementi qualsiasi è la radice quadrata del prodotto del primo dei relativi termini della varianza eterogenea.

$$\begin{bmatrix} \lambda_1^2 + d_1 & \lambda_2 \lambda_1 & \lambda_3 \lambda_1 & \lambda_4 \lambda_1 \\ \lambda_2 \lambda_1 & \lambda_2^2 + d_2 & \lambda_3 \lambda_2 & \lambda_4 \lambda_2 \\ \lambda_3 \lambda_1 & \lambda_3 \lambda_2 & \lambda_3^2 + d_3 & \lambda_4 \lambda_3 \\ \lambda_4 \lambda_1 & \lambda_4 \lambda_2 & \lambda_4 \lambda_3 & \lambda_4^2 + d_4 \end{bmatrix}$$

**Huynh-Feldt.** Si tratta di una matrice “circolare” in cui la covarianza tra due elementi qualsiasi è pari alla media delle varianze meno una delle costanti. In questo tipo di struttura non risultano costanti né le varianze né le correlazioni.

$$\begin{bmatrix} \sigma_1^2 & \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_2^2}{2} - \lambda & \sigma_2^2 & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_3^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_3^2}{2} - \lambda & \sigma_3^2 & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda \\ \frac{\sigma_1^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_2^2 + \sigma_4^2}{2} - \lambda & \frac{\sigma_3^2 + \sigma_4^2}{2} - \lambda & \sigma_4^2 \end{bmatrix}$$

**Identità scalata.** Questa struttura ha una varianza costante e si suppone che non ci siano correlazioni tra gli elementi.

$$\sigma^2 \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

**Toeplitz.** Questa struttura della covarianza ha varianze omogenee e correlazioni tra elementi eterogenee. La correlazione tra elementi adiacenti è omogenea nelle coppie di elementi adiacenti. La correlazione tra elementi separati da un terzo elemento è anch'essa omogenea e così via.

$$\sigma^2 \begin{bmatrix} 1 & \rho_1 & \rho_2 & \rho_3 \\ \rho_1 & 1 & \rho_1 & \rho_2 \\ \rho_2 & \rho_1 & 1 & \rho_1 \\ \rho_3 & \rho_2 & \rho_1 & 1 \end{bmatrix}$$

**Toeplitz: eterogenea.** Questa struttura della covarianza ha varianze e correlazioni tra gli elementi eterogenee. La correlazione tra elementi adiacenti è omogenea nelle coppie di elementi adiacenti. La correlazione tra elementi separati da un terzo elemento è anch'essa omogenea e così via.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_1 & \sigma_3 \sigma_1 \rho_2 & \sigma_4 \sigma_1 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_3 \sigma_2 \rho_1 & \sigma_4 \sigma_2 \rho_2 \\ \sigma_3 \sigma_1 \rho_2 & \sigma_3 \sigma_2 \rho_1 & \sigma_3^2 & \sigma_4 \sigma_3 \rho_1 \\ \sigma_4 \sigma_1 \rho_3 & \sigma_4 \sigma_2 \rho_2 & \sigma_4 \sigma_3 \rho_1 & \sigma_4^2 \end{bmatrix}$$

**Non strutturato.** Questo tipo corrisponde a una matrice di covarianza estremamente generica.

$$\begin{bmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{bmatrix}$$

**Non strutturata: metrica di correlazione.** Questa struttura della covarianza ha varianze e correlazioni eterogenee.

$$\begin{bmatrix} \sigma_1^2 & \sigma_2\sigma_1\rho_{21} & \sigma_3\sigma_1\rho_{31} & \sigma_4\sigma_1\rho_{41} \\ \sigma_2\sigma_1\rho_{21} & \sigma_2^2 & \sigma_3\sigma_2\rho_{32} & \sigma_4\sigma_2\rho_{42} \\ \sigma_3\sigma_1\rho_{31} & \sigma_3\sigma_2\rho_{32} & \sigma_3^2 & \sigma_4\sigma_3\rho_{43} \\ \sigma_4\sigma_1\rho_{41} & \sigma_4\sigma_2\rho_{42} & \sigma_4\sigma_3\rho_{43} & \sigma_4^2 \end{bmatrix}$$

