

*IBM SPSS Advanced Statistics 32*

**IBM**

**Nota**

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in [“Informazioni particolari” a pagina 141](#).

**Informazioni sul prodotto**

Questa edizione si applica alla versione 32, release 0, modifica 0 di IBM® SPSS Statistica e a tutte le release e modifiche successive se non diversamente indicato in nuove edizioni.

© **Copyright International Business Machines Corporation .**

# Indice

<b>Capitolo 1. Statistiche avanzate.....</b>	<b>1</b>
Introduzione a statistiche avanzate.....	1
Analisi Multivariata GLM.....	1
Modello Multivariate GLM.....	3
Contrasti Multivariati GLM.....	4
Grafici profilo Multivariate GLM.....	4
GLM Multivariate Post Hoc.....	5
Medie marginali stimate GLM.....	6
GLM – Univariato: Salva.....	6
Opzioni Multivariate GLM.....	7
Funzioni Aggiuntive Comando GLM.....	9
GLM Misure ripetute.....	9
GLM Misure ripetute definiscono i fattori.....	11
Modello Misure ripetute GLM.....	12
GLM Misure Ripetute Contrade.....	13
GLM Misure ripetute Profilo profilo.....	14
GLM Misure Ripetute Post Hoc.....	14
Medie marginali stimate GLM.....	15
GLM Misure Ripetute Salvataggio.....	16
Opzioni Misure ripetute GLM.....	16
Funzioni Aggiuntive Comando GLM.....	17
Analisi Componenti della varianza.....	18
Modello Componenti Variance.....	19
Opzioni Componenti di varianza.....	19
Varianza Componenti Salva su Nuovo file.....	21
VARCOMP Command Aggiuntive Features.....	21
Modelli misti lineari.....	21
Modelli Misti Lineari: Soggetti e Repeati.....	22
Modelli misti lineari a effetti fissi.....	24
Modelli misti lineari a effetti casuali.....	25
Stima dei modelli misti lineari.....	26
Statistiche Modelli Misti Lineari.....	27
Linear Mixed Models EM Mezzo.....	27
Modelli misti lineari Salva.....	28
Modelli misti lineari - Esporta.....	28
MIXED Comando Funzioni aggiuntive.....	28
Modelli lineari generalizzati.....	28
Risposta Modelli Lineari Generalizzati.....	31
Predicatori Modelli Lineari Generalizzati.....	32
Modello Modelli Lineari Generalizzati.....	33
Stima Modelli Lineari Generalizzati.....	33
Statistiche Modelli Lineari Generalizzati.....	35
Modelli Lineari Generalizzati EM Mezzi.....	36
Modelli lineari generalizzati Salva.....	37
Esportazione Modelli Lineari Generalizzati.....	38
Funzioni Aggiuntive di comando GENLIN.....	39
Equazioni di stima generalizzate.....	39
Equazioni di stima generalizzate Tipo di modello.....	41
Stima delle Equazioni generalizzate.....	43
Predicatori Di Equazioni Stimate Generalizzati.....	43
Modello di Equazioni di stima generalizzata.....	44

Stima Delle Equazioni Di Stima Generalizzata.....	45
Statistiche delle Equazioni di stima generalizzate.....	46
Equazioni di stima generalizzate EM Means.....	47
Equazioni di stima generalizzate Salva.....	48
Stima delle Equazioni generalizzate Export.....	49
Funzioni Aggiuntive di comando GENLIN.....	50
Modelli misti lineari generalizzati.....	50
Ottenimento di un modello misto lineare generalizzato.....	52
Obiettivo .....	53
Effetti fissi .....	55
Effetti casuali .....	56
Peso e offset .....	57
Opzioni generali di creazione .....	58
Stima .....	58
Medie stimate .....	59
Salva .....	60
Esporta .....	60
Vista modello .....	60
Selezione modello: Analisi loglineare.....	64
Analisi loglineare Definisci intervallo.....	65
Modello di analisi loglineare.....	65
Opzioni di analisi Loglineare di selezione modello.....	66
HILOGLINEAR Comando Funzioni Aggiuntive.....	66
Analisi loglineare generale.....	66
Modello di analisi Loglineare generale.....	68
Opzioni generali di analisi Loglineare.....	68
Analisi generale di analisi Loglineare.....	69
Funzioni Aggiuntive Di Comando GENLOG.....	69
Analisi loglineare logit.....	69
Modello di analisi Loglineare logit.....	70
Opzioni di analisi Loglineare logit.....	71
Analisi Logit Loglineare Salva.....	72
Funzioni Aggiuntive Di Comando GENLOG.....	72
Tabelle vita.....	72
Tabelle di vita Definisci intervallo.....	74
Opzioni tabelle di vita.....	74
SURVIVAL Comando Funzioni Aggiuntive.....	74
Modelli Temporal Accelerati Parametrici accelerati.....	74
Modelli parametrici di tempo di guasto accelerato: Criteri.....	75
Parametric Accelerated Failure Time Models: Modello.....	76
Parametric Accelerated Failure Time Models: Preventivo.....	76
Parametric Accelerated Failure Time Models: Stampa.....	78
Parametric Accelerated Failure Time Models: Prevedere.....	78
Modelli parametrici di tempo di guasto accelerato: Tracciato.....	79
Modelli parametrici di tempo di guasto accelerato: Esportazione.....	80
Eventi di definizione AFT di sopravvivenza per variabili di stato.....	80
Modelli di tempo di errore accelerati parametrici: Seleziona categoria.....	80
Modelli di fragilità condivisa parametrici.....	81
Modelli di fragilità condivisa parametrica: criteri.....	82
Modelli di fragilità condivisa parametrica: Modello.....	83
Modelli di fragilità condivisa parametrica: Stima.....	83
Modelli di fragilità condivisa parametrica: Stampa.....	85
Modelli di fragilità condivisa parametrici: Prevedi.....	85
Modelli di fragilità parametrici condivisi: grafico.....	86
Modelli di fragilità condivisa parametrici: Esporta.....	87
Modelli di fragilità parametrica condivisa: Definisci eventi.....	87
Modelli di fragilità condivisa parametrici - Esempi.....	88
Modelli di fragilità parametrici condivisi - Un caso di studio per i dati più di recente.....	90

Kaplan - Meier Survival Analysis.....	92
Kaplan - Meier Definisci evento per la variabile di stato.....	93
Kaplan - Meier Confronta Factor Livelli.....	93
Kaplan - Meier Save New Variables.....	94
Kaplan - Opzioni Meier.....	94
Funzioni Aggiuntive di comando KM.....	94
Cox Analisi di regressione.....	94
Cox Regressione Definire le variabili categoriali.....	95
Cox Grafici di regressione.....	96
Cox Regressione Salva nuove variabili.....	96
Cox Opzioni di regressione.....	97
Cox Regressione Definire l'evento per la variabile di stato.....	97
Funzioni Aggiuntive Comando COXREG.....	97
Tempo di calcolo - Covariati dipendenti.....	98
Computing a tempo - Patto dipendente.....	98
Schemi Di Codifica Variabile Categoriale.....	99
Deviazione.....	99
Semplice.....	99
Helmert.....	100
Differenza.....	100
Polinomiale.....	100
Ripetuto.....	101
Speciale.....	101
Indicatore.....	102
Strutture covarianza.....	102
Statistiche bayesiane.....	106
Inferenza a un campione bayesiana: Normale.....	107
Inferenza a un campione bayesiana: Binomiale.....	111
Inferenza a un campione bayesiana: Poisson.....	113
Inferenza campione correlato bayesiana: Normale.....	115
Bayesiano Independent - Inferenza di esempio.....	117
Inferenza Bayessiana su correlazione di Pearson.....	121
Inferenza Bayessiana su modelli di regressione lineari.....	123
ANOVA a una via bayesiana.....	128
Modelli Loglinear Bayesiani.....	131
Bayesi One-way Ripetuto Misure ANOVA Models.....	134
KRR (Kernel Ridge Regression).....	136
Parametri kernel.....	138
Regressione Ridge kernel: opzioni.....	138
<b>Informazioni particolari.....</b>	<b>141</b>
Marchi.....	142

<b>Indice analitico.....</b>	<b>145</b>
------------------------------	------------



---

# Capitolo 1. Statistiche avanzate

Le seguenti funzioni statistiche avanzate sono incluse in SPSS Statistiche Standard Edition o l'opzione Advanced Statistics.

---

## Introduzione a statistiche avanzate

L'opzione Advanced Statistics include procedure che offrono opzioni di modellazione più avanzate di quelle disponibili tramite l'opzione Statistiche Base.

- GLM Multivariate estende il modello lineare generale fornito da GLM Univariate per consentire molteplici variabili dipendenti. Un'ulteriore estensione, GLM Repeated Measures, consente misurazioni ripetute di più variabili dipendenti.
- Variance Components Analysis è uno strumento specifico per decomporre la variabilità in una variabile dipendente in componenti fissi e casuali.
- Linear Mixed Models espande il modello lineare generale in modo che i dati siano consentiti per esibire la variabilità correlata e non costante. Il modello lineare misto, quindi, fornisce la flessibilità di modellazione non solo dei mezzi dei dati ma delle varianze e delle covarianze.
- Generalized Linear Models (GLM) rilassa l'assunzione di normalità per il termine di errore e richiede solo che la variabile dipendente sia linearmente correlata ai predittori attraverso una funzione di trasformazione o di collegamento. Le Equazioni di stima generalizzate (GEE) ampliano GLM per consentire misurazioni ripetute.
- General Loglinear Analysis consente di adattare i modelli per i dati di conteggio incrociati e Model Selection Loglinear Analysis può aiutarti a scegliere tra i modelli.
- Logit Loglinear Analysis consente di adattare modelli loglineari per l'analisi del rapporto tra un dipendente categoriale e uno o più predittori categoriali.
- L'analisi di sopravvivenza è disponibile attraverso le Tabelle di vita per esaminare la distribuzione delle variabili tempo-evento, eventualmente in base ai livelli di una variabile fattore; l'Analisi di sopravvivenza di Kaplan-Meier per esaminare la distribuzione delle variabili tempo-evento, eventualmente in base ai livelli di una variabile fattore o producendo analisi separate in base ai livelli di una variabile di stratificazione; e Cox Regression per modellare il tempo a un evento specifico, in base ai valori di determinate covariate.

---

## Analisi Multivariata GLM

La procedura GLM Multivariate fornisce analisi di regressione e analisi della varianza per variabili dipendenti multiple da una o più variabili fattore o covariate. Le variabili fattore suddividono la popolazione in gruppi. Utilizzando questa procedura di modello lineare generale, è possibile testare ipotesi nulle sugli effetti delle variabili del fattore sui mezzi di vari raggruppamenti di una distribuzione congiunta di variabili dipendenti. È possibile indagare le interazioni tra fattori oltre che 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, le somme di quadrati dovute agli effetti nel modello e le somme di errore delle piazze sono in forma matrice piuttosto che la forma scalare trovata in analisi univariata. Queste matrici sono chiamate SSCP (somme - di - quadrati e prodotti incrociati) matrici. Se si specifica più di una variabile dipendente, viene fornita l'analisi multivariata della varianza che utilizza la traccia di Pillai, la lambda di Wilks, la traccia di Hotelling e il principale criterio di root di Roy con statistica approssimata  $F$ , nonché l'analisi univariata della varianza per ogni variabile dipendente. Oltre a testare ipotesi, GLM Multivariate produce stime di parametri.

I contrasti di *a priori* sono disponibili per eseguire test di ipotesi. 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. I mezzi marginali stimati danno stime dei valori medi previsti per le celle nel modello, e i tracciati di profili (tracciati di interazione) di questi mezzi consentono di visualizzare facilmente alcune relazioni. I test di confronto multipli post hoc vengono eseguiti per ogni variabile dipendente separatamente.

Residui, valori attesi, distanza di Cook e valori di leva possono essere salvati come variabili nel file di dati per la verifica di ipotesi. Disponibile anche una matrice SSCP residua, ovvero una matrice quadrata di somme di quadrati e cross - products di residui, una matrice di covarianza residua, ovvero la matrice SSCP residua divisa dai gradi di libertà dei residui, e la matrice di correlazione residua, ovvero la forma standardizzata della matrice di covarianza residua.

WLS Weight consente di specificare una variabile usata per dare osservazioni differenti pesi per un'analisi WLS (ponderata), magari per compensare una diversa precisione di misurazione.

**Esempio.** Un produttore di plastica misura tre proprietà del film di plastica: resistenza a strappo, gloss e opacità. Si provano due tassi di estrusione e due diverse quantità di additivo e le tre proprietà sono misurate sotto ogni combinazione di tasso di estrusione e di additivo. Il produttore rileva che il tasso di estrusione e la quantità di additivo producono singolarmente risultati significativi ma che l'interazione dei due fattori non è significativa.

**Metodi.** Le somme di quadrati di tipo I, Type II, Type III e Type IV possono essere utilizzate per valutare diverse ipotesi. Il metodo predefinito è il Tipo III.

**Statistiche.** Test di gamma post hoc e confronti multipli: differenza meno significativa, Bonferroni, Sidak, Scheffé, Ryan - Einot - Gabriel - Welsch multiplo  $F$ , Ryan - Einot - Gabriel - Welsch multipli, Student - Newman - Keuls, Tukey's onestamente significative, Tukey's  $b$ , Duncan, Hochberg's GT2, Gabriel, Waller Duncan  $t$  test, Dunnett (uno - sided e due - schierati), Tamhane's T2, Dunnett's T3, Games - Howell, and Dunnett's C. Statistiche descrittive: medie osservate, deviazioni standard e conta per tutte le variabili dipendenti in tutte le celle; il test di Levene per l'omogeneità della varianza; il test di Box  $M$  dell'omogeneità delle matrici di covarianza delle variabili dipendenti; e la prova di sfericità di Bartlett.

**Grafici.** Spread - versus - livello, residuo e profilo (interazione).

Considerazioni Dati Multivariate GLM

**Dati.** Le variabili dipendenti dovrebbero essere quantitative. I fattori sono categoriali e possono avere valori numerici o valori di stringa. Le covariate sono variabili quantitative correlate alla variabile dipendente.

**Ipotesi.** Per le variabili dipendenti, i dati sono un campione casuale di vettori provenienti da una popolazione normale multivariata; nella popolazione le matrici di varianza - covarianza per tutte le cellule sono le stesse. L'analisi della varianza è uno stimatore robusto degli scostamenti dalla normalità, anche se i dati devono essere simmetrici. Per verificare le ipotesi, è possibile utilizzare l'omogeneità delle prove delle varianze (inclusi i grafici di Box  $M$ ) e i tracciati a livello di diffusione. È inoltre possibile esaminare residui e grafici dei residui.

**Procedure correlate.** Utilizzare la procedura Explore per esaminare i dati prima di effettuare un'analisi della varianza. Per una singola variabile dipendente, utilizzare GLM Univariate. Se si misurano le stesse variabili dipendenti in diverse occasioni per ogni soggetto, utilizzare GLM Repeated Misure.

Ottenimento di Tavole Multivariate GLM

1. Dai menu, scegliere:

**Analizzare > Modello Lineare Generale > Multivariate ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare almeno due variabili dipendenti.

Facoltativamente, è possibile specificare Fixed Factor (s), Covariate (s) e WLS Peso.

## Modello Multivariate GLM

**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 dei 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 della somma dei quadrati in genere utilizzato con modelli bilanciati o non bilanciati privi di celle mancanti è il Tipo III.

**Includi 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.

## Creazione di termini e termini personalizzati

### Crea termini

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### Crea termini personalizzati

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## 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 noto anche come decomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene adattato 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 adattato 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 della 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 calcola le somme dei quadrati di un effetto nel disegno come la somma dei quadrati adattata per qualsiasi altro effetto che non lo contiene 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 pesate di Yates. Il metodo della 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  $F$  nel 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 della 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.

## Contrasti Multivariati GLM

I contrasti vengono utilizzati per verificare se i livelli di un effetto sono significativamente diversi l'uno dall'altro. È possibile specificare un contrasto per ogni fattore del modello. I contrasti rappresentano combinazioni lineari dei parametri.

L'ipotesi test si basa sull'ipotesi nulla  $\mathbf{LBM} = \mathbf{0}$ , dove  $\mathbf{L}$  è la matrice dei coefficienti di contrasto,  $\mathbf{M}$  è la matrice di identità (che ha dimensione uguale al numero di variabili dipendenti) e  $\mathbf{B}$  è il vettore di parametri. Quando viene specificato un contrasto, viene creata una matrice  $\mathbf{L}$  tale che le colonne corrispondenti al fattore corrispondono al contrasto. Le altre colonne vengono adattate in modo che la matrice  $\mathbf{L}$  possa essere stimata.

Oltre al test univariato utilizzando le statistiche  $F$  e gli intervalli di confidenza simultanei di tipo Bonferroni basati sulla distribuzione di Student's  $t$  per le differenze di contrasto in tutte le variabili dipendenti, vengono forniti i test multivariati che utilizzano la traccia di Pillai, la lambda di Wilks, la traccia di Hotelling e i principali criteri di root 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 dei fattori 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 delle tendenze polinomiali.

## Grafici profilo Multivariate GLM

I grafici di profilo, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un grafico di profilo è un grafico a linee in cui ciascun punto indica la media marginale stimata

di una variabile dipendente (adattata 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 tracciati. I tracciati del profilo vengono creati per ogni variabile dipendente.

Il grafico di 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.

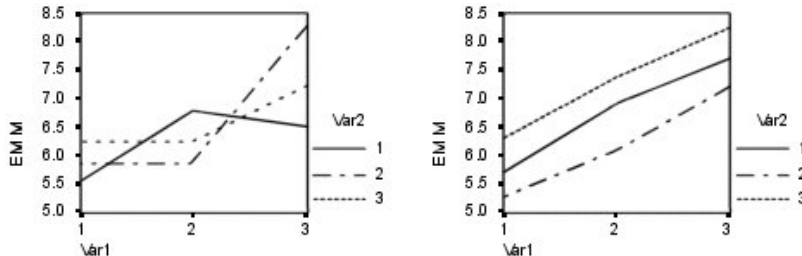


Figura 1. Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)

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 Multivariate Post Hoc

**Test di confronti multipli post hoc.** Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervallo e i confronti pairwise multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stato apportato alcun adattamento. I test post hoc vengono eseguiti per ogni variabile dipendente separatamente.

I test di confronto multiplo usati più di frequente sono il test di Bonferroni e il test HSD di Tukey. Il **test di Bonferroni**, basato sulla statistica  $t$  di Student, consente di adattare il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test  $t$  di Sidak** adatta inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test HSD di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti pairwise tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti pairwise. Quando si eseguono test su un elevato numero di coppie di medie, il test HSD 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 HSD di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test per confronti pairwise 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 pairwise 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 dei fattori (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 a qualsiasi livello del fattore è inferiore a quella della categoria di controllo, selezionare **< Control**. Allo stesso modo, per verificare se la media a qualsiasi livello del fattore è superiore a quella della categoria di controllo, selezionare **> Control**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test di intervallo 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 di 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, utilizzare **T2 di Tamhane** (test per confronti pairwise conservativo basato su un test  $t$ ), **T3 di Dunnett** (test per confronti pairwise basato sui moduli massimi studentizzati), **il test per confronti pairwise di Games-Howell** (a volte liberale) o **C di Dunnett** (test per confronti pairwise basato sull'intervallo studentizzato).

**Il test di intervallo multiplo di Duncan**, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test di 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 di intervallo che usa la media armonica della dimensione del campione nel caso di dimensioni del campione non uguali.

Il livello di significatività del test di **Scheffé** è progettato per consentire il test di tutte le possibili combinazioni lineari delle medie di gruppo, non solo dei confronti pairwise disponibili in questa funzione. Ne risulta che il test di Scheffé è spesso più conservativo di altri test, ciò significa che è richiesta una maggiore differenza tra le medie per la significatività.

Il test di confronto multiplo pairwise 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 adattamento del livello di significatività osservata per confronti multipli.

**Test visualizzati.** I confronti pairwise sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), C di Dunnett e T3 di Dunnett. 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 di intervallo. Il test Differenza significativa di Tukey, GT2 di Hochberg, il test di Gabriel e il test di Scheffé sono sia test di confronto multiplo che test di intervalli.

## Medie marginali stimate GLM

Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono adattate per le eventuali covariate.

### Confronta effetti principali

Consente di eseguire confronti pairwise 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.

### Confronta effetti principali semplici

L'impostazione è abilitata ogni qualvolta l'elenco di destinazione contiene uno o più effetti di prodotto o interazione (ad esempio  $A*B$ ,  $A*B*C$ ). L'impostazione supporta la specifica dei confronti tra semplici effetti principali, che sono effetti principali nidificati all'interno dei livelli di altri fattori.

### Adattamento intervallo di confidenza

Selezionare la differenza meno significativa (LSD), l'adattamento di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo articolo è disponibile solo quando viene selezionato **Confronta gli effetti principali** e / o **Confronta semplici effetti principali**.

## Specifiche delle medie marginali stimate

1. Dai menu scegliere una delle procedure disponibili sotto > **Analizza** > **Modello lineare generale**.
2. Nella finestra di dialogo principale, fare clic su **Medie EM**.

## GLM – Univariato: 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 Statistica, è necessario salvare il file di dati corrente.

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

- *Non standardizzati*. Il valore che il modello prevede per la variabile dipendente.
- *Pesato*. Valori attesi ponderati non standardizzati. Disponibile solo se era stata precedentemente 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.

**Diagnostiche.** 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. Un valore alto del D di Cook indica che l'esclusione di un caso dal calcolo delle statistiche di regressione modifica sostanzialmente i coefficienti.
- *Valori di leva*. Valori di leva non centrati. L'influenza relativa di ogni osservazione sull'adattamento del modello.

**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.
- *Pesato*. Residui ponderati non standardizzati. Disponibile solo se era stata precedentemente selezionata una variabile WLS.
- *Standardizzati*. Il residuo diviso per una stima della sua deviazione standard. I residui standardizzati, noti anche come residui di Pearson, hanno una media pari a 0 e una deviazione standard pari a 1.
- *Studentizzati*. Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza dei valori di ogni caso sulle variabili indipendenti dai mezzi delle variabili indipendenti. A volte indicato come residui internamente studentizzati.
- *Eliminato*. Il residuo per un caso quando tale caso è escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore previsto adattato.

**Statistiche dei coefficienti.** Scrive una matrice della varianza-covarianza delle stime del parametro nel modello in un nuovo dataset della sessione attiva o in un file di dati IBM SPSS Statistica esterno. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime del parametro, una riga di errori standard delle stime del parametro, una riga di valori di significatività per le statistiche *t* corrispondenti alle stime del parametro e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. Quando è selezionato Statistiche congruenti con l'eteroschedasticità (disponibile solo per modelli univariata), la matrice varianza-covarianza è calcolata utilizzando uno stimatore robusto, la riga di errori standard visualizza gli errori standard robusti e i valori di significatività riflettono gli errori robusti. Il file matrice può essere usato in altre procedure che leggono i file matrice.