**Componenti della varianza.** Assegna una struttura di identità scalata a ciascuno degli effetti casuali specificati.



# Notices

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

Patent No. 7,023,453

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

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

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

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

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

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

## COPYRIGHT LICENSE:

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

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

### **Trademarks**

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

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

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

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

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

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

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

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

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

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

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

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



- analisi della covarianza
  - in GLM - Multivariata, 2
- analisi della varianza
  - in componenti della varianza, 33
- analisi di sopravvivenza
  - in Kaplan-Meier, 141
  - in regressione di Cox, 146
  - in tavole di sopravvivenza, 136
  - Regressione di Cox dipendente dal tempo, 153
- analisi loglineare, 121
  - Analisi loglineare generale, 125
  - Analisi loglineare Logit, 130
- Analisi loglineare generale
  - contrasti, 125
  - covariate di cella, 125
  - criteri, 128
  - distribuzione dei conteggi di cella, 125
  - fattori, 125
  - funzioni aggiuntive del comando, 129
  - grafici, 128
  - impostazione di modelli, 127
  - intervalli di confidenza, 128
  - opzioni di visualizzazione, 128
  - residuo, 129
  - salvataggio di valori attesi, 129
  - salvataggio di variabili, 129
  - strutture di cella, 125
- Analisi loglineare Logit, 130
  - contrasti, 130
  - covariate di cella, 130
  - criteri, 133
  - distribuzione dei conteggi di cella, 130
  - fattori, 130
  - grafici, 133
  - impostazione di modelli, 132
  - intervalli di confidenza, 133
  - opzioni di visualizzazione, 133
  - residuo, 134
  - salvataggio di variabili, 134
  - strutture di cella, 130
  - valori attesi, 134
- ANOVA
  - in GLM - Multivariata, 2
  - in GLM a misure ripetute, 15
- ANOVA multivariata, 2
  
- Bonferroni
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- bontà di adattamento
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
  
- C di Dunnett
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- casi troncati
  - in Kaplan-Meier, 141
  - in regressione di Cox, 146
  - in tavole di sopravvivenza, 136
- categoria di riferimento
  - in Equazioni di stima generalizzate, 81, 83
  - in Modelli lineari generalizzati, 55
- classe generatrice
  - in selezione analisi loglineare, 123
- Componenti della varianza, 30
  - funzioni aggiuntive del comando, 35
  - modello, 32
  - opzioni, 33
  - salvataggio di risultati, 35
- contrasti
  - in analisi loglineare generale, 125
  - in analisi loglineare Logit, 130
  - in regressione di Cox, 148
- Convergenza hessiana
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
- convergenza parametri
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
  - in modelli misti lineari, 44
- convergenza verosimiglianza
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
  - in modelli misti lineari, 44
- costruzione di termini, 5, 20, 32, 123, 127, 133
- covariate
  - in regressione di Cox, 148
- covariate dipendenti dal tempo segmentate
  - in regressione di Cox, 153
- covariate stringa
  - in regressione di Cox, 148
- cronologia iterazioni
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
  - in modelli misti lineari, 44
  
- descrittive
  - in GLM - Multivariata, 12
  - in modelli misti lineari, 45
- deviazione standard
  - in GLM - Multivariata, 12

- in GLM a misure ripetute, 27
- differenza meno significativa (LSD)
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- differenza meno significativa (LSD) di Fisher
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- differenza significativa di Tukey
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- dimezzamenti
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
  - in modelli misti lineari, 44
- Distanza di Cook
  - in GLM, 11
  - in GLM a misure ripetute, 26
  - in Modelli lineari generalizzati, 68
- distribuzione binomiale
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione binomiale negativa
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione di Poisson
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione gamma
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione gaussiana inversa
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione multinomiale
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione normale
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- distribuzione Tweedie
  - in equazioni di stima generalizzate, 77
  - in modelli lineari generalizzati, 51
- Dunnett (T3)
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- effetti casuali
  - in modelli misti lineari, 42
- effetti fissi
  - in modelli misti lineari, 40
- eliminazione all'indietro
  - in selezione analisi loglineare, 121
- Equazioni di stima generalizzate, 72
  - categoria di riferimento per la risposta binaria, 81
  - criteri di stima, 86
  - esportazione del modello, 95
  - impostazione di modelli, 84
  - medie marginali stimate, 91
  - opzioni per i fattori categoriali, 83
  - predittori, 82
  - risposta, 80
  - salvataggio delle variabili nell'insieme di dati attivo, 93
  - statistiche, 89
  - tipo di modello, 76
  - valori iniziali, 88
- errore standard
  - in GLM, 11
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 26–27
- eta-quadrato
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- F* multiplo di Ryan-Einot-Gabriel-Welsch
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- fattori
  - in GLM a misure ripetute, 18
- frequenze
  - in selezione analisi loglineare, 124
- funzione di collegamento
  - modelli misti lineari generalizzati, 101
- funzione di collegamento binomiale negativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento Cauchit cumulativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento complemento log
  - in modelli lineari generalizzati, 52
- funzione di collegamento identità
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento log
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento log complementare
  - in equazioni di stima generalizzate, 78
- funzione di collegamento log-log complementare
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento log-log complementare cumulativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento log-log negativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento log-log negativa cumulativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento logit
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52

- funzione di collegamento logit cumulativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento potenza
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento potenza Odd
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento probit
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di collegamento probit cumulativa
  - in equazioni di stima generalizzate, 78
  - in modelli lineari generalizzati, 52
- funzione di sopravvivenza
  - in tavole di sopravvivenza, 136
- funzione funzionale generale
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
- GLM
  - salvataggio di matrici, 11
  - salvataggio di variabili, 11
- GLM - Misure ripetute, 15
  - definizione fattori, 18
  - diagnostici, 27
  - funzioni aggiuntive del comando, 29
  - grafici di profilo, 23
  - medie marginali stimate, 27
  - modello, 19
  - opzioni, 27
  - salvataggio di variabili, 26
  - test Post Hoc, 24
  - visualizzazione, 27
- GLM - Multivariata, 2, 14
  - covariate, 2
  - diagnostici, 12
  - fattori, 2
  - grafici di profilo, 8
  - medie marginali stimate, 12
  - opzioni, 12
  - test Post Hoc, 9
  - variabile dipendente, 2
  - visualizzazione, 12
- GLM multivariata, 2
- GLOR
  - in analisi loglineare generale, 125
- grafici
  - in analisi loglineare generale, 128
  - in analisi loglineare Logit, 133
- grafici dei residui
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- grafici di probabilità normale
  - in selezione analisi loglineare, 124
- grafici di profilo
  - in GLM - Multivariata, 8
  - in GLM a misure ripetute, 23
- grafici di variabilità vs. densità
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- Hochberg (GT2)
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- informazioni a livello di fattore
  - in modelli misti lineari, 45
- informazioni sul modello
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
- intervalli di confidenza
  - in analisi loglineare generale, 128
  - in analisi loglineare Logit, 133
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
  - in modelli misti lineari, 45
- intervallo multiplo di Ryan-Einot-Gabriel-Welsch
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- iterazioni
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
  - in selezione analisi loglineare, 124
- Kaplan-Meier, 141
  - confronto dei livelli del fattore, 143
  - definizione di eventi, 143
  - esempio, 141
  - funzioni aggiuntive del comando, 145
  - grafici, 145
  - quartili, 145
  - salvataggio di nuove variabili, 144
  - statistiche, 141, 145
  - tavole di sopravvivenza, 145
  - tempo di sopravvivenza media e mediana, 145
  - trend lineare per livelli di fattore, 143
  - variabili dello stato di sopravvivenza, 143
- legal notices, 165
- matrice dei coefficienti di contrasto
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
- matrice di correlazione
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
  - in modelli misti lineari, 45