## Opzioni Multivariate GLM

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

**Visualizza.** Selezionare **Statistiche descrittive** per produrre mezzi osservati, deviazioni standard e conta per tutte le variabili dipendenti in tutte le celle. La funzione **Stima della dimensione degli effetti** fornisce un valore eta al quadrato parziale per ciascun effetto e per ciascuna stima del parametro. La statistica eta al 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 del parametro** per ottenere stime del parametro, errori standard, test *t*, intervalli di confidenza e la potenza osservata per ciascun test. È possibile visualizzare l'ipotesi e l'errore **matrici SSCP** e la matrice **Residual SSCP** più la prova di sfericità della matrice di covarianza residua.

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 tra soggetti, solo per i fattori tra soggetti.

Inoltre, i test di omogeneità includono la prova di Box *M* dell'omogeneità delle matrici di covarianza delle variabili dipendenti in tutte le combinazioni di livello dei fattori tra i soggetti. Le opzioni Grafici di diffusione 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 **Residual grafici** per produrre una trama dei residui osservati per ogni variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Selezionare **Lack of fit test** per verificare se il rapporto tra la variabile dipendente e le variabili indipendenti può essere adeguatamente descritto dal modello. **Funzione Generale stimabile (s)** consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

## Visualizza

### Statistiche descrittive

Produce mezzi osservati, deviazioni standard e conta per tutte le variabili dipendenti in tutte le celle.

### Stime della dimensione degli effetti

Dà un valore di tipo beta parziale per ogni effetto e ogni stima di parametro. La statistica eta al quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore.

### Potenza osservata

Ottiene la potenza del test quando l'ipotesi alternativa è impostata in base al valore osservato.

### Stime dei parametri

Produce le stime dei parametri, gli errori standard, i test *t*, gli intervalli di confidenza e la potenza osservata per ogni test.

### Matrici SSCP

Visualizza le matrici SSCP di ipotesi e di errore.

### Matrice SSCP dei residui

Visualizza la matrice SSCP residua di ipotesi e di errore.

### Matrice di trasformazione

Visualizza la prova di sfericità di Bartlett della matrice di covarianza residua.

### Test di omogeneità

Produce il test di Levene dell'omogeneità della varianza per ogni variabile dipendente in tutte le combinazioni di livello dei fattori tra soggetti, per soli fattori tra soggetti. Inoltre, i test di omogeneità includono la prova di Box *M* dell'omogeneità delle matrici di covarianza delle variabili dipendenti in tutte le combinazioni di livello dei fattori tra i soggetti.

### Grafico di confronto tra diffusione e livello

Utile per verificare le ipotesi sui dati per indagare l'assunzione di uguale varianza. Se non è disponibile alcun fattore, questa opzione risulta disattivata.

### Grafico dei residui

Produce una trama dei residui osservati per ogni variabile dipendente. La trama è utile per indagare l'assunzione di uguale varianza.

### Mancanza di adattamento

Verificare se il rapporto tra la variabile dipendente e le variabili indipendenti può essere adeguatamente descritto dal modello.

### Funzioni generali stimabili

Consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

### Livello di significatività

Potrebbe risultare utile adattare 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, il livello associato degli intervalli di confidenza viene visualizzato nella finestra di dialogo.

## Funzioni Aggiuntive Comando GLM

Queste caratteristiche possono applicarsi ad analisi di misure univariate, multivariate o ripetute. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati nel 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 personalizzata **L**, matrice **M** o matrice **K** (utilizzando i sottocomandi 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 nell'elenco dei fattori (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati matrice di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati matrice 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).

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

## GLM Misure ripetute

---

La Procedura GLM Ripetute Misure fornisce l'analisi della varianza quando la stessa misurazione viene effettuata più volte su ogni soggetto o caso. Se i fattori tra - soggetti sono specificati, dividono la popolazione in gruppi. Utilizzando questa procedura di modello lineare generale, è possibile testare ipotesi nulle sugli effetti sia dei fattori tra soggetti che dei fattori all'interno dei soggetti. È possibile indagare le interazioni tra fattori oltre che gli effetti dei singoli fattori. Inoltre, possono essere inclusi gli effetti delle covariate costanti e delle interazioni covariate con i fattori tra i soggetti.

In un design di misure ripetute doppiamente multivariate, le variabili dipendenti rappresentano misurazioni di più di una variabile per i diversi livelli dei fattori all'interno dei soggetti. Ad esempio, si poteva misurare sia il polso che la respirazione in tre momenti diversi su ogni soggetto.

La Procedura GLM Ripetute Misure fornisce analisi sia univariate che multivariate per i dati delle 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, le somme di quadrati dovute agli effetti nel modello e le somme di errore delle piazze sono in forma matrice piuttosto che la forma scalare trovata in analisi univariate. Queste matrici sono chiamate SSCP (somme - di - quadrati e prodotti incrociati) matrici. Oltre alle ipotesi di prova, GLM Repeated Misure produce stime di parametri.

I contrasti di *a priori* sono disponibili per eseguire test di ipotesi sui fattori tra soggetti. 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. I mezzi marginali stimati danno stime dei valori medi previsti per le celle nel modello, e i tracciati di profili (tracciati di interazione) di questi mezzi consentono di visualizzare facilmente alcune relazioni.

Residui, valori attesi, distanza di Cook e valori di leva possono essere salvati come variabili nel file di dati per la verifica di ipotesi. Disponibile anche una matrice SSCP residua, ovvero una matrice quadrata

di somme di quadrati e cross - products di residui, una matrice di covarianza residua, ovvero la matrice SSCP residua divisa dai gradi di libertà dei residui, e la matrice di correlazione residua, ovvero la forma standardizzata della matrice di covarianza residua.

WLS Weight consente di specificare una variabile usata per dare osservazioni differenti pesi per un'analisi WLS (ponderata), magari per compensare una diversa precisione di misurazione.

**Esempio.** Dodici studenti sono assegnati ad un gruppo ad alta o bassa ansia in base ai loro punteggi su un test di ansia. Il rating ansioso è definito un fattore tra soggetti perché divide i soggetti in gruppi. Gli studenti sono ciascuno con quattro prove su un compito di apprendimento, e si registra il numero di errori per ogni sperimentazione. Gli errori per ogni trial sono registrati in variabili distinte, e un fattore di soggetto interno (trial) è definito con quattro livelli per le quattro prove. L'effetto di processo è risultato significativo, mentre l'interazione trial - by - ansia non è significativa.

**Metodi.** Le somme di quadrati di tipo I, Type II, Type III e Type IV possono essere utilizzate per valutare diverse ipotesi. Il metodo predefinito è il Tipo III.

**Statistiche.** Test di gamma post hoc e confronti multipli (per i fattori tra - soggetti): differenza meno significativa, Bonferroni, Sidak, Scheffé, Ryan - Einot - Gabriel - Welsch multiplo  $F$ , Ryan - Einot - Gabriel - Welsch multiplo, Student - Newman - Keuls, la differenza onestamente significativa di Tukey, bdi Tukey, Duncan, Hochberg GT2, Gabriel, Waller DuncanTest  $t$ , Dunnett (uno - sided e due - sided), il T2di Tamhane, il T3di Dunnett, i giochi - Howell e il Cdi Dunnett. Statistiche descrittive: medie osservate, deviazioni standard e conta per tutte le variabili dipendenti in tutte le cellule; il test di Levene per omogeneità di varianza; Box  $M$ ; e la prova di sfericità di Mauchly.

**Grafici.** Spread - versus - livello, residuo e profilo (interazione).

GLM Ripetute Misure Dati Considerazioni

**Dati.** Le variabili dipendenti dovrebbero essere quantitative. I fattori tra soggetti dividono il campione in sottogruppi discreti, come il maschio e la femmina. Questi fattori sono categoriali e possono avere valori numerici o valori di stringa. I fattori all'interno di soggetti sono definiti nella finestra di dialogo Repeated Misure Definisci Factor. Le covariate sono variabili quantitative correlate alla variabile dipendente. Per un'analisi delle misure ripetute, queste dovrebbero rimanere costanti ad ogni livello di una variabile all'interno.

Il file dei dati dovrebbe contenere una serie di variabili per ogni gruppo di misurazioni sui soggetti. Il set ha una variabile per ogni ripetizione della misurazione all'interno del gruppo. Un fattore all'interno dei soggetti è definito per il gruppo con il numero di livelli pari al numero di ripetizioni. Per esempio, le misurazioni del peso potrebbero essere assunte nei giorni diversi. Se le misurazioni della stessa proprietà sono state prese in cinque giorni, il fattore interno - soggetti potrebbe essere specificato come *giorno* con cinque livelli.

Per molteplici fattori all'interno dei soggetti, il numero di misurazioni per ogni soggetto è pari al prodotto del numero di livelli di ciascun fattore. Ad esempio, se le misurazioni sono state effettuate in tre momenti diversi ogni giorno per quattro giorni, il numero totale di misurazioni è di 12 per ogni soggetto. I fattori all'interno dei soggetti potrebbero essere specificati come *giorno* (4) e *ora* (3).

**Ipotesi.** Un'analisi delle misure ripetute può essere avvicinata in due modi, univariate e multivariate.

L'approccio univariato (noto anche come approccio split - plot o mixed - model) considera le variabili dipendenti come risposte ai livelli dei fattori all'interno dei soggetti. Le misurazioni su un soggetto dovrebbero essere un campione da una distribuzione normale multivariata e le matrici di varianza - covarianza sono le stesse attraverso le cellule formate dagli effetti tra soggetti. Alcune ipotesi sono fatte sulla matrice di varianza - covarianza delle variabili dipendenti. La validità della statistica  $F$  utilizzata nell'approccio univariato può essere assicurata se la matrice di varianza - covarianza è circolare in forma (Huynh e Mandeville, 1979).

Per testare questo presupposto si può utilizzare il test di sfericità di Mauchly, che esegue una prova di sfericità sulla matrice di varianza - covarianza di una variabile dipendente trasformata di ortonormalizzato. Il test di Mauchly viene visualizzato automaticamente per un'analisi delle misure ripetute. Per le piccole dimensioni del campione, questo test non è molto potente. Per le grandi dimensioni del campione, il test può essere significativo anche quando l'impatto della partenza sui risultati è piccolo. Se il significato del test è grande, si può ipotizza l'ipotesi di sfericità. Tuttavia, se il significato è piccolo e l'assunzione

di sfericità sembra violata, si può effettuare una regolazione del numeratore e dei gradi di libertà denominatori per convalidare la statistica univariata  $F$ . Tre le stime di questa regolazione, che si chiama **epsilon**, sono disponibili nella procedura GLM Repeated Measures. Sia il numeratore che il denominatore gradi di libertà devono essere moltiplicati per epsilon, e il significato del rapporto  $F$  deve essere valutato con i nuovi gradi di libertà.

L'approccio multivariato considera le misurazioni su un soggetto di essere un campione da una distribuzione normale multivariata e le matrici di varianza - covarianza sono le stesse attraverso le cellule formate dagli effetti tra soggetti. Per verificare se le matrici di varianza - covarianza attraverso le celle sono le stesse, è possibile utilizzare il test di Box  $M$ .

**Procedure correlate.** Utilizzare la procedura Explore per esaminare i dati prima di effettuare un'analisi della varianza. Se ci sono *non* misurazioni ripetute su ogni soggetto, utilizzare GLM Univariate o GLM Multivariato. Se ci sono solo due misurazioni per ogni soggetto (ad esempio pre - test e post - test) e non ci sono fattori tra - soggetti, è possibile utilizzare la procedura Pandate - Samples T Test.

Ottenere GLM Misure ripetute

1. Dai menu, scegliere:

**Analizzare > Modello Lineare Generale > Misure ripetute ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Digitare un nome fattore all'interno e il suo numero di livelli.

3. Fare clic su **Aggiungi**.

4. Ripetere questi passaggi per ogni fattore all'interno dei soggetti.

Per definire i fattori di misura per un design di misure ripetute doppiamente multivariate:

5. Digitare il nome di misura.

6. Fare clic su **Aggiungi**.

Dopo aver definito tutti i tuoi fattori e misure:

7. Fare clic su **Definisci**.

8. Selezionare una variabile dipendente che corrisponde ad ogni combinazione di fattori all'interno dei soggetti (e facoltativamente, misure) nella lista.

Per modificare le posizioni delle variabili, utilizzare le frecce su e giù.

Per apportare modifiche ai fattori all'interno dei soggetti, è possibile riaprire la finestra di dialogo Repeated Measures Definisci Factor senza chiudere la finestra di dialogo principale. Facoltativamente, è possibile specificare tra - soggetti fattori e covariati.

## GLM Misure ripetute definiscono i fattori

GLM Ripetute Misure analizza gruppi di variabili dipendenti correlate che rappresentano diverse misurazioni dello stesso attributo. Questa finestra di dialogo consente di definire uno o più fattori all'interno dei soggetti per l'utilizzo in GLM Repeated Measures. Si noti che l'ordine in cui si specificano i fattori all'interno dei soggetti è importante. Ogni fattore costituisce un livello all'interno del fattore precedente.

Per utilizzare le Misure Repeate è necessario impostare correttamente i dati. È necessario definire i fattori all'interno dei soggetti in questa finestra di dialogo. Notate che questi fattori non sono variabili esistenti nei vostri dati ma piuttosto fattori che si definiscono qui.

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

**Esempio.** In uno studio sulla perdita di peso, supponiamo che i pesi di diverse persone siano misurati ogni settimana per cinque settimane. Nel file dei dati, ogni persona è un soggetto o un caso. I pesi per le settimane vengono registrati nelle variabili *weight1*, *weight2* e così via. Il sesso di ogni persona viene registrato in un'altra variabile. I pesi, misurati per ogni soggetto ripetutamente, possono essere

raggruppati definendo un fattore all'interno dei soggetti. Il fattore potrebbe essere chiamato *settimana*, definito per avere cinque livelli. Nella finestra di dialogo principale, le variabili *weight1*, ..., *weight5* vengono utilizzati per assegnare i cinque livelli di *settimana*. La variabile nel file dati che raggruppa maschi e femmine (*sesto*) può essere specificata come un fattore tra - soggetti per studiare le differenze tra maschi e femmine.

**Misure.** Se i soggetti sono stati sottoposti a test su più misure in ogni momento, definire le misure. Ad esempio, il battito e la velocità di respirazione potrebbero essere misurati su ogni soggetto ogni giorno per una settimana. Queste misure non esistono come variabili nel file dati ma sono definite qui. Un modello con più di una misura viene talvolta definito un modello di misure ripetute doppiamente multivariate.

## Modello Misure ripetute GLM

**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 dei fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

**Tra - Subito.** Sono elencati i fattori tra i soggetti e le covariate.

**Modello.** Il modello varia in base alla natura dei dati in uso. Dopo aver selezionato **Custom**, è possibile selezionare gli effetti e le interazioni all'interno e gli effetti tra i soggetti e le interazioni che sono di interesse per la tua analisi.

**Somma dei quadrati.** Il metodo di calcolo delle somme di quadrati per il modello tra soggetti. Per i modelli bilanciati o sbilanciati tra - soggetti senza celle mancanti, il metodo della somma dei quadrati di tipo III è quello più utilizzato.

## Creazione di termini e termini personalizzati

### Crea termini

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### Crea termini personalizzati

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## 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 noto anche come decomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene adattato 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 adattato 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 della 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 calcola le somme dei quadrati di un effetto nel disegno come la somma dei quadrati adattata per qualsiasi altro effetto che non lo contiene 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 pesate di Yates. Il metodo della 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  $F$  nel 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 della 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 Contrade

I contrasti sono utilizzati per testare le differenze tra i livelli di un fattore tra soggetti. È possibile specificare un contrasto per ogni fattore tra - soggetti nel modello. I contrasti rappresentano combinazioni lineari dei parametri.

Il test sull'ipotesi è basato sull'ipotesi null  $\mathbf{LBM} = 0$ , dove  $\mathbf{L}$  è la matrice dei coefficienti di contrasto,  $\mathbf{B}$  è il vettore del parametro e  $\mathbf{M}$  è la matrice media che corrisponde alla trasformazione media per la variabile dipendente. È possibile visualizzare questa matrice di trasformazione selezionando **matrice di trasformazione** nella finestra di dialogo Opzioni di misure Repeate. Se per i fattori entro soggetti sono disponibili, ad esempio, quattro variabili dipendenti, un fattore entro soggetti di quattro livelli e contrasti polinomiali (valore predefinito), la matrice  $\mathbf{M}$  sarà  $(0,5 \ 0,5 \ 0,5 \ 0,5)'$ . Quando viene specificato un contrasto, viene creata una matrice  $\mathbf{L}$  tale che le colonne corrispondenti al fattore tra soggetti corrispondono al contrasto. Le altre colonne vengono adattate 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 dei fattori 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 delle tendenze polinomiali.

## GLM Misure ripetute Profilo profilo

I grafici di profilo, o grafici di interazione, risultano utili per il confronto delle medie marginali di un modello. Un grafico di profilo è un grafico a linee in cui ciascun punto indica la media marginale stimata di una variabile dipendente (adattata 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 tracciati. I tracciati del profilo vengono creati per ogni variabile dipendente. Sia i fattori tra soggetti che i fattori all'interno dei soggetti possono essere utilizzati nei tracciati del profilo.

Il grafico di 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.

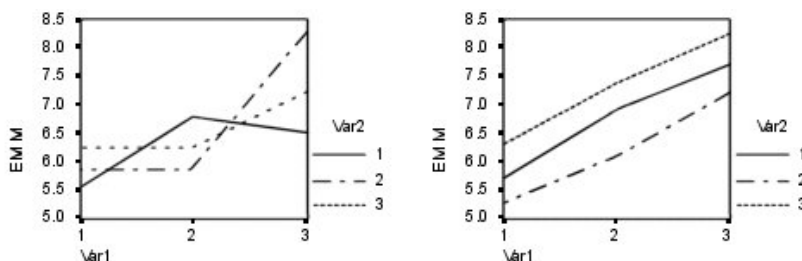


Figura 2. Grafico con linee non parallele (a sinistra) e grafico con linee parallele (a destra)

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 Misure Ripetute Post Hoc

**Test di confronti multipli post hoc.** Dopo aver determinato l'esistenza di differenze tra le medie, i test post hoc di intervallo e i confronti pairwise multipli consentono di determinare quale media differisce dalle altre. I confronti vengono eseguiti su valori a cui non è stato apportato alcun adattamento. Questi test non sono disponibili se non ci sono fattori tra - soggetti, e i test di confronto multipli post hoc vengono eseguiti per la media sui livelli dei fattori all'interno dei soggetti.

I test di confronto multiplo usati più di frequente sono il test di Bonferroni e il test HSD di Tukey. Il **test di Bonferroni**, basato sulla statistica  $t$  di Student, consente di adattare il livello di significatività osservato in base al fatto che vengono eseguiti confronti multipli. Il **test  $t$  di Sidak** adatta inoltre il livello di significatività ed è più restrittivo del test di Bonferroni. Il **test HSD di Tukey** utilizza la statistica di intervallo studentizzato per effettuare tutti i confronti pairwise tra gruppi e imposta il tasso di errore sperimentale sul valore del tasso di errore per l'insieme di tutti i confronti pairwise. Quando si eseguono test su un elevato numero di coppie di medie, il test HSD 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 HSD di Tukey, ma utilizza il modulo massimo studentizzato. Il test di Tukey risulta in genere più efficace. Anche il **test per confronti pairwise 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 pairwise 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 dei fattori (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 a qualsiasi livello del fattore è inferiore a quella della categoria di controllo, selezionare **< Control**. Allo stesso modo, per verificare se la media a qualsiasi livello del fattore è superiore a quella della categoria di controllo, selezionare **> Control**.

Ryan, Einot, Gabriel e Welsch (R-E-G-W) hanno sviluppato due test di intervallo 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 di 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, utilizzare **T2 di Tamhane** (test per confronti pairwise conservativo basato su un test  $t$ ), **T3 di Dunnett** (test per confronti pairwise basato sui moduli massimi studentizzati), **il test per confronti pairwise di Games-Howell** (a volte liberale) o **C di Dunnett** (test per confronti pairwise basato sull'intervallo studentizzato).

**Il test di intervallo multiplo di Duncan**, il test di Student-Newman-Keuls (**S-N-K**) e il test **b di Tukey** sono test di 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 di intervallo che usa la media armonica della dimensione del campione nel caso di dimensioni del campione non uguali.

Il livello di significatività del test di **Scheffé** è progettato per consentire il test di tutte le possibili combinazioni lineari delle medie di gruppo, non solo dei confronti pairwise disponibili in questa funzione. Ne risulta che il test di Scheffé è spesso più conservativo di altri test, ciò significa che è richiesta una maggiore differenza tra le medie per la significatività.

Il test di confronto multiplo pairwise 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 adattamento del livello di significatività osservata per confronti multipli.

**Test visualizzati.** I confronti pairwise sono disponibili per i test LSD, Sidak, Bonferroni, Games-Howell, Tamhane (T2) e (T3), C di Dunnett e T3 di Dunnett. 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 di intervallo. Il test Differenza significativa di Tukey, GT2 di Hochberg, il test di Gabriel e il test di Scheffé sono sia test di confronto multiplo che test di intervalli.

## Medie marginali stimate GLM

Selezionare i fattori e le interazioni per cui si desiderano le stime delle medie marginali della popolazione nelle celle. Queste medie vengono adattate per le eventuali covariate.

### Confronta effetti principali

Consente di eseguire confronti pairwise 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.

### Confronta effetti principali semplici

L'impostazione è abilitata ogni qualvolta l'elenco di destinazione contiene uno o più effetti di prodotto o interazione (ad esempio  $A*B$ ,  $A*B*C$ ). L'impostazione supporta la specifica dei confronti tra semplici effetti principali, che sono effetti principali nidificati all'interno dei livelli di altri fattori.

### Adattamento intervallo di confidenza

Selezionare la differenza meno significativa (LSD), l'adattamento di Bonferroni o di Sidak agli intervalli di confidenza e la significatività. Questo articolo è disponibile solo quando viene selezionato **Confronta gli effetti principali** e / o **Confronta semplici effetti principali**.

## Specifica delle medie marginali stimate

1. Dai menu scegliere una delle procedure disponibili sotto > **Analizza** > **Modello lineare generale**.
2. Nella finestra di dialogo principale, fare clic su **Medie EM**.

## GLM Misure Ripetute Salvataggio

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 Statistica, è necessario salvare il file di dati corrente.

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

- *Non standardizzati.* Il valore che il modello prevede per la variabile dipendente.
- *Errore standard.* Una stima della deviazione standard del valore medio della variabile dipendente per i casi che hanno gli stessi valori delle variabili indipendenti.

**Diagnostiche.** 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. Disponibili sono la distanza di Cook e i valori di leva 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. Un valore alto del D di Cook indica che l'esclusione di un caso dal calcolo delle statistiche di regressione modifica sostanzialmente i coefficienti.
- *Valori di leva.* Valori di leva non centrati. L'influenza relativa di ogni osservazione sull'adattamento del modello.

**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 sua deviazione standard. I residui standardizzati, noti anche come residui di Pearson, hanno una media pari a 0 e una deviazione standard pari a 1.
- *Studentizzati.* Il residuo diviso per una stima della sua deviazione standard che varia da caso a caso, a seconda della distanza dei valori di ogni caso sulle variabili indipendenti dai mezzi delle variabili indipendenti. A volte indicato come residui internamente studentizzati.
- *Eliminato.* Il residuo per un caso quando tale caso è escluso dal calcolo dei coefficienti di regressione. È la differenza tra il valore della variabile dipendente e il valore previsto adattato.

**Statistiche dei coefficienti.** Salva una matrice di varianza - covarianza delle stime dei parametri in un dataset o un file di dati. Per ciascuna variabile dipendente è inoltre disponibile una riga di stime del parametro, una riga di valori di significatività per le statistiche *t* corrispondenti alle stime del parametro e una riga di gradi di libertà dei residui. Per ciascuna variabile dipendente di modelli multivariati sono disponibili righe simili. È possibile utilizzare questi dati di matrice in altre procedure che leggono i file di matrice. I dataset possono anche essere riutilizzati nella stessa sessione, ma non vengono salvati come file a meno che siano stati salvati come tali al termine della sessione. I nomi dei dataset devono essere conformi alle regole dei nomi delle variabili.

## Opzioni Misure ripetute GLM

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

### Visualizza

#### Statistiche descrittive

Produce mezzi osservati, deviazioni standard e conta per tutte le variabili dipendenti in tutte le celle.

**Stime della dimensione degli effetti**

Dà un valore di tipo beta parziale per ogni effetto e ogni stima di parametro. La statistica eta al quadrato consente di ottenere la proporzione della variabilità totale attribuibile a un fattore.

**Potenza osservata**

Ottiene la potenza del test quando l'ipotesi alternativa è impostata in base al valore osservato.

**Stime dei parametri**

Produce le stime dei parametri, gli errori standard, i test  $t$ , gli intervalli di confidenza e la potenza osservata per ogni test.

**Matrici SSCP**

Visualizzate le ipotesi e le matrici SSCP di errore.

**Matrice SSCP dei residui**

Visualizza la matrice Residual SSCP.

**Matrice di trasformazione**

Visualizza la prova di sfericità di Bartlett della matrice di covarianza residua.

**Test di omogeneità**

Produce il test di Levene dell'omogeneità della varianza per ogni variabile dipendente in tutte le combinazioni di livello dei fattori tra soggetti, per soli fattori tra soggetti. Inoltre, i test di omogeneità includono la prova di Box  $M$  dell'omogeneità delle matrici di covarianza delle variabili dipendenti in tutte le combinazioni di livello dei fattori tra i soggetti.

**Grafico di confronto tra diffusione e livello**

Utile per verificare le ipotesi sui dati. Questa opzione è disabilitata se non ci sono fattori.

**Grafico dei residui**

Produce una trama dei residui osservati per ogni variabile dipendente. Questi grafici risultano utili per l'analisi dell'ipotesi di uguaglianza della varianza. Questa opzione è disabilitata se non ci sono fattori.

**Mancanza di adattamento**

Verifica se il rapporto tra la variabile dipendente e le variabili indipendenti può essere adeguatamente descritto dal modello.

**Funzioni generali stimabili**

Consente di costruire test di ipotesi personalizzati in base alle funzioni generali stimabili. Le righe in qualsiasi matrice di coefficiente di contrasto sono combinazioni lineari della funzione generale stimabile (s).

**Livello di significatività**

Potrebbe risultare utile adattare 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, il livello associato degli intervalli di confidenza viene visualizzato nella finestra di dialogo.

## Funzioni Aggiuntive Comando GLM

Queste caratteristiche possono applicarsi ad analisi di misure univariate, multivariate o ripetute. Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare gli effetti nidificati nel 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).
- Creare una matrice **L**, una matrice **M** o una matrice **K** personalizzata (utilizzando i sottocomandi LMATRIX, MMATRIX e 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 nell'elenco dei fattori (tramite il sottocomando EMMEANS).
- Assegnare un nome alle variabili temporanee (tramite il sottocomando SAVE).
- Costruire un file di dati matrice di correlazione (tramite il sottocomando OUTFILE).
- Costruire un file di dati matrice 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).

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

## Analisi Componenti della varianza

La Procedura Variance Components, per modelli a effetti misti, stima il contributo di ogni effetto casuale alla varianza della variabile dipendente. Questa procedura è particolarmente interessante per l'analisi di modelli misti come split plot, misure ripetute univariate e disegni di blocco casuali. Calcolando i componenti della varianza, è possibile determinare dove focalizzare l'attenzione per ridurre la varianza.

Sono disponibili quattro diversi metodi per stimare i componenti della varianza: minimo norm quadratico stimatore non biased (MINQUE), analisi della varianza (ANOVA), probabilità massima (ML) e probabilità massima limitata (REML). Diverse specifiche sono disponibili per i diversi metodi.

L'output predefinito per tutti i metodi include le stime dei componenti di varianza. Se viene utilizzato il metodo ML o il metodo REML viene visualizzata anche una tabella di matrice di covarianza asintotica. Altra uscita disponibile include una tabella ANOVA e le piazze medie previste per il metodo ANOVA e una storia di iterazione per i metodi ML e REML. La Procedura Di Varianza Componenti è pienamente compatibile con la procedura GLM Univariate.

WLS Weight consente di specificare una variabile usata per dare osservazioni differenti pesi per un'analisi ponderata, magari per compensare variazioni di precisione di misurazione.

**Esempio.** In una scuola agricola, dopo un mese si misurano i guadagni di peso per i maiali in sei diversi lettini. La variabile litter è un fattore casuale con sei livelli. (I sei litri studiati sono un campione casuale di una grande popolazione di suoli di suino). L'investigatore scopre che la varianza in aumento di peso è attribuibile alla differenza in litri molto più che alla differenza nei suini all'interno di una lettiera.

Considerazioni Dati Componenti di varianza

**Dati.** La variabile dipendente è quantitativa. I fattori sono categoriali. Possono avere valori numerici o valori di stringa fino a otto byte. Almeno uno dei fattori deve essere casuale. Ovvero i livelli del fattore devono essere un campione casuale di livelli possibili. Le covariate sono variabili quantitative correlate alla variabile dipendente.

**Ipotesi.** Tutti i metodi ipotizza che i parametri del modello di un effetto casuale abbiano zero mezzi e varianze costanti finite e si uniscono reciprocamente. Anche i parametri del modello da diversi effetti casuali sono poco correlati.

Il termine residuo ha anche una varianza media zero e finita. È slegato con parametri di modello di qualsiasi effetto casuale. Si ipotizza che i termini residui da diverse osservazioni siano non correlati.

In base a queste ipotesi, le osservazioni dallo stesso livello di un fattore casuale sono correlate. Questo fatto distingue un modello di componente variance da un modello lineare generale.

ANOVA e MINQUE non richiedono ipotesi di normalità. Sono entrambi robusti per le partenze moderate dall'assunzione di normalità.

ML e REML richiedono il parametro del modello e il termine residuo da distribuire normalmente.

**Procedure correlate.** Utilizzare la procedura Explore per esaminare i dati prima di fare analisi dei componenti di varianza. Per i test di ipotesi, utilizzare GLM Univariate, GLM Multivariate e GLM Repeated Measures.

Ottenimento Tabelle Componenti di varianza

1. Dai menu, scegliere:

**Analizzare > Modello Lineare Generale > Componenti di varianza ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile dipendente.

3. Selezionare le variabili per l'opzione Fattori fissi, Fattori casuali o Covariate, a seconda dei dati in uso. Per specificare una variabile di peso, utilizzare WLS Peso.

## Modello Componenti Variance

**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 dei fattori e covariate. È necessario indicare tutti i termini da includere nel modello.

**Fattori & Covariati.** 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 contenere un fattore casuale.

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.

**Includi intercetta nel modello.** Di solito l'intercettazione è inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercetta può essere esclusa.

## Creazione di termini e termini personalizzati

### Crea termini

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### Crea termini personalizzati

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## Opzioni Componenti di varianza

**Metodo.** È possibile scegliere uno dei quattro metodi per stimare i componenti della varianza.

- **MINQUE** (stimatore non biased quadratico minimo norm) produce stime invarianti rispetto agli effetti fissi. Se i dati sono normalmente distribuiti e le stime sono corrette, questo metodo produce la minima varianza tra tutti gli estimatori non biased. È possibile scegliere un metodo per i pesi precedenti ad effetto casuale.
- **ANOVA** (analisi della varianza) calcola preventivi unbiased utilizzando le somme di quadrati di Tipo I o di Tipo III per ogni effetto. Il metodo ANOVA a volte produce stime di varianza negativa, che possono indicare un modello errato, un metodo di stima inappropriato o un bisogno di più dati.
- **Massimo verosimiglianza** (ML) produce stime che sarebbero più coerenti con i dati effettivamente osservati, utilizzando iterazioni. Queste stime possono essere biased. Questo metodo è asintoticamente normale. Le stime ML e REML sono invarianti sotto la traduzione. Questo metodo non tiene conto dei gradi di libertà utilizzati per stimare gli effetti fissi.
- **Limitate probabilità massima** (REML) le stime riducono le stime ANOVA per molti (se non tutti) casi di dati bilanciati. Poiché questo metodo viene regolato per gli effetti fissi, dovrebbe avere errori standard più piccoli rispetto al metodo ML. Questo metodo tiene conto dei gradi di libertà utilizzati per stimare gli effetti fissi.

**Priori a effetto casuale. Uniforme** implica che tutti gli effetti casuali e il termine residuo abbiano un impatto uguale sulle osservazioni. Lo schema **Zero** equivale ad assumere varianze a effetto casuale zero. Disponibile solo per il metodo MINQUE.

**Somma dei quadrati. Tipo I** somme di quadrati sono utilizzate per il modello gerarchico, spesso utilizzato nella letteratura componente varianza. Se si sceglie **Tipo III**, il default in GLM, le stime della varianza possono essere utilizzate in GLM Univariate per l'ipotesi test con somme di quadrati di tipo III. Disponibile solo per 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.

**Visualizza.** Per il metodo ANOVA è possibile scegliere di visualizzare somme di quadrati e quadrati medi previsti. Se hai selezionato **Massimo verosimiglianza** o **Limitazione massima limitata**, è possibile visualizzare una cronologia delle iterazioni.

## Somma di Squares (Componenti di variazione)

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

**Tipo I.** Questo metodo è noto anche come decomposizione gerarchica del metodo Somma dei quadrati. Ciascun termine viene adattato solo per i termini del modello che lo precedono. Il metodo della somma dei quadrati di tipo I è comunemente utilizzato per:

- 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 calcola le somme di quadrati di un effetto nella progettazione come le somme di quadrati regolate per qualsiasi altro effetto che non la contengono e ortogonale a eventuali effetti (se presenti) 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. Pertanto, questo tipo viene spesso considerato utile per un modello sbilanciato senza celle mancanti. In un disegno fattoriale privo di celle mancanti, questo metodo equivale alla tecnica dei quadrati delle medie pesate di Yates. Il metodo della somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- Eventuali modelli elencati in Tipo I.

- Qualsiasi modello equilibrato o sbilanciato senza celle vuote.

## Varianza Componenti Salva su Nuovo file

È possibile salvare alcuni risultati di questa procedura ad un nuovo file di dati IBM SPSS Statistica .

**stime dei componenti di variazione.** Salva le stime dei componenti della varianza e stima le etichette ad un file di dati o un dataset. Questi possono essere utilizzati nel calcolo di più statistiche o in ulteriori analisi nelle procedure GLM. Ad esempio, è possibile utilizzarli per calcolare intervalli di confidenza o ipotesi di prova.

**Covarianza del componente.** Salva una matrice di varianza - covarianza o una matrice di correlazione in un file di dati o un dataset. Disponibile solo se è stato specificato **Massimo probabilità** o **Limitazione massima limitata** .

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

È possibile utilizzare il comando MATRIX per estrarre i dati necessari dal file dei dati e quindi calcolare gli intervalli di confidenza o eseguire i test.

## VARCOMP Command Aggiuntive Features

Il linguaggio della sintassi dei comandi consente inoltre di:

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

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

## Modelli misti lineari

---

La Procedura Lineare Mixed Models espande il modello lineare generale in modo che i dati siano consentiti per esibire la variabilità correlata e non costante. Il modello lineare misto, quindi, fornisce la flessibilità di modellazione non solo dei mezzi dei dati ma delle loro varianze e covarianze.

La Procedura Lineare Misti Modelli è anche uno strumento flessibile per l'adattamento di altri modelli che possono essere formulati come modelli lineari misti. Tali modelli includono modelli multilivello, modelli lineari gerarchici e modelli di coefficiente casuali.

### Esempio

Una catena di negozi di alimentari è interessata agli effetti di vari coupon sulla spesa dei clienti. Prendendo un campione casuale dei loro clienti regolari, seguono la spesa di ogni cliente per 10 settimane. In ogni settimana, un coupon diverso viene spedito ai clienti. Linear Mixed Models viene utilizzato per stimare l'effetto di diverse cedole sulla spesa mentre si adegua per la correlazione a causa di osservazioni ripetute su ogni soggetto nelle 10 settimane.

### Metodi

Probabilità massima di stima (ML) e di massima probabilità limitata (REML).

### Statistiche

Statistiche descrittive: dimensioni campionarie, medie e deviazioni standard della variabile dipendente e covariate per ogni combinazione di livello distinto dei fattori. Informazioni di livello del fattore: valori ordinati dei livelli di ciascun fattore e delle loro frequenze. Inoltre, le stime dei parametri e gli intervalli di confidenza per gli effetti fissi e le prove Wald e gli intervalli di confidenza per i parametri delle matrici di covarianza. Le somme di quadrati di tipo I e di tipo III possono essere utilizzate per valutare diverse ipotesi. Il metodo predefinito è il Tipo III.

## considerazioni sui dati Modelli misti lineari

### Dati

La variabile dipendente dovrebbe essere quantitativa. I fattori dovrebbero essere categoriali e possono avere valori numerici o valori di stringa. Le covariate e la variabile peso dovrebbero essere quantitative. I soggetti e le variabili ripetute possono essere di qualsiasi tipo.

### Ipotesi

La variabile dipendente si assume che sia linearmente correlata ai fattori fissi, fattori casuali e covariati. Gli effetti fissi modellano la media della variabile dipendente. Gli effetti casuali modellano la struttura di covarianza della variabile dipendente. Molteplici effetti casuali sono considerati indipendenti tra loro, e le matrici di covarianza separate saranno calcolate per ciascuno; tuttavia, i termini del modello specificati sullo stesso effetto casuale possono essere correlati. Le misure ripetute modellano la struttura di covarianza dei residui. Anche la variabile dipendente viene assunta per provenire da una normale distribuzione.

### Procedure correlate

Utilizzare la procedura Explore per esaminare i dati prima di eseguire un'analisi. Se non si sospetta che ci sia una variabilità correlata o non costante, è possibile utilizzare la procedura GLM Univariate o GLM Repeated Measures. È possibile utilizzare in alternativa la procedura di Analisi Componenti di variazione se gli effetti casuali hanno una struttura di covarianza componenti varianze e non ci sono misure ripetute.

## Ottenimento di un Linear Mixed Models Analysis

1. Dai menu, scegliere:

**Analizza > Modelli > Lineare...**

2. Facoltativamente, selezionare una o più variabili di soggetto.
3. Facoltativamente, selezionare una o più variabili ripetute. Se vengono definite variabili ripetute, selezionare un **Tipo di Covarianza Ripetuto** dall'elenco a discesa.
4. Facoltativamente, selezionare una o più variabili di misura Kronecker.
5. Opzionalmente, selezionare una struttura di covarianza residua.
6. Fare clic su **Continua**.
7. Selezionare una variabile dipendente.
8. Selezionare almeno un fattore o covariato.
9. Fare clic su **Fixed** o **Random** e specificare almeno un modello a effetti fissi o casuali.

Opzionalmente, selezionare una variabile di ponderazione.

## Modelli Misti Lineari: Soggetti e Repeati

Questa finestra di dialogo consente di selezionare le variabili che definiscono i soggetti, le osservazioni ripetute, le misure Kronecker e di scegliere una struttura di covarianza per i residui.

### Soggetti

Un soggetto è un'unità osservazionale che può essere considerata indipendente da 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 modellare la correlazione tra queste osservazioni. Ci si potrebbe aspettare, per esempio, che le letture della pressione sanguigna di un paziente effettuate in occasione di visite al medico consecutive siano correlate.

I soggetti possono anche essere definiti dalla combinazione di livello fattoriale di più variabili; ad esempio, è possibile specificare *Gender* e *Categoria Età* come variabili di soggetto per modellare la convinzione che *maschi sopra i 65 anni* siano simili tra loro ma indipendenti da *maschi under 65* e *femmine*.

Tutte le variabili specificate nell'elenco **Subjects** vengono utilizzate per definire i soggetti per la struttura di covarianza residua. È possibile utilizzare alcune o tutte le variabili per definire i soggetti per la struttura di covarianza a effetti casuali.

### **Ripetuto**

Le variabili specificate in questo elenco sono utilizzate per identificare osservazioni ripetute. Per esempio, una singola variabile *Settimana* potrebbe identificare 10 settimane di osservazioni in uno studio medico, mentre *Mese* e *Giorno* potrebbero essere utilizzati insieme per identificare le osservazioni giornaliere nel corso di un anno.

### **Tipo di covarianza ripetuta**

Specifica la struttura di covarianza per i residui. Le strutture disponibili sono le seguenti:

- Ante dipendenza: primo ordine
- AR(1)
- AR1 prodotto diretto (UN\_AR1)
- Prodotto diretto non strutturato (UN\_UN)
- Simmetria del composto prodotto diretto (UN\_CS)
- AR(1): eterogeneo
- ARMA(1,1)
- Simmetria del composto
- Simmetria del composto: Metrica di correlazione
- Simmetria del composto: Eterogeneo
- Diagonale
- Analitica fattoriale: primo ordine
- Analitica fattoriale: primo ordine, eterogenea
- Huynh-Feldt
- Identità scalata
- Toeplitz
- Toeplitz: eterogeneo
- Non strutturato
- Non strutturato: Metrica di correlazione
- Spaziale: potenza
- Spaziale: esponenziale
- Spaziale: gaussiano
- Spaziale: lineare
- Spaziale: log lineare
- Spaziale: sferico

### **Misure Kronecker**

Selezionare le variabili che specificano la struttura del soggetto o le misurazioni della covarianza di Kronecker e determinare in che modo vengono corretti gli errori di misurazione. Il campo è disponibile solo quando viene selezionato uno dei seguenti **Tipo di covarianza ripetuta**:

- AR1 prodotto diretto (UN\_AR1)
- Prodotto diretto non strutturato (UN\_UN)
- Simmetria del composto prodotto diretto (UN\_CS)

### **Coordinate covarianza spaziale**

Le variabili in questo elenco specificano le coordinate delle osservazioni ripetute quando uno dei tipi di covarianza spaziale viene selezionato per il tipo di covarianza ripetuta.

Per ulteriori informazioni, consultare l'argomento [“Strutture covarianza”](#) a pagina 102.

## Modelli misti lineari a effetti fissi

**Effetti Fissi.** Non esiste un modello predefinito, quindi è necessario specificare esplicitamente gli effetti fissi. In alternativa, è possibile costruire termini nidificati o non nidificati.

**Includi intercettazione.** 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 Type III è più comunemente utilizzato.

### Crea Termini Non Nidificati

Per i fattori e le covariate selezionati:

**Factoriale.** Crea 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.

### Crea 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 negozi di generi alimentari può seguire la spesa dei propri clienti in diverse località del negozio. Poiché ogni cliente frequenta solo una di quelle località, l'effetto *Cliente* può essere detto **nidificato all'interno** l'effetto *Posizione negozio*.

Inoltre, è possibile includere effetti di interazione o aggiungere più livelli di nidificazione al termine nidificato.

**Limitazione.** I termini nidificati sono sottoposti alle seguenti restrizioni:

- Tutti i fattori compresi in un'interazione devono essere univoci. Di conseguenza, se  $A$  è un fattore, non è possibile specificare  $A*A$ .
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Di conseguenza, se  $A$  è un fattore, non è possibile specificare  $A(A)$ .
- Nessun effetto può essere nidificato all'interno di una covariata. Di conseguenza, 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 noto anche come decomposizione gerarchica del metodo Somma dei quadrati. Ogni termine viene regolato solo per il termine che lo precede nel modello. 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 calcola le somme di quadrati di un effetto nella progettazione come le somme di quadrati regolate per qualsiasi altro effetto che non la contengono e ortogonale a eventuali effetti (se presenti) 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 pesate di Yates. Il metodo della somma dei quadrati Tipo III è in genere usato con i seguenti elementi:

- Eventuali modelli elencati in Tipo I.
- Qualsiasi modello equilibrato o sbilanciato senza celle vuote.

## Modelli misti lineari a effetti casuali

**Tipo di covarianza.** Questo consente di specificare la struttura di covarianza per il modello a effetti casuali. Viene stimata una matrice di covarianza separata per ogni effetto casuale. Le strutture disponibili sono le seguenti:

- Ante dipendenza: primo ordine
- AR(1)
- AR(1): eterogeneo
- ARMA(1,1)
- Simmetria del composto
- Simmetria del composto: Matrice di correlazione
- Simmetria del composto: Eterogeneo
- Diagonale
- Analitica fattoriale: primo ordine
- Analitica fattoriale: primo ordine, eterogenea
- Huynh-Feldt
- Identità scalata
- Toeplitz
- Toeplitz: eterogeneo
- Non strutturato
- Non strutturato: Matrice di correlazione
- Componenti della varianza

Per ulteriori informazioni, consultare l'argomento [“Strutture covarianza”](#) a pagina 102.

**Effetti casuali.** Non esiste un modello predefinito, quindi è necessario specificare esplicitamente gli effetti casuali. In alternativa, è possibile costruire termini nidificati o non nidificati. Puoi anche scegliere di includere un termine di intercettazione nel modello a effetti casuali.

È possibile specificare più modelli a effetti casuali. Dopo aver costruito il primo modello, fare clic su **Avanti** per costruire il modello successivo. Clicca su **Previous** per scorrere indietro attraverso i modelli esistenti. Ogni modello a effetto casuale è assunto per essere indipendente da ogni altro modello a effetto casuale; cioè le matrici di covarianza separate verranno calcolate per ciascuno. I termini specificati nello stesso modello a effetto casuale possono essere correlati.

**Raggruppamenti di soggetti.** Le variabili elencate sono quelle selezionate come variabili di soggetto nella finestra di dialogo Seleziona Argomenti / Variabili ripetute. Scegliete alcuni o tutti questi per definire i soggetti per il modello di effetti casuali.

**Visualizza le previsioni dei parametri per questa serie di effetti casuali.** Specifica di visualizzare le stime dei parametri ad effetti casuali.

## Stima dei modelli misti lineari

### Metodo

Selezionare la stima della massima verosimiglianza o della massima verosimiglianza riservata.

### Gradi di libertà

Fornisce opzioni per la definizione dei gradi di libertà per tutti i test.

### Metodo dei residui

Il metodo dei residui ha un determinato grado di libertà per tutti i test. È utile se la dimensione del campione è sufficientemente grande, se i dati sono bilanciati oppure se il modello usa un tipo di covarianza più semplice (ad esempio identità scalata o diagonale).