- matrice di covarianza
  - in Equazioni di stima generalizzate, 86, 90
  - in GLM, 11
  - in Modelli lineari generalizzati, 60, 64
  - in modelli misti lineari, 45
- matrice di covarianza ad effetto casuale
  - in modelli misti lineari, 45
- matrice di covarianza del parametro
  - in modelli misti lineari, 45
- matrice di covarianza residua
  - in modelli misti lineari, 45
- matrice L
  - in Equazioni di stima generalizzate, 90
  - in Modelli lineari generalizzati, 64
- medie marginali stimate
  - in Equazioni di stima generalizzate, 91
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
  - in Modelli lineari generalizzati, 65
  - in modelli misti lineari, 46
- medie osservate
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- metodo Newton-Raphson
  - in analisi loglineare generale, 125
  - in analisi loglineare Logit, 130
- MINQUE
  - in componenti della varianza, 33
- modelli fattoriali completi
  - in componenti della varianza, 32
  - in GLM a misure ripetute, 19
- Modelli lineari generalizzati, 49
  - categoria di riferimento per la risposta binaria, 55
  - criteri di stima, 60
  - distribuzione, 49
  - esportazione del modello, 69
  - funzione di collegamento, 49
  - impostazione di modelli, 58
  - medie marginali stimate, 65
  - opzioni per i fattori categoriali, 57
  - predittori, 56
  - risposta, 54
  - salvataggio delle variabili nell'insieme di dati attivo, 67
  - statistiche, 63
  - tipi di modello, 49
  - valori iniziali, 62
- modelli logit multinomiali, 130
- modelli loglineari gerarchici, 121
- modelli misti
  - lineari, 36
- Modelli misti lineari, 36, 161
  - costruzione di termini, 40-41
  - criteri di stima, 44
  - effetti casuali, 42
  - effetti fissi, 40
  - funzioni aggiuntive del comando, 48
  - medie marginali stimate, 46
  - modello, 45
  - salvataggio di variabili, 47
  - struttura della covarianza, 161
  - termini di interazione, 40
- modelli misti lineari generalizzati, 98
  - annotazioni, 117-118
  - blocco effetti casuali, 108
  - distribuzione obiettivo, 101
  - effetti casuali, 107
  - effetti fissi, 104
  - esportazione del modello, 114
  - funzione di collegamento, 101
  - medie marginali stimate, 112
  - medie stimate, 119
  - offset, 110
  - peso analisi, 110
  - previsioni e osservazioni, 116
  - riepilogo del modello, 115
  - salva campi, 114
  - struttura dati, 116
  - tabella di classificazione, 116
  - termini personalizzati, 105
  - vista del modello, 115
- modelli personalizzati
  - in componenti della varianza, 32
  - in GLM a misure ripetute, 19
  - in selezione analisi loglineare, 123
- modelli saturi
  - in selezione analisi loglineare, 123
- modello di proporzionalità del rischio
  - in regressione di Cox, 146
- Newman-Keuls
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- parametro scala
  - in Equazioni di stima generalizzate, 86
  - in Modelli lineari generalizzati, 60
- pesi degli effetti casuali
  - in componenti della varianza, 33
- prodotti incrociati
  - matrici di ipotesi ed errore, 12
- punteggio
  - in modelli misti lineari, 44
- Punteggio Fisher
  - in modelli misti lineari, 44
- R-E-G-W  $F$ 
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- R-E-G-W  $Q$ 
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24

- rapporto logaritmico generalizzato  
in analisi loglineare generale, 125
- Rapporto Odd  
in analisi loglineare generale, 125
- regressione con multivariate, 2
- Regressione di Cox, 146  
contrasti, 148  
covariate, 146  
covariate categoriche, 148  
covariate dipendenti dal tempo, 153–154  
covariate stringa, 148  
definizione evento, 152  
DiffBeta, 150  
esempio, 146  
funzione di rischio, 150  
funzione di sopravvivenza, 150  
funzioni aggiuntive del comando, 152  
funzioni di base, 151  
grafici, 149  
inserimento e rimozione per passi, 151  
iterazioni, 151  
residui parziali, 150  
salvataggio di nuove variabili, 150  
statistiche, 146, 151  
variabile dello stato di sopravvivenza, 152
- Regressione di Poisson  
in analisi loglineare generale, 125
- residui  
in Equazioni di stima generalizzate, 94  
in Modelli lineari generalizzati, 68
- residui cancellati  
in GLM, 11  
in GLM a misure ripetute, 26
- residui di devianza  
in Modelli lineari generalizzati, 68
- residui di Pearson  
in Equazioni di stima generalizzate, 94  
in Modelli lineari generalizzati, 68
- residui di verosimiglianza  
in Modelli lineari generalizzati, 68
- residui non standardizzati  
in GLM, 11  
in GLM a misure ripetute, 26
- residui standardizzati  
in GLM, 11  
in GLM a misure ripetute, 26
- residuo  
in analisi loglineare generale, 129  
in analisi loglineare Logit, 134  
in modelli misti lineari, 47  
in selezione analisi loglineare, 124
- riassunto dell'elaborazione casi  
in Equazioni di stima generalizzate, 90  
in Modelli lineari generalizzati, 64
- scomposizione gerarchica, 5, 20  
in componenti della varianza, 34
- Selezione Analisi Loglineare, 121  
definizione di intervalli di fattori, 122  
funzioni aggiuntive del comando, 124  
modelli, 123  
opzioni, 124
- separazione  
in Equazioni di stima generalizzate, 86  
in Modelli lineari generalizzati, 60
- somma dei quadrati, 5, 20  
in componenti della varianza, 34  
in modelli misti lineari, 41  
matrici di ipotesi ed errore, 12
- SSCP  
in GLM - Multivariata, 12  
in GLM a misure ripetute, 27
- SSCP dei residui  
in GLM - Multivariata, 12  
in GLM a misure ripetute, 27
- statistica di Wald  
in analisi loglineare generale, 125  
in analisi loglineare Logit, 130
- statistiche descrittive  
in Equazioni di stima generalizzate, 90  
in GLM a misure ripetute, 27  
in Modelli lineari generalizzati, 64
- stima massima verosimiglianza  
in componenti della varianza, 33
- stima massima verosimiglianza ristretta  
in componenti della varianza, 33
- stime dei parametri  
in analisi loglineare generale, 125  
in analisi loglineare Logit, 130  
in Equazioni di stima generalizzate, 90  
in GLM - Multivariata, 12  
in GLM a misure ripetute, 27  
in Modelli lineari generalizzati, 64  
in modelli misti lineari, 45  
in selezione analisi loglineare, 124
- stime effetto-dimensioni  
in GLM - Multivariata, 12  
in GLM a misure ripetute, 27
- stime potenza  
in GLM - Multivariata, 12  
in GLM a misure ripetute, 27
- strutture della covarianza, 161  
in modelli misti lineari, 161
- Student-Newman-Keuls  
in GLM - Multivariata, 9  
in GLM a misure ripetute, 24
- Tamhane (T2)  
in GLM - Multivariata, 9  
in GLM a misure ripetute, 24
- tasso di rischio  
in tavole di sopravvivenza, 136
- tavole di contingenza  
in analisi loglineare generale, 125