### approssimazione di Satterthwaite

Il metodo di Satterthwaite ha un campo gradi di libertà attraverso i test. È utile nel caso in cui la dimensione del campione è piccola oppure i dati non sono bilanciati oppure il modello usa un tipo di covarianza complesso (ad esempio non strutturato).

### Approssimazione di Kenward-Roger

Il metodo di Kenward-Roger offre uno stimatore a campioni piccoli più preciso per la varianza-covarianza dei parametri degli effetti fissi e dei gradi di libertà approssimativi del denominatore nei t-test e negli F-test. Il metodo introduce un fattore di scala per la statistica F e stima questo e il grado di libertà del denominatore utilizzando un'espansione della serie Taylor per la struttura casuale stimata all'interno dei dati.

**Nota:** Il metodo Kenward-Roger è utilizzato nella covarianza basata sul modello (invece della covarianza robusta). Quando vengono selezionati il metodo Kenward-Roger e la covarianza robusta, il metodo Kenward-Roger viene applicato alla covarianza basata su modello e viene presentato il seguente avviso: "Poiché il metodo Kenward-Roger è selezionato, il metodo di covarianza robusta viene modificato nel metodo di covarianza basato sul modello".

### Iterazioni

Sono disponibili le seguenti opzioni:

#### Numero massimo di iterazioni

Specificare un intero non negativo.

#### Numero massimo di dimezzamenti

Ad ogni iterazione, la dimensione di passo viene ridotta di un fattore di 0,5 finché la verosimiglianza log non aumenta o non si raggiunge il numero massimo di dimezzamenti. Specificare un intero positivo.

#### Stampa cronologia delle iterazioni per ogni n fasi

Mostra una tabella contenente il valore e le stime dei parametri della funzione di verosimiglianza per ciascuna iterazione  $n$  che inizia con l'iterazione 0<sup>esima</sup> (stime iniziali). Se si sceglie di stampare la cronologia dell'iterazione, l'ultima iterazione viene sempre stampata a prescindere dal valore  $n$ .

### Convergenza logaritmo della verosimiglianza

Si presume la convergenza se la modifica assoluta o la modifica relativa nella funzione di verosimiglianza log sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Convergenza parametri

Si presume la convergenza se la modifica assoluta massima o la modifica relativa massima delle stime del parametro sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Convergenza hessiana

Per la specifica **Absolute**, la convergenza viene presunta se la statistica basata sulla convergenza hessiana è inferiore al valore specificato. Per la specifica **Relativa**, si presume la convergenza se la statistica è minore del prodotto tra il valore specificato e il valore assoluto della log-verosimiglianza. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Fasi del punteggio massimo

Richieste di utilizzare l'algoritmo di punteggio Fisher fino all'iterazione numero  $n$ . Specificare un intero non negativo.

### Tolleranza della singolarità

Questo valore viene utilizzato come tolleranza nel controllo della singolarità. Specificare un valore positivo.

## Statistiche Modelli Misti Lineari

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

- **Statistiche descrittive.** Visualizza le dimensioni del campione, i mezzi e le deviazioni standard della variabile dipendente e delle covariate (se specificato). Queste statistiche vengono visualizzate per ogni combinazione di livello distinto dei fattori.
- **Riepilogo elaborazione casi.** Visualizza i valori ordinati dei fattori, le variabili di misura ripetute, i soggetti di misura ripetuta e i soggetti a effetti casuali e le loro frequenze.

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

- **Le stime dei parametri per gli effetti fissi.** Visualizza le stime dei parametri a effetto fisso e i relativi errori standard approssimativi.
- **Test per i parametri di covarianza.** Visualizza gli errori standard asintotici e le prove Wald per i parametri di covarianza.
- **Correlazioni delle stime dei parametri.** Visualizza la matrice di correlazione asintotica delle stime dei parametri a effetto fisso.
- **Le covarianze delle stime dei parametri.** Visualizza la matrice di covarianza asintotica delle stime dei parametri a effetto fisso.
- **Le covarianze degli effetti casuali.** Visualizza la matrice di covarianza stimata degli effetti casuali. Questa opzione è disponibile solo quando viene specificato almeno un effetto casuale. Se una variabile soggetto viene specificata per un effetto casuale, viene visualizzato il blocco comune.
- **Le covarianze dei residui.** Visualizza la matrice di covarianza residua stimata. Questa opzione è disponibile solo quando è stata specificata una variabile ripetuta. Se viene specificata una variabile di soggetto, viene visualizzato il blocco comune.
- **Matrice del coefficiente di contrasto.** Questa opzione visualizza le funzioni stimabili utilizzate per la verifica degli effetti fissi e delle ipotesi personalizzate.

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

## Linear Mixed Models EM Mezzo

**Medie marginali stimate dei modelli montati.** Questo gruppo consente di richiedere i mezzi marginali preventivi previsti dalla variabile dipendente nelle celle e i relativi errori standard per i fattori specificati. Inoltre, è possibile richiedere che i livelli di fattore degli effetti principali siano confrontati.

- **Factor (s) e Interazioni di Fattore.** Questo elenco contiene fattori e interazioni dei fattori che sono stati specificati nella finestra di dialogo Fixed, più un termine OVERALL . I termini modello costruiti da covariate sono esclusi da questo elenco.
- **Display per.** La procedura calcolerà i mezzi marginali stimati per i fattori e le interazioni dei fattori selezionati in questo elenco. Se viene selezionato OVERALL , vengono visualizzati i mezzi marginali stimati della variabile dipendente, collassando su tutti i fattori. Si noti che eventuali fattori selezionati o

le interazioni dei fattori rimangono selezionati a meno che una variabile associata non sia stata rimossa dall'elenco dei Fattori nella finestra di dialogo principale.

- **Confronta effetti principali.** Questa opzione consente di richiedere confronti pairwise dei livelli di effetti principali selezionati. L'Aggiustamento Intervallo di confidenza consente di applicare una regolazione agli intervalli di confidenza e valori di significati per rendere conto di più confronti. I metodi disponibili sono LSD (no regolazione), Bonferroni e Sidak. Infine, per ogni fattore, è possibile selezionare una categoria di riferimento a cui si effettuano i confronti. Se non viene selezionata alcuna categoria di riferimento, verranno costruiti tutti i confronti pairwise. Le opzioni per la categoria di riferimento sono prime, ultime o personalizzate (in tal caso si entra nel valore della categoria di riferimento).

## Modelli misti lineari Salva

Questa finestra di dialogo consente di salvare vari risultati del modello sul file di lavoro.

**Valori previsti fissi.** Salva le variabili relative alla regressione significa senza gli effetti.

- **Valori previsti.** La regressione significa senza gli effetti casuali.
- **Errori standard.** Gli errori standard delle stime.
- **Gradi di libertà.** I gradi di libertà associati alle stime.

**Valori previsti & residui.** Salva le variabili relative al valore montato del modello.

- **Valori previsti.** Il modello di valore montato.
- **Errori standard.** Gli errori standard delle stime.
- **Gradi di libertà.** I gradi di libertà associati alle stime.
- **Residui.** Il valore dei dati meno il valore previsto.

## Modelli misti lineari - Esporta

Questa finestra di dialogo consente di esportare il contenuto delle tabelle di output di EBLUP (Empirical Best Linear Unbiased Predictions) in dataset o file .sav . Il pulsante Esporta nella finestra di dialogo Modelli misti lineari è abilitato se viene specificato almeno un effetto casuale tramite la finestra di dialogo Casuale e la casella di controllo "Visualizza previsioni di parametri per questo blocco" è selezionata.

### Esporta EBLUPS

Selezionare una destinazione: **Dataset** o **File di dati**

Facoltativamente, fornire un nome.

Se sono stati specificati più effetti casuali con EBLUP prodotti, deselezionare la casella di spunta per visualizzare ogni tabella di risultati in un dataset o file separato.

## MIXED Comando Funzioni aggiuntive

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).
- Calcola medie marginali stimate per valori specificati di covariate (utilizzando la parola chiave WITH del sottocomando EMMEANS ).
- Confronta semplici effetti principali delle interazioni (utilizzando il sottocomando EMMEANS ).

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

## 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. Grazie alla formulazione estremamente generale del modello, copre i modelli statistici utilizzati di frequente, quali la regressione lineare per le risposte distribuite normalmente, i modelli logistici per i dati binari, i modelli log-lineari per i dati dei conteggi, i modelli doppi logaritmici complementari per i dati di sopravvivenza censurati per intervallo, oltre a molti altri modelli statistici.

**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 della richiesta.

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.

Considerazioni Sui Modelli Lineari Generalizzati

**Dati.** La risposta può essere scala, conta, binario o eventi - in - trial. I fattori sono assunti per essere categoriali. Si ipotizzano le covariate, il peso scala e lo scostamento.

**Ipotesi.** Si ipotizza che i casi siano osservazioni indipendenti.

Per Ottenere un Modello Linear Generalizzato

Dai menu, scegliere:

**Analizzare > Modelli lineari generalizzati > Modelli lineari generalizzati ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Specificare una funzione di distribuzione e di collegamento (vedi sotto per i dettagli sulle varie opzioni).
2. Nella scheda Risposta , selezionare una variabile dipendente.
3. Nella scheda Predittori , selezionare i fattori e le covariate da utilizzare nella previsione della variabile dipendente.
4. Nella scheda Modello , specificare gli effetti del modello utilizzando i fattori selezionati e le covariate.

La scheda Tipo di modello consente di specificare la funzione di distribuzione e collegamento per il proprio modello, fornendo brevi tagli per diversi modelli comuni categorizzati per tipo di risposta.

Tipi di modello

**Risposta di scala.** Sono disponibili le seguenti opzioni:

- **Lineare.** Specifica Normale come la distribuzione e l'Identità come funzione di collegamento.
- **Gamma con collegamento log.** Specifica Gamma come distribuzione e Accedi come funzione di collegamento.

**Risposta ordinaria.** Sono disponibili le seguenti opzioni:

- **logistico Ordinale.** Specifica il multinomiale (ordinale) come la distribuzione e la logit cumulativa come funzione di collegamento.
- **probit ordinaria.** Specifica il multinomiale (ordinale) come la distribuzione e la Probit cumulativa come funzione di collegamento.

**Conteggi.** Sono disponibili le seguenti opzioni:

- **Poisson loglineare.** Specifica Poisson come distribuzione e Accedi come funzione di collegamento.
- **Binomiale negativo con collegamento log.** Specifica Negativo binomiale (con un valore di 1 per il parametro accessorio) come distribuzione e Accedi come funzione di collegamento. Per avere la procedura stima il valore del parametro accessorio, specificare un modello personalizzato con distribuzione binomiale negativa e selezionare **Valore stimato** nel gruppo parametri.

**Risposta binaria o Eventi / Trials Data.** Sono disponibili le seguenti opzioni:

- **Binario logistico.** Specifica Binomiale come distribuzione e Logit come funzione di collegamento.
- **Probit binaria.** Specifica Binomiale come distribuzione e Probit come funzione di collegamento.
- **Sopravvivenza di censura per intervallo.** Specifica Binomiale come la distribuzione e il log - log integrativo come funzione di collegamento.

**Miscela.** Sono disponibili le seguenti opzioni:

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

**Personalizzato.** Specificare la propria combinazione di funzione di distribuzione e collegamento.

Distribuzione

Questa selezione specifica la distribuzione della variabile dipendente. La possibilità di specificare una distribuzione non normale e una funzione di legame di non identità è il vantaggio essenziale del modello lineare generalizzato rispetto al modello lineare 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 adatta solo alle variabili che rappresentano una risposta o un numero di eventi binario.
- **Gamma.** Questa distribuzione è adatta alle variabili con valori di scala positivi asimmetrici verso valori positivi superiori. 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 adatta alle variabili con valori di scala positivi asimmetrici verso valori positivi superiori. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Binomiale negativa.** Questa distribuzione può essere vista come il numero di prove necessario per osservare  $k$  esiti positivi ed è adatta alle 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 accessorio della distribuzione binomiale negativa può essere qualsiasi numero maggiore o uguale a 0; è possibile impostarlo su un valore fisso o consentire di stimarlo con la procedura. Quando il parametro ausiliario è impostato su 0, l'utilizzo di questa distribuzione equivale all'utilizzo della distribuzione di Poisson.
- **Normale.** È adatta alle variabili di scala i cui valori si distribuiscono assumendo una forma simmetrica a campana intorno a un valore centrale (media). La variabile dipendente deve essere un valore numerico.
- **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 appropriata per le variabili che possono essere rappresentate da combinazioni di Poisson di distribuzioni gamma; la distribuzione risulta "mista" nel senso che combina proprietà di distribuzioni continue (accetta valori reali non negativi) e distribuzioni discrete (massa di probabilità positiva a un unico valore, 0). La variabile dipendente deve essere numerica, con valori di dati maggiori o uguali a zero. Se il valore di un dato è minore di zero o mancante, il caso corrispondente non viene utilizzato 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 appropriata per variabili che rappresentano una risposta ordinale. La variabile dipendente può essere un numero o una stringa e deve avere almeno due valori di dati validi distinti.

Funzioni di collegamento

La funzione di legame è una trasformazione della variabile dipendente che consente la stima del modello. Sono disponibili le seguenti funzioni:

- **Identità.**  $f(x)=x$ . La variabile dipendente non viene trasformata. Questo legame può essere utilizzato con qualsiasi distribuzione.
- **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. È adatta solo con la distribuzione multinomiale.
- **Log-log complementare cumulativa.**  $f(x)=\ln(-\ln(1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Logit cumulativa.**  $f(x)=\ln(x / (1-x))$ , applicata alla probabilità cumulativa di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Log-log negativa cumulativa.**  $f(x)=-\ln(-\ln(x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Probit cumulativa.**  $f(x)=\Phi^{-1}(x)$ , applicata alla probabilità cumulativa di ciascuna categoria della risposta, dove  $\Phi^{-1}$  è la funzione di distribuzione cumulativa normale standard inversa. È adatta solo con la distribuzione multinomiale.
- **Log.**  $f(x)=\log(x)$ . Questo legame può essere utilizzato con qualsiasi distribuzione.
- **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. È adatta solo con la distribuzione binomiale negativa.
- **Log-log negativa.**  $f(x)=-\log(-\log(x))$ . 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 cumulativa 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. Questo legame può essere utilizzato con qualsiasi distribuzione.

## Risposta Modelli Lineari Generalizzati

In molti casi, è possibile specificare semplicemente una variabile dipendente; tuttavia, le variabili che assumono solo due valori e risposte che registrano eventi in trial richiedono un'attenzione extra.

- **Risposta binario.** Quando la variabile dipendente assume solo due valori, è possibile specificare la categoria di riferimento per la stima dei parametri. Una variabile di risposta binaria può essere stringa o numerica.
- **Numero di eventi che si verificano in una serie di prove.** Quando la risposta è un numero di eventi che si verificano in una serie di prove, la variabile dipendente contiene il numero di eventi e si può selezionare una variabile aggiuntiva contenente il numero di prove. In alternativa, se il numero delle prove è uguale per tutti i soggetti, le prove possono essere specificate mediante un valore fisso. Il numero di prove deve essere maggiore o uguale al numero di eventi per ogni caso. Gli eventi devono essere numeri interi non negativi, mentre le prove devono essere numeri interi positivi.

Per i modelli multinazionali ordinali è possibile specificare l'ordine di categoria della risposta: ascendente, decrescente o dati (ordine dei dati significa che il primo valore riscontrato nei dati definisce la prima categoria, l'ultimo valore riscontrato definisce l'ultima categoria).

**Peso 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.

## Generalized Linear Models Reference Categoria

Per la risposta binaria, è possibile scegliere la categoria di riferimento della variabile dipendente. Ciò può influire su alcuni output, quali le stime dei parametri e i valori salvati, ma non dovrebbe modificare l'adattamento del modello. Per esempio, se la risposta binaria prende i valori 0 e 1:

- Per default, la procedura rende l'ultima categoria (con valore più alto), o 1, la categoria di riferimento. In questa situazione, le probabilità salvate del modello stimano la possibilità che un dato caso prenda il valore 0 e le stime dei parametri devono essere interpretate come relative alla verosimiglianza di categoria 0.
- Se si specifica la prima categoria (con valore più basso), o 0, come categoria di riferimento, le probabilità salvate del modello stimano la possibilità che un dato caso prenda il valore 1.
- Se si specifica la categoria personalizzata e per la variabile sono definite delle etichette, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Ciò può essere utile quando, durante la specifica di un modello, non si ricorda esattamente come era codificata una variabile particolare.

## Predicatori Modelli Lineari Generalizzati

Scheda Predittori consente di specificare i fattori e le covariate utilizzate per costruire effetti di modello e di specificare uno scostamento opzionale.

**Fattori.** I fattori sono predittori categoriali; possono essere numerici o stringa.

**Covariati.** Le covariate sono predittori di scala; devono essere numeriche.

*Nota:* quando la risposta è binomiale con formato binario, la procedura calcola le statistiche di devianza e di bontà di tipo chi - quadrato da sottopopolazioni che si basano sulla classificazione incrociata dei valori osservati dei fattori selezionati e delle covariate. Si deve tenere la stessa serie di predittori attraverso molteplici esecuzioni della procedura per garantire un numero consistente di sottopopolazioni.

**Offset.** Il termine offset è un predittore "strutturale". Il suo coefficiente non è stimato dal modello ma si presume che abbia il valore 1; pertanto, i valori dell'offset vengono semplicemente aggiunti al predittore lineare dell'obiettivo. Ciò è particolarmente utile nei modelli di regressione di Poisson, nei quali ogni caso può avere diversi livelli di esposizione all'evento di interesse.

Per esempio, nella modellazione dei tassi di incidente per singolo conducente esiste una differenza significativa tra un autista responsabile di un incidente in tre anni di esperienza e un autista responsabile di un incidente in 25 anni. Il numero di incidenti può essere rappresentato come una risposta di Poisson o binomiale negativa con un collegamento log se il logaritmo naturale dell'esperienza del conducente viene incluso come termine di offset.

Altre combinazioni dei tipi di distribuzione e di collegamento richiederebbero altre trasformazioni della variabile di offset.

## Opzioni Modelli Lineari Generalizzati

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

**Valori mancanti definiti dall'utente.** I fattori devono avere valori validi per un caso da inserire nell'analisi. Questi controlli consentono di decidere se i valori mancanti degli utenti sono trattati come validi tra le variabili del fattore.

**Ordine di categoria.** Ciò è rilevante per determinare l'ultimo livello di un fattore, che può essere associato ad un parametro ridondante nell'algoritmo di stima. La modifica dell'ordine di categoria può modificare i valori degli effetti a livello di fattore, poiché queste stime dei parametri sono calcolate rispetto al livello "ultimo". I fattori possono essere ordinati in ordine crescente dal valore più basso al più alto, in ordine decrescente dal valore più alto al valore più basso, oppure in "ordine dati". Ciò significa che il primo valore riscontrato nei dati definisce la prima categoria e l'ultimo valore univoco riscontrato definisce l'ultima categoria.

## Modello Modelli Lineari Generalizzati

### Specifica effetti del modello

Il modello predefinito è *intercept - only*, quindi è necessario specificare esplicitamente altri effetti di modello. In alternativa, è possibile costruire 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

Crea il termine di interazione ad alto livello per tutte le variabili selezionate.

#### Fattoriale

Crea 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. Per esempio, una catena di supermercati può seguire le abitudini di spesa dei propri clienti in più negozi. Poiché ogni cliente frequenta un solo negozio, l'effetto *Cliente* può definirsi **nidificato** all'interno dell'effetto *Negozi*.

È inoltre possibile includere effetti di interazione, ad esempio termini polinomiali che interessano la stessa covariata, 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. Di conseguenza, se  $A$  è un fattore, non è possibile specificare  $A*A$ .
- Tutti i fattori compresi in un effetto nidificato devono essere univoci. Di conseguenza, se  $A$  è un fattore, non è possibile specificare  $A(A)$ .
- Nessun effetto può essere nidificato all'interno di una covariata. Di conseguenza, se  $A$  è un fattore e  $X$  è una covariata, non è possibile specificare  $A(X)$ .

## Stima Modelli Lineari Generalizzati

**Stima dei parametri.** I controlli in questo gruppo consentono di specificare i metodi di stima e di fornire valori iniziali per le stime dei parametri.

- **Metodo.** È possibile selezionare un metodo di stima dei parametri. Scegliere tra Newton-Raphson, calcolo del punteggio di Fisher o un metodo ibrido in cui le iterazioni del calcolo del punteggio di Fisher vengono eseguite prima di passare al metodo di Newton-Raphson. Se durante la fase di Fisher-scoring del metodo ibrido, prima del raggiungimento del numero massimo di iterazioni di Fisher, viene raggiunta la convergenza, l'algoritmo continua con il metodo di Newton-Raphson.
- **Metodo del parametro di scala.** È possibile selezionare il metodo di stima del parametro di 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 di devianza e chi-quadrato di Pearson stimano il parametro di scala a partire dal valore di queste statistiche. In alternativa, è possibile specificare un valore fisso per il parametro di scala.

- **Valori iniziali.** La procedura calcola automaticamente i valori iniziali dei parametri. In alternativa, è possibile specificare i valori iniziali per le stime dei parametri.

**Matrice di covarianza.** Lo stimatore basato su modello è il negativo dell'inversa generalizzata 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.** Sono disponibili le seguenti opzioni:

- **Numero massimo di iterazioni.** Numero massimo di iterazioni eseguite dall'algoritmo. Specificare un intero non negativo.
- **Max dimezzamenti.** Ad ogni iterazione, la dimensione di passo viene ridotta di un fattore di 0,5 finché la verosimiglianza log non aumenta o non si raggiunge il numero massimo di dimezzamenti. Specificare un intero positivo.
- **Verificare la separazione dei punti dati.** Quando è selezionato, l'algoritmo esegue dei test per garantire che le stime dei parametri abbiano valori univoci. La separazione si verifica quando la procedura può generare un modello che classifica correttamente ogni caso. Questa opzione è disponibile per risposte multinomiali e risposte binomiali con formato binario.

**Criteri di convergenza.** Sono disponibili le seguenti opzioni

- **Convergenza parametri.** Quando è selezionato, l'algoritmo si arresta dopo un'iterazione nella quale la variazione assoluta o relativa nelle stime dei parametri è minore del valore specificato, che deve essere positivo.
- **Convergenza logaritmo della verosimiglianza.** Quando è selezionato, l'algoritmo si arresta dopo un'iterazione nella quale la variazione assoluta o relativa nella funzione di verosimiglianza log è minore del valore specificato, che deve essere positivo.
- **Convergenza hessiana.** Per la specifica assoluta, si presume la convergenza se una statistica basata sulla convergenza hessiana è minore del valore positivo specificato. Per la specifica relativa, si presume la convergenza se la statistica è minore del prodotto tra il valore positivo specificato e il valore assoluto della verosimiglianza log.

**Tolleranza della singolarità.** Le matrici singolari (o non invertibili) dispongono di colonne dipendenti in modo lineare, che possono causare gravi problemi all'algoritmo di stima. Anche le matrici quasi singolari possono generare risultati imprecisi, pertanto la procedura tratta come singolare qualsiasi matrice il cui determinante sia inferiore alla tolleranza. Specificare un valore positivo.

## Valori Iniziali Modelli generalizzati

Se vengono specificati valori iniziali, devono essere forniti per tutti i parametri (inclusi i parametri ridondanti) nel modello. Nel dataset, l'ordine delle variabili da sinistra a destra deve essere: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., dove *RowType\_* e *VarName\_* sono variabili stringa e *P1*, *P2*, ... sono variabili numeriche corrispondenti a un elenco ordinato di parametri.

- I valori iniziali vengono forniti su un record con valore *EST* per la variabile *RowType\_*; i valori iniziali effettivi sono indicati sotto le variabili *P1*, *P2*, .... La procedura ignora tutti i record per i quali *RowType\_* ha un valore diverso da *EST* e tutti i record oltre la prima ricorrenza di *RowType\_* uguale a *EST*.
- L'intercettazione, se inclusa nel modello, o 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 Split File è in vigore, allora le variabili devono iniziare con la variabile split - file o variabili nell'ordine specificato quando si crea il File di Split, seguito da *RowType\_*, *VarName\_*, *P1*, *P2*, ... come sopra. Le suddivise devono avvenire nel dataset specificato nello stesso ordine del dataset originale.

*Nota:* i nomi variabili  $P1, P2, \dots$  non sono richiesti; la procedura accetterà qualsiasi nome variabile valido per i parametri in quanto la mappatura delle variabili ai parametri è basata sulla variabile posizione, non nome variabile. Eventuali variabili oltre l'ultimo parametro vengono ignorate.

La struttura dei file per i valori iniziali è la stessa utilizzata quando si esportano il modello come dati; quindi, è possibile utilizzare i valori finali da un'esecuzione della procedura come input in una successiva esecuzione.

## Statistiche Modelli Lineari Generalizzati

**Effetti del modello.** Sono disponibili le seguenti opzioni:

- **Tipo di analisi.** Specificare il tipo di analisi da eseguire. L'analisi di tipo I è generalmente appropriata quando si hanno dei motivi a priori per ordinare i predittori nel modello, mentre l'analisi di tipo III trova un'applicazione più generale. Le statistiche di Wald o rapporti di verosimiglianza sono calcolati in base alla selezione nel gruppo di statistiche chi-quadrato.
- **Intervalli di confidenza.** Specificare un livello di confidenza maggiore di 50 e minore di 100. Gli intervalli di wald si basano sul presupposto che i parametri hanno una distribuzione normale asintotica; gli intervalli di probabilità del profilo sono più accurati ma possono essere computazionalmente costosi. Il livello di tolleranza per gli intervalli di verosimiglianza dei profili è il criterio usato per arrestare l'algoritmo iterativo utilizzato per calcolare gli intervalli.
- **Funzione di verosimiglianza logaritmica.** Controlla il formato di visualizzazione della funzione di verosimiglianza log. La funzione completa include un termine aggiuntivo costante rispetto alla stima dei parametri; non ha effetti sulla stima dei parametri ed è escluso dalla visualizzazione in alcuni prodotti software.

**Stampa.** È disponibile il seguente output:

- **Riepilogo elaborazione casi.** Visualizza il numero e la percentuale dei casi inclusi ed esclusi dall'analisi e la tabella di riepilogo dei dati correlati.
- **Statistiche descrittive.** Visualizza le statistiche descrittive e le informazioni di riepilogo su variabile dipendente, covariate e fattori.
- **Informazioni sul modello.** Visualizza il nome dell'insieme di dati, la variabile dipendente o le variabili eventi e prove, la variabile offset, la variabile peso di scala, la distribuzione della probabilità e la funzione di legame.
- **Statistiche della bontà di adattamento.** Visualizza la devianza e la devianza scalata, il chi-quadrato di Pearson e il chi-quadrato di Pearson scalato, la log-verosimiglianza, il criterio di informazione di Akaike (AIC), l'AIC corretto in campioni finiti (AICC), il criterio di informazione bayesiano (BIC) e l'AIC coerente (CAIC).
- **Statistiche di riepilogo del modello.** Visualizza i test di adattamento del modello, incluse le statistiche del rapporto di verosimiglianza per il test omnibus di adattamento del modello e le statistiche per i contrasti di tipo I o III di ogni effetto.
- **Stime dei parametri.** Visualizza le stime dei parametri e le statistiche di test e gli intervalli di confidenza corrispondenti. Se lo si desidera, è possibile visualizzare le stime dei parametri esponenziati oltre alle stime dei parametri grezzi.
- **Matrice di covarianza per le stime dei parametri.** Visualizza la matrice di covarianza dei parametri stimati.
- **Matrice di correlazione per le stime dei parametri.** Visualizza la matrice di correlazione dei parametri stimati.
- **Matrici del coefficiente di contrasto (L).** Visualizza i coefficienti di contrasto degli effetti predefiniti e delle medie marginali stimate, se richiesto nella scheda Medie marginali.
- **Funzioni stimabili generali.** Visualizza le matrici per la generazione delle matrici del coefficiente di contrasto (L).
- **Cronologia iterazioni.** Visualizza la cronologia delle iterazioni per le stime dei parametri e la verosimiglianza log e stampa l'ultima valutazione del vettore gradiente e della matrice hessiana. La tabella della cronologia delle iterazioni visualizza le stime dei parametri per per ciascuna iterazione  $n$

esima che inizia con l'iterazione 0<sup>esima</sup> (stime iniziali), dove  $n$  è il valore dell'intervallo di stampa. Se è richiesta la cronologia delle iterazioni, l'ultima iterazione viene sempre visualizzata indipendentemente da  $n$ .

- **Prova moltiplicatore di Lagrange test di scala o parametro accessorio binomiale negativo.** Visualizza le statistiche di prova del moltiplicatore di Lagrange per valutare la validità di un parametro di scala che viene calcolato utilizzando la devianza o la Pearson chi - quadrato o impostata ad un numero fisso, per le distribuzioni normali, gamma, inversa gaussiane e Tweedie. Per la distribuzione binomiale negativa, viene verificato il parametro ausiliario fisso.

## Modelli Lineari Generalizzati EM Mezzi

Questa scheda consente di visualizzare le medie marginali stimate per i livelli di fattori e le interazioni dei fattori. È anche possibile richiedere che venga visualizzata la media stimata complessiva. I mezzi marginali stimati non sono disponibili per i modelli multinazionali ordinali.

**Fattori e Interazioni.** Questo elenco contiene i fattori specificati nella scheda Predittori e le interazioni dei fattori specificati nella scheda Modello. Le covariate sono escluse da questa lista. I termini possono essere selezionati direttamente da questo elenco o combinati in un termine di interazione utilizzando il pulsante **By \***.

**Display Mezzi Per.** Le medie stimate vengono calcolate per i fattori e le interazioni dei fattori selezionati. Il contrasto determina la modalità di impostazione dei test di ipotesi per il confronto delle medie stimate. Il semplice contrasto richiede una categoria di riferimento o un livello di fattore rispetto al quale si confronta gli altri.

- **A coppie.** I confronti pairwise vengono calcolati per combinazioni di tutti i livelli dei fattori specificati o impliciti. Questo 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.** Ogni livello del fattore viene confrontato con il grand mean. I contrasti di deviazione non sono ortogonali.
- *Differenza.* Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Sono a volte definiti contrasti inversi 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 stimare le tendenze polinomiali.

**Scala.** I mezzi marginali stimati possono essere calcolati per la risposta, in base alla scala originale della variabile dipendente o per il predicatore lineare, in base alla variabile dipendente come trasformata dalla funzione di collegamento.

**Regolazione per Confronti multipli.** Quando si eseguono test di ipotesi con più contrasti, il livello di significatività generale può essere adattato in base ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di regolazione.

- **Differenza meno significativa.** Questo metodo non controlla la probabilità generale di rifiuto delle ipotesi che alcuni contrasti lineari siano diversi dai valori di ipotesi null.
- *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 decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.

## Modelli lineari generalizzati Salva

Gli articoli controllati vengono salvati con il nome specificato; è possibile scegliere di sovrascrivere le variabili esistenti con lo stesso nome delle nuove variabili o evitare i conflitti di nome tramite suffissi appendici per rendere unici i nuovi nomi variabili.

### Valore previsto della media della risposta

Salva i valori previsti modello per ogni caso nella metrica di risposta originale. Quando la distribuzione di risposta è binomiale e la variabile dipendente è binaria, la procedura salva le probabilità previste. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Cumulative probabilità prevista** e la procedura salva la probabilità prevista cumulativa per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

### Limite inferiore dell'intervallo di confidenza per la media della risposta

Salva il limite inferiore dell'intervallo di confidenza per la media della risposta. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Bassa di intervallo di confidenza per la probabilità prevista cumulativa** e la procedura salva il limite inferiore per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

### Limite superiore dell'intervallo di confidenza per la media della risposta

Salva il limite superiore dell'intervallo di confidenza per la media della risposta. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Upper bound of confidence interval per probabilità cumulativa** e la procedura salva il limite superiore per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

### Categoria prevista

Per i modelli con distribuzione binomiale e variabile dipendente binario, o distribuzione multinomiale, questo salva la categoria di risposta prevista per ogni caso. Questa opzione non è disponibile per altre distribuzioni di risposta.

### Valore previsto del predittore lineare

Salva i valori previsti dal modello per ogni caso nella metrica del predittore lineare (risposta trasformata tramite la funzione di collegamento specificato). Quando la distribuzione di risposta è multinomiale, la procedura salva il valore previsto per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

### Errore standard stimato del valore previsto del predittore lineare

Quando la distribuzione di risposta è multinomiale, la procedura salva l'errore standard stimato per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

I seguenti elementi non sono disponibili quando la distribuzione di 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. Un valore alto del D di Cook indica che l'esclusione di un caso dal calcolo delle statistiche di regressione modifica sostanzialmente i coefficienti.

### Valore di leva

Misura l'influenza di un punto sull'adattamento della regressione. La leva centrata è compresa tra 0 (nessuna influenza sull'adattamento) e  $(N-1)/N$ .

### Residuo

La differenza tra un valore osservato e il valore stimato dal modello.

### Residuo di Pearson

La radice quadrata del contributo di un caso alla statistica di Pearson chi - quadrato, con il segno della residua residua.

### Residuo di Pearson standardizzato

Il residuo Pearson moltiplicato per la radice quadrata dell'inverso del prodotto del parametro di scala e  $1 - \text{leverage}$  per il caso.

### Residuo di devianza

La radice quadrata del contributo di un caso alla statistica Devianza, con il segno della residua residua.

### Residuo di devianza standardizzato

Il Devianza residuo moltiplicato per la radice quadrata dell'inverso del prodotto del parametro di scala e  $1 - \text{leverage}$  per il caso.

### Residuo della verosimiglianza

La radice quadrata di una media ponderata (basata sulla leva del caso) delle piazze del Pearson standardizzato e dei residui di Deviance standardizzati, con il segno della residua residua.

## Esportazione Modelli Lineari Generalizzati

**Modello di esportazione come dati.** Scrive un dataset in formato IBM SPSS Statistica contenente la matrice di correlazione o la matrice di covarianza con le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file di matrice è il seguente.

- **Split variabili.** Se utilizzato, eventuali variabili che definiscono la suddivisione.
- **RowType\_.** Prende i valori (e le etichette dei valori) *COV* (covarianze), *CORR* (correlazioni), *EST* (stime dei parametri), *SE* (errori standard), *SIG* (livelli di significatività) e *DF* (gradi di libertà del disegno di campionamento). Esiste un caso separato con tipo di riga *COV* (o *CORR*) per ogni parametro modello, più un caso separato per ciascuno degli altri tipi di riga.
- **VarName\_.** Assume valori *P1, P2, ...*, corrispondente ad un elenco ordinato di tutti i parametri di modello stimati (ad eccezione dei parametri binomiali di scala o negativi), per i tipi di righe *COV* o *CORR*, con etichette di valore corrispondenti alle stringhe di parametri mostrate nella tabella delle stime dei parametri. Le celle sono vuote per altri tipi di righe.
- **P1, P2, ...** Queste variabili corrispondono ad un elenco ordinato di tutti i parametri del modello (inclusa la scala e i parametri binomiali negativi, a seconda dei casi), con etichette variabili corrispondenti alle stringhe di parametro mostrate nella tabella delle stime dei parametri e assumono valori in base al tipo di riga.

Per i parametri ridondanti, tutte le covarianze sono impostate a zero, le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate a zero; e tutti gli errori standard, i livelli di significatività e i gradi residui di libertà 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 è stimato tramite la massima verosimiglianza, viene dato l'errore standard; altrimenti 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 probabilità, viene dato l'errore standard; altrimenti viene impostato sul valore mancante di sistema.

Se ci sono le divise, allora l'elenco dei parametri deve essere accumulato in tutte le divise. In una determinata suddivisione, alcuni parametri possono essere irrilevanti; questo non è uguale a quello ridondante. Per parametri irrilevanti, tutte le covarianze o correlazioni, le stime dei parametri, gli errori standard, i livelli di significatività e i gradi di libertà sono impostati sul valore mancante di sistema.

È possibile utilizzare questo file di matrice come valori iniziali per un'ulteriore stima del modello; notare che questo file non è immediatamente utilizzabile per ulteriori analisi in altre procedure che leggono un file di matrice a meno che tali procedure non accettino tutti i tipi di riga esportati qui. Anche allora, bisogna fare attenzione che tutti i parametri presenti in questo file di matrice abbiano lo stesso significato per la procedura di lettura del file.

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

## Funzioni Aggiuntive di comando GENLIN

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i valori iniziali per le stime dei parametri come elenco dei numeri (utilizzando il comando CRITERIA).
- Fissare covariate a valori diversi dai loro mezzi quando si calcolano i mezzi marginali stimati (utilizzando il comando EMMEANS).
- Specificare contrasti polinomiali personalizzati per i mezzi marginali stimati (utilizzando il comando EMMEANS).
- Specificare un sottoinsieme dei fattori per i quali vengono visualizzati i mezzi marginali stimati da confrontare 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 Generalized Estimating Equations estende il modello lineare generalizzato per consentire l'analisi di misurazioni ripetute o altre osservazioni correlate, come ad esempio i dati di clustered.

**Esempio.** I funzionari della Sanità Pubblica possono utilizzare le equazioni di stima generalizzate per adattarsi a misure ripetute di regressione logistica per studiare gli effetti dell'inquinamento dell'aria sui bambini.

Generalizzazioni Stimate Equazioni Dati Considerazioni

**Dati.** La risposta può essere scala, conta, binario o eventi - in - trial. I fattori sono assunti per essere categoriali. Si ipotizzano le covariate, il peso scala e lo scostamento. Le variabili utilizzate per definire i soggetti o le misurazioni ripetute all'interno del soggetto non possono essere utilizzate per definire la risposta ma possono servire altri ruoli nel modello.

**Ipotesi.** Si ipotizza che i casi siano dipendenti all'interno di soggetti e indipendenti tra i soggetti. La matrice di correlazione che rappresenta le dipendenze all'interno è stimata come parte del modello.

Ottenimento di Equazioni Generalizzate

Dai menu, scegliere:

**Analizzare > Modelli lineari generalizzati > Equazioni di stima generalizzate ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare una o più variabili di soggetto (vedi sotto per ulteriori opzioni).

La combinazione di valori delle variabili specificate dovrebbe definire univocamente **soggetti** all'interno del dataset. Ad esempio, una singola variabile *ID paziente* dovrebbe essere sufficiente per definire i soggetti in un singolo ospedale, ma la combinazione di *ID ospedaliero* e *ID paziente* può essere necessaria se i numeri di identificazione del paziente non sono univocamente negli ospedali. In un'impostazione di misure ripetute vengono registrate più osservazioni per ogni soggetto, quindi ogni soggetto può occupare più casi nel dataset.

2. Nella scheda Tipo di modello, specificare una funzione di distribuzione e collegamento.
3. Nella scheda Risposta, selezionare una variabile dipendente.
4. Nella scheda Predittori, selezionare i fattori e le covariate da utilizzare nella previsione della variabile dipendente.
5. Nella scheda Modello, specificare gli effetti del modello utilizzando i fattori selezionati e le covariate.

Facoltativamente, nella scheda Repeated è possibile specificare:

**Variabili entro - soggetto.** La combinazione di valori delle variabili all'interno definisce l'ordinamento delle misurazioni all'interno dei soggetti; quindi, la combinazione di variabili di soggetto e di soggetto

definisce univocamente ogni misurazione. Ad esempio, la combinazione di *Periodo*, *ID ospedaliero* e *ID paziente* definisce, per ogni caso, una particolare visita d'ufficio per un determinato paziente all'interno di un determinato ospedale.

Se il dataset è già ordinato in modo che le misurazioni ripetute di ogni soggetto si verifichino in un blocco contiguo di casi e nel corretto ordine, non è strettamente necessario specificare una variabile all'interno dei soggetti ed è possibile deselezionare **Ordina i casi per soggetto e variabili all'interno** e salvare il tempo di elaborazione richiesto per eseguire il (temporaneo) ordinamento. Generalmente è consigliabile fare uso di variabili all'interno del soggetto per garantire un corretto ordinamento delle misurazioni.

Le variabili soggette e all'interno non possono essere utilizzate per definire la risposta, ma possono svolgere altre funzioni nel modello. Ad esempio, *ID Hospital* potrebbe essere utilizzato come fattore nel modello.

**Matrice di covarianza.** Lo stimatore basato su modello è il negativo dell'inversa generalizzata della matrice hessiana. Il robusto stimatore (chiamato anche stimatore Huber / White/sandwich) è un stimatore basato su modello "corretto" che fornisce una stima coerente della covarianza, anche quando la matrice di correlazione di lavoro è missata. Questa specifica si applica ai parametri nel modello lineare parte delle equazioni di stima generalizzate, mentre la specifica sulla scheda Stima si applica solo al modello lineare generalizzato iniziale.

**Matrice di correlazione di lavoro.** Questa matrice di correlazione rappresenta le dipendenze all'interno. La sua dimensione è determinata dal numero di misurazioni e quindi dalla combinazione di valori di variabili all'interno. È possibile specificare una delle seguenti strutture:

- **Indipendente.** Le misurazioni ripetute sono non correlate.
- **AR (1).** Le misurazioni ripetute hanno una relazione autoregressiva di primo ordine. La correlazione tra due elementi è uguale a  $\rho$  per gli elementi adiacenti,  $\rho^2$  per gli elementi separati da un terzo e così via. È vincolato in modo che  $-1 < \rho < 1$ .
- **Espmodificabile.** Questa struttura ha correlazioni omogenee tra elementi. È conosciuta anche come struttura di simmetria composta.
- **M-dipendente.** Le misurazioni consecutive hanno un coefficiente di correlazione comune, le coppie di misurazioni separate da un terzo hanno un coefficiente di correlazione comune e così via, attraverso coppie di misurazioni separate da  $m-1$  altre misurazioni. Ad esempio, se si sottopongono gli studenti a test standardizzati ogni anno dalla terza alla settima classe. Questa struttura presuppone che i punteggi della terza e quarta, quarta e quinta, quinta e sesta e sesta e settima classe abbiano la stessa correlazione; la terza e la quinta, la quarta e la sesta e la quinta e la settima avranno la stessa correlazione; la terza e la sesta e la quarta e la settima avranno la stessa correlazione. Le misurazioni con separazioni superiori a  $m$  sono ipotizzabili. Quando si sceglie questa struttura, specificare un valore di  $m$  inferiore all'ordine della matrice di correlazione di lavoro.
- **Non strutturato.** Si tratta di una matrice di correlazione completamente generale.

Per impostazione predefinita, la procedura adeguerà le stime di correlazione per numero di parametri non ridondanti. La rimozione di questa regolazione può essere auspicabile se si desidera che le stime siano invariante a modifiche di replica di livello soggetto.

- **Numero massimo di iterazioni.** Verrà eseguito il numero massimo di iterazioni l'algoritmo di equazioni stimate generalizzato. Specificare un intero non negativo. Questa specifica si applica ai parametri nel modello lineare parte delle equazioni di stima generalizzate, mentre la specifica sulla scheda Stima si applica solo al modello lineare generalizzato iniziale.
- **La matrice di aggiornamento.** Gli elementi nella matrice di correlazione di lavoro sono stimati in base alle stime dei parametri, che vengono aggiornati in ogni iterazione dell'algoritmo. Se la matrice di correlazione di lavoro non è affatto aggiornata, la matrice di correlazione di lavoro iniziale viene utilizzata per tutto il processo di stima. Se la matrice viene aggiornata, è possibile specificare l'intervallo di iterazione in cui aggiornare elementi di matrice di correlazione di lavoro. Specificando un valore superiore al 1 maggio si riduce il tempo di elaborazione.

**Criteri di convergenza.** Queste specifiche si applicano ai parametri nel modello lineare parte delle equazioni di stima generalizzate, mentre la specifica sulla scheda Stima si applica solo al modello lineare generalizzato iniziale.

- **Convergenza parametri.** Quando è selezionato, l'algoritmo si arresta dopo un'iterazione nella quale la variazione assoluta o relativa nelle stime dei parametri è minore del valore specificato, che deve essere positivo.
- **Convergenza hessiana.** Si ipotizza la convergenza se una statistica basata sull'Hessian è inferiore al valore specificato, che deve essere positivo.

## Equazioni di stima generalizzate Tipo di modello

La scheda Tipo di modello consente di specificare la funzione di distribuzione e collegamento per il proprio modello, fornendo collegamenti rapidi per diversi modelli comuni categorizzati per tipo di risposta.

Tipi di modello

**Risposta di scala.** Sono disponibili le seguenti opzioni:

- **Lineare.** Specifica Normale come la distribuzione e l'Identità come funzione di collegamento.
- **Gamma con collegamento log.** Specifica Gamma come distribuzione e Accedi come funzione di collegamento.

**Risposta ordinaria.** Sono disponibili le seguenti opzioni:

- **logistico Ordinale.** Specifica il multinomiale (ordinale) come la distribuzione e la logit cumulativa come funzione di collegamento.
- **probit ordinaria.** Specifica il multinomiale (ordinale) come la distribuzione e la Probit cumulativa come funzione di collegamento.

**Conteggi.** Sono disponibili le seguenti opzioni:

- **Poisson loglineare.** Specifica Poisson come distribuzione e Accedi come funzione di collegamento.
- **Binomiale negativo con collegamento log.** Specifica Negativo binomiale (con un valore di 1 per il parametro accessorio) come distribuzione e Accedi come funzione di collegamento. Per avere la procedura stima il valore del parametro accessorio, specificare un modello personalizzato con distribuzione binomiale negativa e selezionare **Valore stimato** nel gruppo parametri.

**Risposta binaria o Eventi / Trials Data.** Sono disponibili le seguenti opzioni:

- **Binario logistico.** Specifica Binomiale come distribuzione e Logit come funzione di collegamento.
- **Probit binaria.** Specifica Binomiale come distribuzione e Probit come funzione di collegamento.
- **Sopravvivenza di censura per intervallo.** Specifica Binomiale come la distribuzione e il log - log integrativo come funzione di collegamento.

**Miscela.** Sono disponibili le seguenti opzioni:

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

**Personalizzato.** Specificare la propria combinazione di funzione di distribuzione e collegamento.

Distribuzione

Questa selezione specifica la distribuzione della variabile dipendente. La possibilità di specificare una distribuzione non normale e una funzione di legame di non identità è il vantaggio essenziale del modello lineare generalizzato rispetto al modello lineare 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 adatta solo alle variabili che rappresentano una risposta o un numero di eventi binario.