- in selezione analisi loglineare, 121
- Tavole di sopravvivenza, 136
  - confronto dei livelli del fattore, 139
  - disattivazione della visualizzazione delle tavole, 139
  - esempio, 136
  - funzione di sopravvivenza, 136
  - funzioni aggiuntive del comando, 139
  - grafici, 139
  - statistiche, 136
  - tasso di rischio, 136
  - test di Wilcoxon (Gehan), 139
  - variabili dello stato di sopravvivenza, 138
  - variabili fattore, 138
- termini di interazione, 5, 20, 32, 123, 127, 133
  - in modelli misti lineari, 40
- termini nidificati
  - in Equazioni di stima generalizzate, 84
  - in Modelli lineari generalizzati, 58
  - in modelli misti lineari, 41
- test a intervallo multiplo di Duncan
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test del moltiplicatore di Lagrange
  - in Modelli lineari generalizzati, 64
- test del rango logaritmico
  - in Kaplan-Meier, 143
- test di Breslow
  - in Kaplan-Meier, 143
- Test *di* Dunnett
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test di Gehan
  - in tavole di sopravvivenza, 139
- test di Levene
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- test di omogeneità della varianza
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- test di Scheffé
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test di sfericità di Bartlett
  - in GLM - Multivariata, 12
- test di sfericità di Mauchly
  - in GLM a misure ripetute, 27
- Test di Sidak
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test di Tarone-Ware
  - in Kaplan-Meier, 143
- Test *di* Tukey
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- Test *di* Waller-Duncan
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test di Wilcoxon
  - in tavole di sopravvivenza, 139
- test M di Box
  - in GLM - Multivariata, 12
- Test per confronti a coppie di Gabriel
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- Test per confronti a coppie di Games e Howell
  - in GLM - Multivariata, 9
  - in GLM a misure ripetute, 24
- test per parametri di covarianza
  - in modelli misti lineari, 45
- test t
  - in GLM - Multivariata, 12
  - in GLM a misure ripetute, 27
- tolleranza singolarità
  - in modelli misti lineari, 44
- trademarks, 166
- valori attesi
  - in analisi loglineare generale, 129
  - in analisi loglineare Logit, 134
  - in modelli misti lineari, 47
- valori attesi fissi
  - in modelli misti lineari, 47
- valori attesi ponderati
  - in GLM, 11
  - in GLM a misure ripetute, 26
- valori d'influenza
  - in GLM, 11
  - in GLM a misure ripetute, 26
  - in Modelli lineari generalizzati, 68
- variabili di misure ripetute
  - in modelli misti lineari, 38
- variabili di soggetti
  - in modelli misti lineari, 38
- vista del modello
  - nei modelli misti lineari generalizzati, 115