- **Gamma.** Questa distribuzione è adatta alle variabili con valori di scala positivi asimmetrici verso valori positivi superiori. 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 adatta alle variabili con valori di scala positivi asimmetrici verso valori positivi superiori. Se il valore dei dati è inferiore o uguale a 0 o è mancante, il caso corrispondente non viene usato nell'analisi.
- **Binomiale negativa.** Questa distribuzione può essere vista come il numero di prove necessario per osservare  $k$  esiti positivi ed è adatta alle 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 accessorio della distribuzione binomiale negativa può essere qualsiasi numero maggiore o uguale a 0; è possibile impostarlo su un valore fisso o consentire di stimarlo con la procedura. Quando il parametro ausiliario è impostato su 0, l'utilizzo di questa distribuzione equivale all'utilizzo della distribuzione di Poisson.
- **Normale.** È adatta alle variabili di scala i cui valori si distribuiscono assumendo una forma simmetrica a campana intorno a un valore centrale (media). La variabile dipendente deve essere un valore numerico.
- **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 appropriata per le variabili che possono essere rappresentate da combinazioni di Poisson di distribuzioni gamma; la distribuzione risulta "mista" nel senso che combina proprietà di distribuzioni continue (accetta valori reali non negativi) e distribuzioni discrete (massa di probabilità positiva a un unico valore, 0). La variabile dipendente deve essere numerica, con valori di dati maggiori o uguali a zero. Se il valore di un dato è minore di zero o mancante, il caso corrispondente non viene utilizzato 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 appropriata per variabili che rappresentano una risposta ordinale. La variabile dipendente può essere un numero o una stringa e deve avere almeno due valori di dati validi distinti.

#### Funzione di collegamento

La funzione di legame è una trasformazione della variabile dipendente che consente la stima del modello. Sono disponibili le seguenti funzioni:

- **Identità.**  $f(x)=x$ . La variabile dipendente non viene trasformata. Questo legame può essere utilizzato con qualsiasi distribuzione.
- **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. È adatta solo con la distribuzione multinomiale.
- **Log-log complementare cumulativa.**  $f(x)=\ln(-\ln(1-x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Logit cumulativa.**  $f(x)=\ln(x / (1-x))$ , applicata alla probabilità cumulativa di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Log-log negativa cumulativa.**  $f(x)=-\ln(-\ln(x))$ , applicata alla probabilità cumulata di ciascuna categoria della risposta. È adatta solo con la distribuzione multinomiale.
- **Probit cumulativa.**  $f(x)=\Phi^{-1}(x)$ , applicata alla probabilità cumulativa di ciascuna categoria della risposta, dove  $\Phi^{-1}$  è la funzione di distribuzione cumulativa normale standard inversa. È adatta solo con la distribuzione multinomiale.
- **Log.**  $f(x)=\log(x)$ . Questo legame può essere utilizzato con qualsiasi distribuzione.
- **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. È adatta solo con la distribuzione binomiale negativa.
- **Log-log negativa.**  $f(x)=-\log(-\log(x))$ . 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 cumulativa 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. Questo legame può essere utilizzato con qualsiasi distribuzione.

## Stima delle Equazioni generalizzate

In molti casi, è possibile specificare semplicemente una variabile dipendente; tuttavia, le variabili che assumono solo due valori e risposte che registrano eventi in trial richiedono un'attenzione extra.

- **Risposta binario.** Quando la variabile dipendente assume solo due valori, è possibile specificare la categoria di riferimento per la stima dei parametri. Una variabile di risposta binaria può essere stringa o numerica.
- **Numero di eventi che si verificano in una serie di prove.** Quando la risposta è un numero di eventi che si verificano in una serie di prove, la variabile dipendente contiene il numero di eventi e si può selezionare una variabile aggiuntiva contenente il numero di prove. In alternativa, se il numero delle prove è uguale per tutti i soggetti, le prove possono essere specificate mediante un valore fisso. Il numero di prove deve essere maggiore o uguale al numero di eventi per ogni caso. Gli eventi devono essere numeri interi non negativi, mentre le prove devono essere numeri interi positivi.

Per i modelli multinazionali ordinali è possibile specificare l'ordine di categoria della risposta: ascendente, decrescente o dati (ordine dei dati significa che il primo valore riscontrato nei dati definisce la prima categoria, l'ultimo valore riscontrato definisce l'ultima categoria).

**Peso 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.

## Categorie di riferimento delle Equazioni generalizzate

Per la risposta binaria, è possibile scegliere la categoria di riferimento della variabile dipendente. Ciò può influire su alcuni output, quali le stime dei parametri e i valori salvati, ma non dovrebbe modificare l'adattamento del modello. Per esempio, se la risposta binaria prende i valori 0 e 1:

- Per default, la procedura rende l'ultima categoria (con valore più alto), o 1, la categoria di riferimento. In questa situazione, le probabilità salvate del modello stimano la possibilità che un dato caso prenda il valore 0 e le stime dei parametri devono essere interpretate come relative alla verosimiglianza di categoria 0.
- Se si specifica la prima categoria (con valore più basso), o 0, come categoria di riferimento, le probabilità salvate del modello stimano la possibilità che un dato caso prenda il valore 1.
- Se si specifica la categoria personalizzata e per la variabile sono definite delle etichette, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Ciò può essere utile quando, durante la specifica di un modello, non si ricorda esattamente come era codificata una variabile particolare.

## Predicatori Di Equazioni Stimate Generalizzati

Scheda Predittori consente di specificare i fattori e le covariate utilizzate per costruire effetti di modello e di specificare uno scostamento opzionale.

**Fattori.** I fattori sono predittori categoriali; possono essere numerici o stringa.

**Covariati.** Le covariate sono predittori di scala; devono essere numeriche.

*Nota:* quando la risposta è binomiale con formato binario, la procedura calcola le statistiche di devianza e di bontà di tipo chi - quadrato da sottopopolazioni che si basano sulla classificazione incrociata dei valori osservati dei fattori selezionati e delle covariate. Si deve tenere la stessa serie di predittori attraverso molteplici esecuzioni della procedura per garantire un numero consistente di sottopopolazioni.

**Offset.** Il termine offset è un predittore "strutturale". Il suo coefficiente non è stimato dal modello ma si presume che abbia il valore 1; pertanto, i valori dell'offset vengono semplicemente aggiunti al predittore lineare dell'obiettivo. Ciò è particolarmente utile nei modelli di regressione di Poisson, nei quali ogni caso può avere diversi livelli di esposizione all'evento di interesse.

Per esempio, nella modellazione dei tassi di incidente per singolo conducente esiste una differenza significativa tra un autista responsabile di un incidente in tre anni di esperienza e un autista responsabile di un incidente in 25 anni. Il numero di incidenti può essere rappresentato come una risposta di Poisson o binomiale negativa con un collegamento log se il logaritmo naturale dell'esperienza del conducente viene incluso come termine di offset.

Altre combinazioni dei tipi di distribuzione e di collegamento richiederebbero altre trasformazioni della variabile di offset.

## Opzioni di Equazioni stimate generalizzate

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

**Valori mancanti definiti dall'utente.** I fattori devono avere valori validi per un caso da inserire nell'analisi. Questi controlli consentono di decidere se i valori mancanti degli utenti sono trattati come validi tra le variabili del fattore.

**Ordine di categoria.** Ciò è rilevante per determinare l'ultimo livello di un fattore, che può essere associato ad un parametro ridondante nell'algoritmo di stima. La modifica dell'ordine di categoria può modificare i valori degli effetti a livello di fattore, poiché queste stime dei parametri sono calcolate rispetto al livello "ultimo". I fattori possono essere ordinati in ordine crescente dal valore più basso al più alto, in ordine decrescente dal valore più alto al valore più basso, oppure in "ordine dati". Ciò significa che il primo valore riscontrato nei dati definisce la prima categoria e l'ultimo valore univoco riscontrato definisce l'ultima categoria.

## Modello di Equazioni di stima generalizzata

**Specifica effetti del modello.** Il modello predefinito è intercept - only, quindi è necessario specificare esplicitamente altri effetti di modello. In alternativa, è possibile costruire 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.** Crea il termine di interazione ad alto livello per tutte le variabili selezionate.

**Factoriale.** Crea 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. Per esempio, una catena di supermercati può seguire le abitudini di spesa dei propri clienti in più

negozi. Poiché ogni cliente frequenta un solo negozio, l'effetto *Cliente* può definirsi **nidificato** all'interno dell'effetto *Negozio*.

Inoltre, è possibile includere effetti di interazione o aggiungere più livelli di nidificazione al termine nidificato.

**Limitazione.** I termini nidificati sono sottoposti alle seguenti restrizioni:

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

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

I modelli con la distribuzione ordinale multinomiale non hanno un unico termine di intercettazione; invece ci sono parametri di soglia che definiscono i punti di transizione tra le categorie adiacenti. Le soglie sono sempre incluse nel modello.

## Stima Delle Equazioni Di Stima Generalizzata

**Stima dei parametri.** I controlli in questo gruppo consentono di specificare i metodi di stima e di fornire valori iniziali per le stime dei parametri.

- **Metodo.** È possibile selezionare un metodo di stima dei parametri; scegliere tra Newton - Raphson, Fisher scoring o un metodo ibrido in cui vengono eseguite le iterazioni di punteggio Fisher prima di passare al metodo Newton - Raphson. Se durante la fase di Fisher-scoring del metodo ibrido, prima del raggiungimento del numero massimo di iterazioni di Fisher, viene raggiunta la convergenza, l'algoritmo continua con il metodo di Newton-Raphson.

- **Metodo del parametro di scala.** È possibile selezionare il metodo di stima del parametro di scala.

La massima verosimiglianza stima il parametro di scala insieme agli effetti del modello; si noti che questa opzione non è valida se la risposta ha una distribuzione binomiale negativa, di Poisson o binomiale. Poiché il concetto di verosimiglianza non entra in equazioni di stima generalizzate, questa specifica si applica solo al modello lineare generalizzato iniziale; questa stima dei parametri di scala viene poi passata alle equazioni di stima generalizzate, che aggiornano il parametro di scala da parte della Pearson chi - quadrato diviso per i suoi gradi di libertà.

Le opzioni di devianza e Pearson chi - quadrato stimano il parametro di scala dal valore di quelle statistiche nel modello lineare generalizzato iniziale; questa stima dei parametri di scala viene quindi passata alle equazioni di stima generalizzate, che lo trattano come fisso.

In alternativa, specificare un valore fisso per il parametro di scala. Sarà trattato come fisso nella stima del modello lineare generalizzato iniziale e delle 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 dei parametri.

Le iterazioni e i criteri di convergenza specificati su questa scheda sono applicabili solo al modello lineare generalizzato iniziale. Per i criteri di stima utilizzati nel montaggio delle equazioni di stima generalizzate, consultare la scheda Repeated.

**Iterazioni.** Sono disponibili le seguenti opzioni:

- **Numero massimo di iterazioni.** Numero massimo di iterazioni eseguite dall'algoritmo. Specificare un intero non negativo.
- **Max dimezzamenti.** Ad ogni iterazione, la dimensione di passo viene ridotta di un fattore di 0,5 finché la verosimiglianza log non aumenta o non si raggiunge il numero massimo di dimezzamenti. Specificare un intero positivo.

- **Verificare la separazione dei punti dati.** Quando è selezionato, l'algoritmo esegue dei test per garantire che le stime dei parametri abbiano valori univoci. La separazione si verifica quando la procedura può generare un modello che classifica correttamente ogni caso. Questa opzione è disponibile per risposte multinomiali e risposte binomiali con formato binario.

**Criteri di convergenza.** Sono disponibili le seguenti opzioni

- **Convergenza parametri.** Quando è selezionato, l'algoritmo si arresta dopo un'iterazione nella quale la variazione assoluta o relativa nelle stime dei parametri è minore del valore specificato, che deve essere positivo.
- **Convergenza logaritmo della verosimiglianza.** Quando è selezionato, l'algoritmo si arresta dopo un'iterazione nella quale la variazione assoluta o relativa nella funzione di verosimiglianza log è minore del valore specificato, che deve essere positivo.
- **Convergenza hessiana.** Per la specifica assoluta, si presume la convergenza se una statistica basata sulla convergenza hessiana è minore del valore positivo specificato. Per la specifica relativa, si presume la convergenza se la statistica è minore del prodotto tra il valore positivo specificato e il valore assoluto della verosimiglianza log.

**Tolleranza della singolarità.** Le matrici singolari (o non invertibili) dispongono di colonne dipendenti in modo lineare, che possono causare gravi problemi all'algoritmo di stima. Anche le matrici quasi singolari possono generare risultati imprecisi, pertanto la procedura tratta come singolare qualsiasi matrice il cui determinante sia inferiore alla tolleranza. Specificare un valore positivo.

## Equazioni di stima generalizzate Valori iniziali

La procedura stima un modello lineare generalizzato iniziale e le stime di questo modello vengono utilizzate come valori iniziali per le stime dei parametri nel modello lineare parte delle equazioni di stima generalizzate. I valori iniziali non sono necessari per la matrice di correlazione di lavoro perché gli elementi di matrice sono basati sulle stime dei parametri. I valori iniziali specificati su questa finestra di dialogo vengono utilizzati come punto di partenza per il modello lineare generalizzato iniziale, non le equazioni di stima generalizzate, a meno che le iterazioni massime sulla scheda Stima siano impostate a 0.

Se vengono specificati valori iniziali, devono essere forniti per tutti i parametri (inclusi i parametri ridondanti) nel modello. Nel dataset, l'ordine delle variabili da sinistra a destra deve essere: *RowType\_*, *VarName\_*, *P1*, *P2*, ..., dove *RowType\_* e *VarName\_* sono variabili stringa e *P1*, *P2*, ... sono variabili numeriche corrispondenti a un elenco ordinato di parametri.

- I valori iniziali vengono forniti su un record con valore *EST* per la variabile *RowType\_*; i valori iniziali effettivi sono indicati sotto le variabili *P1*, *P2*, .... La procedura ignora tutti i record per i quali *RowType\_* ha un valore diverso da *EST* e tutti i record oltre la prima ricorrenza di *RowType\_* uguale a *EST*.
- L'intercettazione, se inclusa nel modello, o 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 Split File è in vigore, allora le variabili devono iniziare con la variabile split - file o variabili nell'ordine specificato quando si crea il File di Split, seguito da *RowType\_*, *VarName\_*, *P1*, *P2*, ... come sopra. Le suddivise devono avvenire nel dataset specificato nello stesso ordine del dataset originale.

*Nota:* i nomi variabili *P1*, *P2*, ... non sono richiesti; la procedura accetterà qualsiasi nome variabile valido per i parametri in quanto la mappatura delle variabili ai parametri è basata sulla variabile posizione, non nome variabile. Eventuali variabili oltre l'ultimo parametro vengono ignorate.

La struttura dei file per i valori iniziali è la stessa utilizzata quando si esportano il modello come dati; quindi, è possibile utilizzare i valori finali da un'esecuzione della procedura come input in una successiva esecuzione.

## Statistiche delle Equazioni di stima generalizzate

**Effetti del modello.** Sono disponibili le seguenti opzioni:

- **Tipo di analisi.** Specificare il tipo di analisi da produrre per verificare gli effetti del modello. L'analisi di tipo I è generalmente appropriata quando si hanno dei motivi a priori per ordinare i predittori nel modello, mentre l'analisi di tipo III trova un'applicazione più generale. Le statistiche di Wald o di punteggio generalizzato sono calcolate in base alla selezione nel gruppo Chi - Square Statistics.
- **Intervalli di confidenza.** Specificare un livello di confidenza maggiore di 50 e minore di 100. Gli intervalli di Wald sono sempre prodotti indipendentemente dal tipo di statistiche di chi - quadrato selezionati, e si basano sul presupposto che i parametri abbiano una distribuzione normale asintotica.
- **Funzione di quasi - verosimigliamento.** Questo controlla il formato di visualizzazione della funzione di quasi - verosimigliamento del log. La funzione completa include un termine aggiuntivo costante rispetto alla stima dei parametri; non ha effetti sulla stima dei parametri ed è escluso dalla visualizzazione in alcuni prodotti software.

**Stampa.** È disponibile la seguente emissione.

- **Riepilogo elaborazione casi.** Visualizza il numero e la percentuale dei casi inclusi ed esclusi dall'analisi e la tabella di riepilogo dei dati correlati.
- **Statistiche descrittive.** Visualizza le statistiche descrittive e le informazioni di riepilogo su variabile dipendente, covariate e fattori.
- **Informazioni sul modello.** Visualizza il nome dell'insieme di dati, la variabile dipendente o le variabili eventi e prove, la variabile offset, la variabile peso di scala, la distribuzione della probabilità e la funzione di legame.
- **Statistiche della bontà di adattamento.** Visualizza due estensioni del Criterio informativo di Akaike per la selezione del modello: la quasi - verosimiglianza sotto il criterio del modello di indipendenza (QIC) per la scelta della migliore struttura di correlazione e un'altra misura QIC per la scelta del miglior sottoinsieme di predittori.
- **Statistiche di riepilogo del modello.** Visualizza i test di adattamento del modello, incluse le statistiche del rapporto di verosimiglianza per il test omnibus di adattamento del modello e le statistiche per i contrasti di tipo I o III di ogni effetto.
- **Stime dei parametri.** Visualizza le stime dei parametri e le statistiche di test e gli intervalli di confidenza corrispondenti. Se lo si desidera, è possibile visualizzare le stime dei parametri esponenziati oltre alle stime dei parametri grezzi.
- **Matrice di covarianza per le stime dei parametri.** Visualizza la matrice di covarianza dei parametri stimati.
- **Matrice di correlazione per le stime dei parametri.** Visualizza la matrice di correlazione dei parametri stimati.
- **Matrici del coefficiente di contrasto (L).** Visualizza i coefficienti di contrasto degli effetti predefiniti e delle medie marginali stimate, se richiesto nella scheda Medie marginali.
- **Funzione generale stimabile (s).** Visualizza le matrici per la generazione delle matrici del coefficiente di contrasto (L).
- **Cronologia iterazioni.** Visualizza la cronologia delle iterazioni per le stime dei parametri e la verosimiglianza log e stampa l'ultima valutazione del vettore gradiente e della matrice hessiana. La tabella della cronologia delle iterazioni visualizza le stime dei parametri per per ciascuna iterazione  $n$  esima che inizia con l'iterazione 0<sup>esima</sup> (stime iniziali), dove  $n$  è il valore dell'intervallo di stampa. Se è richiesta la cronologia delle iterazioni, l'ultima iterazione viene sempre visualizzata indipendentemente da  $n$ .
- **matrice di correlazione di lavoro.** Visualizza i valori della matrice che rappresentano le dipendenze all'interno. La sua struttura dipende dalle specifiche nella scheda [Repeated](#) .

## Equazioni di stima generalizzate EM Means

Questa scheda consente di visualizzare le medie marginali stimate per i livelli di fattori e le interazioni dei fattori. È anche possibile richiedere che venga visualizzata la media stimata complessiva. I mezzi marginali stimati non sono disponibili per i modelli multinazionali ordinali.

**Fattori e Interazioni.** Questo elenco contiene i fattori specificati nella scheda Predittori e le interazioni dei fattori specificati nella scheda Modello. Le covariate sono escluse da questa lista. I termini possono essere selezionati direttamente da questo elenco o combinati in un termine di interazione utilizzando il pulsante **By \***.

**Display Mezzi Per.** Le medie stimate vengono calcolate per i fattori e le interazioni dei fattori selezionati. Il contrasto determina la modalità di impostazione dei test di ipotesi per il confronto delle medie stimate. Il semplice contrasto richiede una categoria di riferimento o un livello di fattore rispetto al quale si confronta gli altri.

- **A coppie.** I confronti pairwise vengono calcolati per combinazioni di tutti i livelli dei fattori specificati o impliciti. Questo 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.** Ogni livello del fattore viene confrontato con il grand mean. I contrasti di deviazione non sono ortogonali.
- *Differenza.* Consente di confrontare la media di ciascun livello (a eccezione del primo) con la media dei livelli precedenti. Sono a volte definiti contrasti inversi 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 stimare le tendenze polinomiali.

**Scala.** I mezzi marginali stimati possono essere calcolati per la risposta, in base alla scala originale della variabile dipendente o per il predicatore lineare, in base alla variabile dipendente come trasformata dalla funzione di collegamento.

**Regolazione per Confronti multipli.** Quando si eseguono test di ipotesi con più contrasti, il livello di significatività generale può essere adattato in base ai livelli di significatività dei contrasti inclusi. Questo gruppo consente di scegliere il metodo di regolazione.

- **Differenza meno significativa.** Questo metodo non controlla la probabilità generale di rifiuto delle ipotesi che alcuni contrasti lineari siano diversi dai valori di ipotesi null.
- *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 decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.

## Equazioni di stima generalizzate Salva

Gli articoli controllati vengono salvati con il nome specificato; è possibile scegliere di sovrascrivere le variabili esistenti con lo stesso nome delle nuove variabili o evitare i conflitti di nome tramite suffissi appendici per rendere unici i nuovi nomi variabili.

- **Valore previsto di media di risposta.** Salva i valori previsti modello per ogni caso nella metrica di risposta originale. Quando la distribuzione di risposta è binomiale e la variabile dipendente è binaria, la procedura salva le probabilità previste. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Cumulative probabilità prevista** e la procedura salva la probabilità prevista

cumulativa per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

- **Abbassa l'intervallo di confidenza per la media della risposta.** Salva il limite inferiore dell'intervallo di confidenza per la media della risposta. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Bassa di intervallo di confidenza per la probabilità prevista cumulativa** e la procedura salva il limite inferiore per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.
- **Alto limite di intervallo di confidenza per la media della risposta.** Salva il limite superiore dell'intervallo di confidenza per la media della risposta. Quando la distribuzione di risposta è multinomiale, l'etichetta dell'articolo diventa **Upper bound of confidence interval per probabilità cumulativa** e la procedura salva il limite superiore per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.
- **Categoria prevista.** Per i modelli con distribuzione binomiale e variabile dipendente binario, o distribuzione multinomiale, questo salva la categoria di risposta prevista per ogni caso. Questa opzione non è disponibile per altre distribuzioni di risposta.
- **Valore previsto di predittore lineare.** Salva i valori previsti dal modello per ogni caso nella metrica del predittore lineare (risposta trasformata tramite la funzione di collegamento specificato). Quando la distribuzione di risposta è multinomiale, la procedura salva il valore previsto per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.
- **Errore standard stimato del valore previsto di predittore lineare.** Quando la distribuzione di risposta è multinomiale, la procedura salva l'errore standard stimato per ogni categoria della risposta, tranne l'ultima, fino al numero di categorie specificate da salvare.

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

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

## Stima delle Equazioni generalizzate Export

**Modello di esportazione come dati.** Scrive un dataset in formato IBM SPSS Statistica contenente la matrice di correlazione o la matrice di covarianza con le stime dei parametri, gli errori standard, i valori di significatività e i gradi di libertà. L'ordine delle variabili nel file di matrice è il seguente.

- **Split variabili.** Se utilizzato, eventuali variabili che definiscono la suddivisione.
- **RowType\_.** Prende i valori (e le etichette dei valori) *COV* (covarianze), *CORR* (correlazioni), *EST* (stime dei parametri), *SE* (errori standard), *SIG* (livelli di significatività) e *DF* (gradi di libertà del disegno di campionamento). Esiste un caso separato con tipo di riga *COV* (o *CORR*) per ogni parametro modello, più un caso separato per ciascuno degli altri tipi di riga.
- **VarName\_.** Assume valori *P1, P2, ...,* corrispondente ad un elenco ordinato di tutti i parametri di modello stimati (ad eccezione dei parametri binomiali di scala o negativi), per i tipi di righe *COV* o *CORR*, con etichette di valore corrispondenti alle stringhe di parametri mostrate nella tabella delle stime dei parametri. Le celle sono vuote per altri tipi di righe.
- **P1, P2, ...** Queste variabili corrispondono ad un elenco ordinato di tutti i parametri del modello (inclusa la scala e i parametri binomiali negativi, a seconda dei casi), con etichette variabili corrispondenti alle stringhe di parametro mostrate nella tabella delle stime dei parametri e assumono valori in base al tipo di riga.

Per i parametri ridondanti, tutte le covarianze sono impostate a zero, le correlazioni sono impostate sul valore mancante di sistema; tutte le stime dei parametri sono impostate a zero; e tutti gli errori standard, i livelli di significatività e i gradi residui di libertà 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 è stimato tramite la massima verosimiglianza, viene dato l'errore standard; altrimenti 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 probabilità, viene dato l'errore standard; altrimenti viene impostato sul valore mancante di sistema.

Se ci sono le divise, allora l'elenco dei parametri deve essere accumulato in tutte le divise. In una determinata suddivisione, alcuni parametri possono essere irrilevanti; questo non è uguale a quello ridondante. Per parametri irrilevanti, tutte le covarianze o correlazioni, le stime dei parametri, gli errori standard, i livelli di significatività e i gradi di libertà sono impostati sul valore mancante di sistema.

È possibile utilizzare questo file di matrice come valori iniziali per un'ulteriore stima del modello; notare che questo file non è immediatamente utilizzabile per ulteriori analisi in altre procedure che leggono un file di matrice a meno che tali procedure non accettino tutti i tipi di riga esportati qui. Anche allora, bisogna fare attenzione che tutti i parametri presenti in questo file di matrice abbiano lo stesso significato per la procedura di lettura del file.

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

## Funzioni Aggiuntive di comando GENLIN

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i valori iniziali per le stime dei parametri come elenco dei numeri (utilizzando il comando CRITERIA).
- Specificare una matrice di correlazione di lavoro fissa (utilizzando il sottocomando REPEATED).
- Fissare covariate a valori diversi dai loro mezzi quando si calcolano i mezzi marginali stimati (utilizzando il comando EMMEANS).
- Specificare contrasti polinomiali personalizzati per i mezzi marginali stimati (utilizzando il comando EMMEANS).
- Specificare un sottoinsieme dei fattori per i quali vengono visualizzati i mezzi marginali stimati da confrontare 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 GLMM 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 GLMM includono un'ampia gamma di modelli, dalla regressione lineare semplice ai modelli multilivello complessi per i dati longitudinali non normali.

### Esempi

Un provveditorato agli studi regionale può utilizzare il modello GLMM per stabilire se un metodo di insegnamento sperimentale è efficace nel migliorare i risultati in matematica degli studenti. Gli studenti della stessa classe devono essere correlati in quanto seguono lo stesso insegnante e anche le classi della stessa scuola possono essere correlate, in modo da poter includere gli effetti casuali a livello di scuola e di classe per tenere conto delle diverse sorgenti di variabilità.

I ricercatori in campo medico possono utilizzare un modello lineare generalizzato misto per determinare se un nuovo medicinale anticonvulsivo può ridurre la frequenza di crisi epilettiche in un paziente. Misurazioni ripetute dello stesso paziente sono generalmente correlate positivamente, pertanto un modello misto con alcuni effetti casuali dovrebbe essere appropriato. Il campo obiettivo,

il numero di crisi, acquisisce valori interi positivi, pertanto un modello lineare generalizzato misto con una distribuzione di Poisson e collegamento logaritmico potrebbe essere appropriato.

I dirigenti di una società fornitrice di servizi TV, telefono e Internet via cavo possono servirsi di un modello misto lineare generalizzato per scoprire più informazioni in merito ai potenziali clienti. Poiché le risposte possibili hanno livelli di misurazione nominali, gli analisti della società utilizzano un modello misto logit generalizzato con un'intercettazione casuale per acquisire la correlazione tra le risposte alle domande in merito all'utilizzo dei vari tipi di servizi (tv, telefono, Internet) nelle risposte di un determinato rispondente all'indagine.

La scheda Struttura dati consente di specificare le relazioni strutturali tra i record di un insieme di dati quando le osservazioni sono correlate. Se i record dell'insieme di dati rappresentano osservazioni indipendenti, non è necessario specificare alcun valore nella scheda.

## Opzioni degli effetti

### Soggetti

La combinazione di valori dei campi categoriali specificati deve definire in modo univoco i soggetti all'interno dell'insieme di dati. Per esempio, un singolo campo *ID paziente* dovrebbe essere sufficiente a definire i soggetti in un ospedale, ma la combinazione di *ID ospedale* e *ID paziente* potrebbe essere necessaria qualora i numeri di identificazione dei pazienti non fossero univoci nei diversi ospedali. 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 desideri modellare la correlazione tra queste osservazioni. Ci si potrebbe aspettare, per esempio, che le letture della pressione sanguigna di un paziente effettuate in occasione di visite al medico consecutive siano correlate.

Tutti i campi specificati come **Soggetti** nella finestra di dialogo Variabili vengono utilizzati per definire i soggetti per la struttura di covarianza residua e fornire l'elenco dei campi possibili per la definizione dei soggetti per le strutture di covarianza ad effetti casuali nel [Blocco effetti casuali](#).

### Misure ripetute

I campi specificati qui vengono utilizzati per identificare le osservazioni ripetute. Per esempio, una singola variabile *Settimana* potrebbe identificare 10 settimane di osservazioni in uno studio medico, mentre *Mese* e *Giorno* potrebbero essere utilizzati insieme per identificare le osservazioni giornaliere nel corso di un anno.

## Opzioni di covarianza

### Definisci i gruppi di covarianza per

I campi categorici specificati in questo punto definiscono insiemi indipendenti di parametri di covarianza a effetti ripetuti; uno per ciascuna 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 ripetuto

Specifica la struttura di covarianza per i residui. Diverse opzioni di covarianza sono disponibili in base al selezionato **Tipo di covarianza Repeated**. Sono disponibili le seguenti strutture:

- Autoregressivo di primo ordine (AR1)
- AR1 prodotto diretto (UN\_AR1)
- Prodotto diretto non strutturato (UN\_UN)
- Simmetria del composto prodotto diretto (UN\_CS)
- Simmetria composta eterogenea (CSH)
- Autoregressiva eterogenea (ARH1)

- Media mobile autoregressiva (1,1) (ARMA11)
- Simmetria del composto
- Diagonale
- Identità scalata
- Toeplitz
- Non strutturato
- Componenti della varianza
- Spaziale: potenza
- Spaziale: esponenziale
- Spaziale: gaussiano
- Spaziale: lineare
- Spaziale: log lineare
- Spaziale: sferico

### Misure Kronecker

Selezionare le variabili che specificano la struttura del soggetto o le misurazioni della covarianza di Kronecker e determinare in che modo vengono corretti gli errori di misurazione. Il campo è disponibile solo quando viene selezionato uno dei seguenti **Tipo di covarianza ripetuta**:

- AR1 prodotto diretto (UN\_AR1)
- Prodotto diretto non strutturato (UN\_UN)
- Simmetria del composto prodotto diretto (UN\_CS)

### Coordinate covarianza spaziale

Le variabili in questo elenco specificano le coordinate delle osservazioni ripetute quando uno dei tipi di covarianza spaziale viene selezionato per il tipo di covarianza ripetuta.

Per ulteriori informazioni, consultare l'argomento [“Strutture covarianza”](#) a pagina 102.

### Pseudo misure $R^2$

Le misure pseudo- $R^2$  e il coefficiente di correlazione all'interno della classe sono inclusi nell'output GLMM (quando appropriato). Le misure pseudo- $R^2$  si basano interamente sulle stime finali e vengono prodotte dopo che la stima è stata completata. Il coefficiente di determinazione  $R^2$  è una statistica comunemente riportata, perché rappresenta la proporzione di varianza spiegata da un modello lineare. Il coefficiente di correlazione intra-class (ICC) è una statistica correlata che quantifica la proporzione di varianza spiegata da un fattore di raggruppamento (casuale) in dati multilevel / gerarchici.

## Ottenimento di un modello misto lineare generalizzato

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Advanced Statistics.

Dai menu, scegliere:

### Analizzare > Modelli Misti > Linear Linear ...

1. Definire la struttura oggetto del tuo dataset sulla scheda **Struttura dati**. La scheda fornisce opzioni per specificare sia gli effetti del modello che il tipo di covarianza ripetuta.
2. Nella scheda **Campi e Effetti** ci deve essere un unico obiettivo, che può avere qualsiasi livello di misurazione, oppure una specifica di eventi / trial, nel qual caso le specifiche degli eventi e delle prove devono essere continue. Facoltativamente specificare la sua funzione di distribuzione e collegamento, gli effetti fissi e eventuali blocchi di effetti casuali, offset o pesi di analisi.
3. Fare clic su **Opzioni di creazione** per specificare le impostazioni di build opzionali.
4. Fare clic su **Opzioni modello** per salvare i punteggi nel dataset attivo ed esportare il modello in un file esterno.

5. Fare clic su **Esegui** per eseguire la procedura e creare gli oggetti Modello.

## Obiettivo

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. Per 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. Per 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 vanno interpretate come relative alla probabilità della categoria 0 o 1 *relativa* alla probabilità della categoria 2. Se si specifica una categoria personalizzata e la propria destinazione ha definito etichette, è possibile impostare la categoria di riferimento scegliendo un valore dall'elenco. Questo può risultare comodo 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 diversi modelli comuni oppure è possibile scegliere un'impostazione **Personalizzata** se esiste una particolare combinazione delle funzioni di distribuzione e collegamento che si desidera adattare e che non è presente 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 da utilizzare quando l'obiettivo è una risposta multi-categoria. Utilizza un collegamento logit cumulativo (risultati ordinali) o un collegamento logit generalizzato (risposte nominali 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 appropriata per un obiettivo che rappresenta una risposta a più categorie. La forma del modello varia a seconda del livello di misurazione dell'obiettivo.

Un obiettivo **nominale** produrrà un modello multinomiale nominale in cui per ogni categoria dell'obiettivo (ad eccezione della categoria di riferimento) viene effettuata la stima di un insieme separato di parametri del modello. Le stime dei parametri per un dato predittore mostrano la relazione tra quel predittore e la probabilità di ogni categoria dell'obiettivo, relativamente alla categoria di riferimento.

Un obiettivo **ordinale** produrrà un modello multinomiale ordinale in cui il termine di intercettazione tradizionale viene sostituito con un insieme di parametri di **soglia** correlati alla probabilità cumulativa delle categorie obiettivo.

- **Binomiale negativa.** La regressione binomiale negativa utilizza una distribuzione binomiale negativa con un collegamento logaritmico, da utilizzare quando l'obiettivo rappresenta un numero di ricorrenze con varianza elevata.
- **Normale.** Questa distribuzione è indicata per un target 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.**  $f(x)=\log(-\log(1-x))$ . È adatta solo con la distribuzione binomiale o multinomiale.
- **Cauchit.**  $f(x) = \tan(\pi (x - 0.5))$ . È adatta solo con la distribuzione binomiale o multinomiale.
- **Log.**  $f(x)=\log(x)$ . Questa funzione di collegamento può essere usata per tutti i tipi di distribuzioni, tranne che per quella multinomiale.
- **Complemento log.**  $f(x)=\log(1-x)$ . Questa funzione è indicata solo per la distribuzione binomiale.
- **Logit.**  $f(x)=\log(x / (1-x))$ . È adatta solo con la distribuzione binomiale o multinomiale.
- **Log-log negativa.**  $f(x)=-\log(-\log(x))$ . È adatta solo con la distribuzione binomiale o multinomiale.

- **Probit.**  $f(x)=\Phi^{-1}(x)$ , dove  $\Phi^{-1}$  è la funzione di distribuzione cumulativa normale standard inversa. È adatta solo con la distribuzione binomiale o multinomiale.
- **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





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.

Immettere gli effetti nel modello selezionando uno o più campi nell'elenco di sorgenti 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.
- **a 2 vie.** Tutte le possibili coppie dei campi rilasciati vengono visualizzate come interazioni a 2 vie in fondo all'elenco degli effetti.
- **a 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 eseguire diverse azioni.

Tabella 1. Descrizioni dei pulsanti del generatore di effetti

Icona	Descrizione
	Eliminare i termini del modello a effetti fissi selezionando i termini che si desidera eliminare e facendo clic sul pulsante di eliminazione.
 	Riordinare i termini all'interno del modello a effetti fissi selezionando i termini che si desidera riordinare e facendo clic sulla freccia rivolta verso l'alto o verso il basso.
	Aggiungere termini nidificati al modello utilizzando la finestra di dialogo <u>"Aggiungi un termine personalizzato"</u> a pagina 55, facendo clic sul pulsante <u>Aggiungi un termine personalizzato</u> .

**Includi intercettazione.** Generalmente, l'intercettazione è inclusa nel modello. Se è possibile presumere che i dati passino attraverso l'origine, l'intercettazione può essere esclusa.

## Aggiungi un 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. Per esempio, una catena di supermercati può seguire le abitudini di spesa dei propri clienti in più negozi. Poiché ogni cliente frequenta un solo negozio, l'effetto *Cliente* può definirsi **nidificato** all'interno dell'effetto *Negozi*.

È inoltre possibile includere effetti di interazione, ad esempio termini polinomiali che interessano la stessa covariata, o aggiungere più livelli di nidificazione al termine nidificato.

**Limitazione.** I termini nidificati sono sottoposti alle seguenti restrizioni:

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

Creazione di un termine nidificato

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

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

## 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. Possono essere utilizzati per descrivere la variabilità di eccesso 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. Per esempio, se sono stati selezionati i soggetti Scuola, Classe e Studente nella scheda Struttura dati, vengono creati automaticamente i seguenti blocchi di effetti casuali:

- Effetto casuale 1: il soggetto è scuola (nessun effetto, solo intercettazione)
- Effetto casuale 2: il soggetto è scuola \* classe (nessun effetto, solo intercettazione)

È possibile utilizzare i blocchi di effetti casuali nei seguenti modi:

1. Per aggiungere un nuovo blocco, fare clic su **Aggiungi blocco ...** Questo apre la finestra di dialogo ["Blocco effetti casuali "](#) a pagina 56 .
2. Per modificare un blocco esistente, selezionare il blocco che si desidera modificare e fare clic su **Modifica blocco ...** Questo apre la finestra di dialogo ["Blocco effetti casuali "](#) a pagina 56 .
3. Per eliminare uno o più blocchi, selezionare i blocchi da eliminare e fare clic sul pulsante di eliminazione.

## Blocco effetti casuali

Immettere gli effetti nel modello selezionando uno o più campi nell'elenco di sorgenti e aggiungendoli all'elenco **degli effetti**.

Il tipo di effetto creato dipende da quale **Tipo** si è selezionato. I campi categoriali (nominali e ordinali) vengono utilizzati come fattori nel modello e i campi continui vengono utilizzati come covariate.

### Interazione

La combinazione di tutti i campi rilasciati viene visualizzata come una singola interazione nelle parte inferiore dell'elenco degli effetti.

### Effetti principali

I campi rilasciati vengono visualizzati come effetti principali separati in fondo all'elenco degli effetti.

### Tutti - 2 vie

Tutte le possibili coppie dei campi rilasciati vengono visualizzate come interazioni a 2 vie in fondo all'elenco degli effetti.

### Tutti - 3 vie

Tutti i possibili gruppi di tre dei campi rilasciati vengono visualizzati come interazioni a 3 vie in fondo all'elenco degli effetti.

### Tutti - 4 vie

Tutti i possibili gruppi di tre dei campi rilasciati vengono visualizzati come interazioni a 4 vie nell'elenco degli effetti.

### Tutti - 5 vie

Tutti i possibili gruppi di tre dei campi rilasciati vengono visualizzati come interazioni a 5 vie in fondo all'elenco degli effetti.

### Includi intercettazione

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

### Visualizza previsioni del parametro per questo blocco

Specifica di visualizzare le stime dei parametri ad effetti casuali.

### Combinazione soggetti

Consente di specificare soggetti degli effetti casuali da combinazioni predefinite di soggetti dalla scheda Variables dialog. Per esempio, se *Scuola*, *Classe*, e *Studente* sono definiti come soggetti nella scheda Variables dialog, e in questo ordiner, l'elenco a discesa Combinazione Soggetti includerà le opzioni **Nessuna**, **Scuola**, **Scuola \* Classe**, e **Scuola \* Classe \* Studente**.

### Tipo di covarianza effetto casuale

Specifica la struttura di covarianza per i residui. Sono disponibili le seguenti strutture:

- Autoregressivo di primo ordine (AR1)
- Media mobile autoregressiva (1,1) (ARMA11)
- Simmetria del composto
- Diagonale
- Identità scalata
- Toeplitz
- Non strutturato
- Componenti della varianza

### Definisci i gruppi di covarianza per

I campi categorici specificati in questo punto definiscono insiemi indipendenti di parametri di covarianza a effetti ripetuti; uno per ciascuna 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.

## Peso e offset

**Peso analisi.** Il parametro di scala è un parametro stimato del modello relativo alla varianza della risposta. I pesi dell'analisi sono valori "noti" che possono variare a seconda delle osservazioni. Se il campo Peso 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.** Il termine offset è un predittore "strutturale". Il suo coefficiente non è stimato dal modello ma si presume che abbia il valore 1; pertanto, i valori dell'offset vengono semplicemente aggiunti al predittore lineare dell'obiettivo. Ciò è particolarmente utile nei modelli di regressione di Poisson, nei quali ogni caso può avere diversi livelli di esposizione all'evento di interesse.

Per esempio, nella modellazione dei tassi di incidente per singolo conducente esiste una differenza significativa tra un autista responsabile di un incidente in tre anni di esperienza e un autista responsabile di un incidente in 25 anni. Il numero di incidenti può essere rappresentato come una risposta di Poisson o

binomiale negativa con un collegamento log se il logaritmo naturale dell'esperienza del conducente viene incluso come termine di offset.

Altre combinazioni dei tipi di distribuzione e di collegamento richiederebbero altre trasformazioni della variabile di offset.

## Opzioni generali di creazione

Queste opzioni consentono di specificare alcuni criteri avanzati per la creazione del modello.

### Criterio di ordinamento

Questi comandi determinano l'ordine delle categorie per l'obiettivo e i fattori (input categoriali) allo scopo di determinare l'"ultima" categoria. L'impostazione del criterio di ordinamento dell'obiettivo viene ignorata se l'obiettivo non appartiene ad una categoria o se è specificata una categoria di riferimento personalizzata nelle impostazioni "[Obiettivo](#)" a pagina 53.

### Regole di arresto

È possibile specificare il numero massimo di iterazioni che verranno eseguite nell' algoritmo. L'algoritmo utilizza un processo doppiamente iterativo che consiste di un loop interno e un loop esterno. Il valore specificato per il numero massimo di iterazioni si applica a entrambi i loop. Specificare un intero non negativo. Il valore predefinito è 100.

### Impostazioni post-stima

Queste impostazioni determinano come viene calcolata parte dell'output del modello 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. Il valore predefinito è 95.

### Gradi di libertà

Specifica la modalità di calcolo dei gradi di libertà per i test di significatività. Scegliere **Metodo dei residui** se la dimensione del campione è sufficientemente grande, oppure se i dati sono bilanciati, oppure il modello utilizza un modello di covarianza più semplice (ad esempio identità scalata o diagonale). Questa è l'impostazione predefinita. Scegliere **l'approssimazione di** se la dimensione del campione è piccola, oppure se i dati non sono bilanciati oppure se il modello use un tipo di covarianza complesso (ad esempio non strutturato). Scegliere **l'approssimazione Kenward-Roger** se la dimensione del campione è piccola e si dispone di un modello Limite massimo di verosimiglianza limitato (REML).

### Test dei coefficienti e degli effetti fissi

Si tratta del metodo per il calcolo della matrice di covarianza delle stime dei parametri. Scegliere la stima robusta se si teme che le ipotesi di modello vengano violate.

## Stima

Il modello che crea l'algoritmo utilizza un processo doppiamente iterativo che consiste di un loop interno e un loop esterno. Al loop interno si applicano le seguenti impostazioni.

### Convergenza parametri.

Si presume la convergenza se la modifica assoluta massima o la modifica relativa massima delle stime del parametro sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Convergenza verosimiglianza.

Si presume la convergenza se la modifica assoluta o la modifica relativa nella funzione di verosimiglianza log sono inferiori al valore specificato, che non deve essere negativo. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Convergenza hessiana.

Per la specifica **Absolute**, la convergenza viene presunta se la statistica basata sulla convergenza hessiana è inferiore al valore specificato. Per la specifica **Relativa**, si presume la convergenza se la statistica è minore del prodotto tra il valore specificato e il valore assoluto della log-verosimiglianza. Il criterio non viene utilizzato se il valore specificato è uguale a 0.

### Max fasi di calcolo del punteggio di Fisher.

Specificare un intero non negativo. Il valore 0 specifica il metodo Newton-Raphson. I valori maggiori di 0 indicano di utilizzare l'algoritmo di punteggio di Fisher fino al numero di iterazioni  $n$ , in cui  $n$  è il numero intero specificato e poi il Newton-Raphson.

### Tolleranza della singolarità

Questo valore viene utilizzato come tolleranza nel controllo della singolarità. Specificare un valore positivo.

**Nota:** Per impostazione predefinita, viene utilizzata la convergenza parametri, in cui il valore **Absolute** massimo cambia ad una tolleranza  $1E-6$ . Questa impostazione può produrre dei risultati che differiscono dai risultati ottenuti nelle versioni prima della versione 22. Per riprodurre i risultati dalle versioni precedenti alla 22, utilizzare **Relative** per il criterio di convergenza parametri e mantenere il valore di tolleranza predefinito  $1E-6$ .

## Medie stimate

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

**Termini.** Visualizza l'elenco dei termini del modello negli effetti fissi interamente composti da campi categoriali. Selezionare ogni termine per il quale si desidera che il modello produca medie marginali stimate.

- **Tipo contrasto.** Specifica il tipo di contrasto da utilizzare per i livelli del campo contrasto. Se viene selezionata l'opzione **Nessuno**, non viene prodotto alcun contrasto. **A coppie** produce confronti a coppie per tutte le combinazioni di livelli dei fattori specificati. Questo è l'unico contrasto disponibile per le interazioni tra fattori. I contrasti **Deviazione** confrontano ogni livello del fattore con la media finale. I contrasti **Semplice** confrontano ogni livello del fattore, ad eccezione dell'ultimo, con l'ultimo livello. L'"ultimo" livello è determinato dal criterio di ordinamento per i fattori specificato in Opzioni di costruzione. Si noti 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 stato selezionato il tipo di contrasto **Nessuno**, non sarà possibile (o necessario) selezionare alcun campo contrasto.

**Campi continui.** I campi continui elencati vengono estratti dai termini negli effetti fissi che utilizzano campi continui. Nel calcolo delle medie marginali stimate, le covariate sono fisse ai valori specificati. Selezionare la media o 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. Si noti che se l'obiettivo viene specificato utilizzando l'opzione eventi/prove, vengono calcolate le medie marginali stimate per la parte eventi/prove anziché per il numero di eventi. **Trasformazione funzione collegamento** calcola le medie marginali stimate per il predittore lineare.

**Adatta per confronti multipli utilizzando.** Quando si eseguono test di ipotesi con più contrasti, il livello di significatività generale può essere adattato in base ai livelli di significatività dei contrasti inclusi. Ciò consente di selezionare il metodo di adattamento.

- **Differenza meno significativa.** Questo metodo non controlla la probabilità generale di rifiuto delle ipotesi che alcuni contrasti lineari siano diversi dai valori di ipotesi null.
- **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 decrescenti, molto meno conservativa in termini di rifiuto di singole ipotesi, ma che mantiene lo stesso livello di significatività globale.

Il metodo della differenza meno significativa è meno conservativo del metodo Sidak sequenziale che a sua volta è meno conservativo rispetto al metodo Bonferroni sequenziale. La differenza meno significativa

rifiuterà almeno lo stesso numero di ipotesi individuali del Sidak sequenziale che a sua volta rifiuterà almeno lo stesso numero di ipotesi individuali del Bonferroni sequenziale.

## Salva

Gli articoli controllati vengono salvati con il nome specificato; non sono consentiti conflitti con i nomi di campo esistenti.

**Valori previsti** - Salva il valore previsto dell'obiettivo. Il nome di campo predefinito è *Valoreprevisto*.

**Probabilità prevista per gli 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**. I valori calcolati sono probabilità cumulative per gli obiettivi ordinali. Il nome root predefinito è *PredictedProbability*. Per salvare la probabilità prevista della categoria prevista, salva la fiducia (cfr.

**Intervalli di confidenza** - Salva i limiti superiore e inferiore dell'intervallo di confidenza per il valore previsto o la probabilità prevista. Per tutte le distribuzioni ad eccezione delle multinomiali, questo crea due variabili e il nome root predefinito è *CI*, con *\_Lower* e *\_Upper* come suffissi.

Per la distribuzione multinomiale e un obiettivo nominale, viene creato un campo per ogni categoria variabile dipendente. Questo salva i limiti più bassi e superiori della probabilità prevista per le prime categorie  $n$  fino al valore specificato come **Categorie massime da salvare**. Il nome root predefinito è *CIe* i nomi dei campi predefiniti sono *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* e così via, corrispondente all'ordine delle categorie target.

Per la distribuzione multinomiale e un obiettivo ordinale, viene creato un campo per ogni categoria di variabile dipendente tranne l'ultimo (consultare l'argomento [“Opzioni generali di creazione”](#) a pagina 58 per ulteriori informazioni. ). Ciò salva i limiti inferiore e superiore della probabilità prevista cumulativa per le prime  $n$  categorie, fino all'ultima, ma non fino al valore specificato come **Numero massimo di categorie da salvare**. Il nome root predefinito è *CIe* i nomi campo predefiniti sono *CI\_Lower\_1*, *CI\_Upper\_1*, *CI\_Lower\_2*, *CI\_Upper\_2* e così via, corrispondenti all'ordine delle categorie di destinazione.

**Residui di Pearson** - Salva il residuo di Pearson per ogni record, che può essere utilizzato nella diagnostica post stima dell'adattamento del modello. Il nome del campo di default è *PearsonResidual*.

**Confidenze** - Salva la confidenza nel valore previsto 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 di campo predefinito è *Confidenza*.

## Esporta

**Esporta modello.** Consente di scrivere il modello in un file *.zip* esterno. È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio. Per ulteriori informazioni, consultare l'argomento [Procedura guidata di calcolo del punteggio](#) . Specificare un nome file valido e univoco. Se la specifica del file si riferisce ad un file esistente, il file viene sovrascritto.

### Esporta output

**Esporta EBLUPS.** Questa sezione è abilitata se viene creato almeno un blocco di effetti casuali tramite la finestra di dialogo della voce Effetti casuali nella scheda Campi & Effetti e viene selezionata la casella di spunta per visualizzare le previsioni dei parametri per tale blocco. Se sono stati specificati più blocchi di questo tipo, è possibile scegliere di salvare i risultati in dataset o file separati oppure di combinarli in un singolo dataset o file.

## Vista modello

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

Per default, viene mostrata la visualizzazione Riepilogo modello. Per un'altra visualizzazione del modello, selezionarla dalle miniature delle visualizzazioni.

In alternativa all'oggetto Modello, è possibile generare tabelle e grafici pivot selezionando **Pivot tabelle e grafici** nel gruppo Output Display nella scheda Output della finestra di dialogo Opzioni (Modifica> Opzioni). L'oggetto Modello è descritto negli argomenti riportati di seguito.

## Riepilogo del modello

Questa visualizzazione è una snapshot, un riepilogo del modello e del suo adattamento.

**Tabella.** La tabella identifica l'obiettivo, la distribuzione della probabilità e la funzione di collegamento specificata in Impostazioni obiettivo. Se l'obiettivo è definito da eventi e prove, la cella viene suddivisa in modo da mostrare il campo eventi ed il campo prove o un 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 la selezione e il confronto di modelli misti basata sulla log likelihood -2 (ristretta). I valori più bassi indicano i modelli migliori. AICC corregge AIC per dimensioni del campione piccole. Mano a mano che aumenta la dimensione del campione, AICC converge in AIC.
- *Bayesiano.* Una misura per la selezione e il confronto dei modelli basata sulla -2 log likelihood. I valori più bassi indicano i modelli migliori. Anche BIC "penalizza" i modelli sovraparametrizzati (modelli complessi con un elevato numero di input, ad esempio) ma in modo più rigoroso rispetto ad AIC.

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

## Struttura dati

Questa visualizzazione offre un riepilogo della struttura di 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 di ogni campo soggetto e campo misure ripetute.

## Previsioni e osservazioni

Per i target continui, inclusi gli obiettivi specificati come eventi/prove, viene visualizzato un grafico a dispersione in bin dei valori previsti sull'asse verticale in base ai valori osservati sull'asse orizzontale. Idealmente, i punti devono giacere su una linea di 45 gradi; questa vista può indicare eventuali record previsti in modo particolarmente non corretto dal modello.

## Classificazione

Per gli obiettivi categoriali, visualizza la classificazione incrociata dei valori osservati e previsti in una mappa termica, più la percentuale globale corretta.

**Stili di tabella.** Sono disponibili diversi stili di visualizzazione, a cui è possibile accedere dall'elenco a discesa **Stile**.

- **Percentuali di riga.** Questa opzione visualizza la percentuale di righe (il numero di celle espresso come percentuale del totale di righe) nelle celle. È l'impostazione predefinita.
- **Conteggi delle celle.** Questa opzione visualizza il numero di celle nelle celle. L'ombreggiatura per la mappa termica è comunque basata sulle percentuali di riga.
- **Mappa termica.** Questa opzione non visualizza alcun valore nelle celle, ma solo l'ombreggiatura.
- **Compresso.** Questa opzione non visualizza alcuna intestazione di righe o colonne o valori nelle celle. Può essere utile se l'obiettivo ha molte categorie.

**Mancante.** Se, per alcuni record, non sono presenti valori nell'obiettivo, i valori vengono visualizzati in una riga (**Mancante**) al di sotto di tutte le righe valide. I record con valori mancanti non contribuiscono alla percentuale globale corretta.

**Più obiettivi.** Se sono presenti più obiettivi di categoria, ciascun obiettivo è visualizzato in una tabella separata ed è presente un elenco a discesa **Obiettivo** che controlla l'obiettivo da visualizzare.

**Tabelle di grandi dimensioni.** Se l'obiettivo visualizzato contiene più di 100 categorie, non viene visualizzata alcuna tabella.

## Effetti fissi

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

**Stili.** Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Diagramma.** Questo è un grafico in cui gli effetti sono disposti dall'alto verso il basso in base all'ordine in cui sono stati specificati nelle impostazioni Effetti fissi. Le linee di connessione nel diagramma sono pesate in base alla significatività dell'effetto, per cui la larghezza della linea più grande corrisponde agli effetti più significativi (valori  $p$  più piccoli). È l'impostazione predefinita.
- **Tabella.** Si tratta di una tabella ANOVA per gli effetti del modello globale e del modello individuale. I singoli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi.

**Significatività.** Il dispositivo di scorrimento Significatività controlla gli effetti mostrati nella visualizzazione. Gli effetti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sugli effetti più importanti. Per impostazione predefinita, il valore è 1.00, quindi non viene filtrato alcun effetto 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.** Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Diagramma.** Questo grafico visualizza prima l'intercettazione, quindi ordina gli effetti dall'alto verso il basso nell'ordine in cui sono stati specificati nelle impostazioni Effetti fissi. All'interno degli effetti che contengono fattori, i coefficienti sono disposti in ordine crescente di 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 predefinito.
- **Tabella.** Mostra i valori, i test di significatività e gli intervalli di confidenza per i singoli coefficienti del modello. Dopo l'intercettazione, gli effetti vengono ordinati dall'alto verso il basso nell'ordine specificato nelle impostazioni Effetti fissi. All'interno degli effetti che contengono fattori, i coefficienti sono disposti in ordine crescente di valori dei dati.

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

**Esponenziale.** Visualizza le stime del coefficiente esponenziale e gli intervalli di confidenza per alcuni tipi di modelli, inclusi Regressione logistica binaria (distribuzione binomiale e collegamento logit), Regressione logistica nominale (distribuzione multinomiale e collegamento logit), Regressione binomiale negativa (distribuzione binomiale negativa e collegamento log) e Modello log-lineare (distribuzione di Poisson e collegamento log).

**Significatività.** È disponibile un dispositivo di scorrimento Significatività che controlla i coefficienti mostrati nella visualizzazione. I coefficienti con valori di significatività superiori al valore del dispositivo di scorrimento vengono nascosti. Questo non cambia il modello, ma consente semplicemente di concentrare l'attenzione sui coefficienti più importanti. Per impostazione predefinita, il valore è 1.00, quindi non viene filtrato alcun coefficiente in base alla significatività.

## Covarianze di effetti casuali

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

**Stili.** Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

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

**Blocchi.** Se sono presenti più blocchi di effetti casuali, è disponibile un elenco a discesa Blocco per la selezione del blocco da visualizzare.

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

**Multinomiale.** Se è attiva la distribuzione multinomiale, l'elenco a discesa Multinomiale controlla la categoria obiettivo da 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 dei residui (**R**) e di effetti casuali (**G**), la classificazione (numero di colonne) nelle matrici di progetto degli effetti fissi (**X**) e degli effetti casuali (**Z**) ed il numero di oggetti definiti mediante i campi soggetto che definiscono la struttura dei dati.

**Tabella Parametri di covarianza.** Per l'effetto selezionato, per ciascun parametri di covarianza sono visualizzati la stima, l'errore standard e l'intervallo di confidenza. Il numero di parametri visualizzati varia a seconda della struttura di covarianza per l'effetto e, per i blocchi di effetti casuali, del numero di effetti del blocco. Se si verifica che i parametri esterni alla diagonale non sono significativi, è possibile utilizzare una struttura di covarianza più semplice.

**Effetti.** Se sono presenti blocchi di effetti casuali, è disponibile un elenco a discesa Effetto per la selezione del blocco di effetti casuali o residui da visualizzare. L'effetto residuo è sempre disponibile.

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

**Multinomiale.** Se è attiva la distribuzione multinomiale, l'elenco a discesa Multinomiale controlla la categoria obiettivo da 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 effetti fissi per tutti i fattori "più significativi", iniziando con le interazioni a tre vie, seguite dalle interazioni a due vie e infine 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. Offre una visualizzazione utile degli effetti dei coefficienti di ciascun predittore sull'obiettivo. Si noti che se non sono presenti predittori significativi, non vengono generate medie stimate.

**Confidenza.** Visualizza i limiti di confidenza superiore ed inferiore per le medie marginali, utilizzando il livello di confidenza specificato come parte delle opzioni di creazione.

## Medie stimate: effetti personalizzati

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

**Stili.** Sono disponibili diversi stili di visualizzazione, che sono accessibili dall'elenco a discesa **Stile**.

- **Diagramma.** Lo stile visualizza un grafico a linee del valore stimato del modello del target sull'asse verticale per ciascun valore dell'effetto principale (o primo effetto elencato in un'interazione) sull'asse orizzontale; viene creata una linea separata per ciascun valore del secondo effetto elencato in un'interazione; viene creato un grafico separato per ciascun valore del terzo effetto elencato in un'interazione a tre vie; tutti gli altri predittori vengono mantenuti 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 significative dal punto di vista statistico; le linee nere corrispondono alle differenze non significative. Passando il cursore del mouse su una linea nella rete viene visualizzato un suggerimento con la significatività adattata della differenza tra i nodi connessi 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 stimato del modello dell'obiettivo, del relativo errore standard e dell'intervallo di confidenza per ciascuna combinazione di livello dei campi nell'effetto; tutti gli altri predittori vengono mantenuti 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 ed inferiore per le medie marginali, utilizzando il livello di confidenza specificato come parte delle opzioni di creazione.

**Layout.** Attiva o disattiva il layout 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 modello: Analisi loglineare

La procedura di analisi Loglineare di selezione Modello analizza le crosstabulazioni multivie (tabelle di contingenza). Si adatta ai modelli loglineari gerarchici a crosstabulazioni multidimensionali utilizzando un algoritmo di adattamento proporzionale iterativo. Questa procedura ti aiuta a scoprire quali variabili categoriali sono associate. Per costruire modelli, sono disponibili i metodi di ingresso e di eliminazione all'indietro. Per i modelli saturi è possibile richiedere le stime dei parametri e i test di associazione parziale. Un modello saturo aggiunge 0.5 a tutte le celle.

**Esempio.** In uno studio di preferenza dell'utente per uno di due detergenti per bucato, i ricercatori hanno contato persone in ogni gruppo, combinando varie categorie di morbidezza dell'acqua (morbida, media o

dura), precedente l'uso di uno dei marchi, e la temperatura di lavaggio (fredda o calda). Hanno scoperto come la temperatura sia correlata alla morbidezza dell'acqua e anche alla preferenza del brand.

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

Modello Selezione Logica di selezione Analisi Dati

**Dati.** Le variabili del fattore sono categoriali. Tutte le variabili da analizzare devono essere numeriche. Le variabili di stringa categoriali possono essere recuperate su variabili numeriche prima di iniziare l'analisi di selezione del modello.

Evitare di specificare molte variabili con molti livelli. Tali specifiche possono portare a una situazione in cui molte cellule hanno piccoli numeri di osservazioni, e i valori di chi - quadrato potrebbero non essere utili.

**Procedure correlate.** La procedura di selezione Modello può aiutare a identificare i termini necessari nel modello. Poi è possibile continuare a valutare il modello utilizzando Analisi Loglineare generale o Analisi Loglineare Logit. È possibile utilizzare Autorecode per codificare le variabili di stringa. Se una variabile numerica ha categorie vuote, utilizzare Recode per creare valori interi consecutivi.

Ottenimento di una Model Selection Loglinear Analysis

Dai menu, scegliere:

**Analisi > Loglineare > Modello selezione ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

1. Selezionare due o più fattori categoriali numerici.
2. Selezionare una o più variabili di fattore nell'elenco di Fattore (s) e fare clic su **Definisci intervallo**.
3. Definire la gamma di valori per ogni variabile fattore.
4. Selezionare un'opzione nel gruppo Model Building.

Opzionalmente è possibile selezionare una variabile di peso cellulare per specificare zeri strutturali.

## Analisi loglineare Definisci intervallo

È necessario indicare la gamma di categorie per ogni variabile fattore. I valori per il minimo e il massimo corrispondono alle categorie più basse e più alte della variabile fattore. Entrambi i valori devono essere interi e il valore minimo deve essere inferiore al valore massimo. Sono esclusi i casi con valori al di fuori dei limiti. Ad esempio, se si specifica un valore minimo di 1 e un valore massimo di 3, vengono utilizzati solo i valori 1, 2 e 3. Ripetere questo processo per ogni variabile fattore.

## Modello di analisi loglineare

**Specifica modello.** Un modello saturo contiene tutti gli effetti principali di fattore e tutte le interazioni fattore - per - fattore. Selezionare **Personalizzato** per specificare una classe di generazione per un modello non saturo.

**Generazione Classe.** Una classe di generazione è un elenco dei termini di ordine più alto in cui appaiono i fattori. Un modello gerarchico contiene i termini che definiscono la classe generatrice e tutti i parenti di ordine inferiore. Supponiamo di selezionare le variabili  $A$ ,  $B$  e  $C$  nell'elenco dei Fattori e quindi selezionare **Interazione** dall'elenco a discesa di Build Terms. Il modello risultante conterrà l'interazione specificata a 3 vie  $A*B*C$ , le interazioni a 2 vie  $A*B$ ,  $A*C$  e  $B*C$  e gli effetti principali per  $A$ ,  $B$  e  $C$ . Non specificare i parenti a basso ordine nella classe di generazione.

Per i fattori 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.

## Creazione di termini e termini personalizzati

### Crea termini

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### Crea termini personalizzati

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## Opzioni di analisi Loglineare di selezione modello

**Visualizza.** È possibile scegliere **Frequenza**, **Residualio** entrambi. In un modello saturo, le frequenze osservate e attese sono uguali e i residui sono pari a 0.

**Grafico.** Per i modelli personalizzati è possibile scegliere uno o entrambi i tipi di tracciati, **Residuali** e **Probabilità normale**. Questi aiuteranno a determinare quanto un modello si adatta ai dati.

**Visualizza per modello Saturato.** Per un modello saturato è possibile scegliere **stime dei parametri**. Le stime dei parametri possono aiutare a determinare quali termini possono essere sganciati dal modello. È disponibile anche una tabella di associazione, che elenca i test di associazione parziale. Questa opzione è computazionalmente costosa per tabelle con molti fattori.

**Criteri modello.** Un algoritmo di adattamento proporzionale iterativo viene utilizzato per ottenere le stime dei parametri. È possibile sovrascrivere uno o più criteri di stima specificando **Massimo iterazioni**, **Convergenza** o **Delta** (un valore aggiunto a tutte le frequenze cellulari per i modelli saturi).

## HILOGLINEAR Comando Funzioni Aggiuntive

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare i pesi delle celle in forma matrice (utilizzando il sottocomando CWEIGHT ).
- Generare analisi di diversi modelli con un singolo comando (utilizzando il comando DESIGN ).

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

## Analisi loglineare generale

---

La procedura General Loglinear Analysis analizza la frequenza di conteggio delle osservazioni che rientrano in ogni categoria di classificazione incrociata in una croce o in una tabella di contingenza. Ogni cross - classificazione nella tabella costituisce una cella, e ogni variabile categoriale è chiamata fattore. La variabile dipendente è il numero di casi (frequenza) in una cella della crosstabulazione, e le variabili esplicative sono fattori e covariate. Questa procedura stima i parametri di massima verosimiglianza dei modelli loglineari gerarchici e non gerarchici utilizzando il metodo Newton - Raphson. O un Poisson o una distribuzione multinomiale possono essere analizzati.

È possibile selezionare fino a 10 fattori per definire le celle di una tabella. Una variabile struttura cellulare consente di definire zeri strutturali per tabelle incomplete, includere un termine offset nel modello,

adattare un modello di velocità di registrazione o implementare il metodo di regolazione delle tabelle marginali. Le variabili di contrasto consentono il calcolo dei rapporti di log-odds generalizzati (GLOR).

Le informazioni sul modello e le statistiche di bontà vengono visualizzate automaticamente. È anche possibile visualizzare una varietà di statistiche e complotti o salvare i residui e i valori previsti nel dataset attivo.

**Esempio.** I dati di un rapporto di incidenti automobilistici in Florida sono utilizzati per determinare il rapporto tra indossare una cintura di sicurezza e se un infortunio è stato fatale o non fatale. Il rapporto di odds indica evidenza significativa di una relazione.

**Statistiche.** Frequenze osservate e attese; residui di materie prime, regolate e devianze; matrice di design; stime dei parametri; odds ratio; ratio log - odds; GLOR; Wald statistica; intervalli di confidenza. Tracciati: residui regolati, residui di devianza e probabilità normale.

Considerazioni Generali Loglinear Analysis Data

**Dati.** I fattori sono categoriali e le covariate cellulari sono continue. Quando una covariata è nel modello, il valore medio covariato per i casi in una cella viene applicato a quella cella. Le variabili di contrasto sono continue. Sono utilizzati per calcolare i rapporti di log-odds generalizzati. I valori della variabile di contrasto sono i coefficienti per la combinazione lineare dei tronchi della conta delle cellule previste.

Una variabile struttura cellulare assegna pesi. Ad esempio, se alcune delle celle sono zeri strutturali, la variabile della struttura cellulare ha un valore di 0 o 1. Non utilizzare una struttura a struttura cellulare variabile a peso aggregato. Scegliere invece **Casi di peso** dal menu Dati.

**Ipotesi.** Due distribuzioni sono disponibili in General Loglinear Analysis: Poisson e multinomiale.

Sotto l'assunzione di distribuzione di Poisson:

- La dimensione totale del campione non è fissata prima dello studio o l'analisi non è condizionata alla dimensione totale del campione.
- L'evento di un'osservazione che si trova in una cella è statisticamente indipendente dalla conta delle cellule di altre cellule.

Sotto l'assunzione di distribuzione multinomiale:

- La dimensione totale del campione è fissa, oppure l'analisi è condizionata alla dimensione totale del campione.
- I conteggi delle cellule non sono statisticamente indipendenti.

**Procedure correlate.** Usa la procedura Crosstabs per esaminare le crosstabulazioni. Utilizzare la procedura Loglineare di Logit quando è naturale considerare una o più variabili categoriali come variabili di risposta e le altre come variabili esplicative.

Ottenimento di un'analisi generale Loglineare

1. Dai menu, scegliere:

**Analisi > Loglineare > Generale ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Nella finestra di dialogo Analisi Loglineare generale, selezionare fino a 10 variabili di fattore.

Facoltativamente, è possibile:

- Selezionare covariate di celle.
- Selezionare una variabile di struttura cellulare per definire zeri strutturali o includere un termine di offset.
- Selezionare una variabile di contrasto.

## Modello di analisi Loglineare generale

**Specifica modello.** Un modello saturo contiene tutti gli effetti principali e le interazioni che coinvolgono variabili di fattore. Non contiene termini covariati. Selezionare **Personalizzato** per specificare un solo sottoinsieme di interazioni o interazioni dei fattori e covariate.

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

**Termini in 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.

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.

## Creazione di termini e termini personalizzati

### Crea termini

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### Crea termini personalizzati

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## Opzioni generali di analisi Loglineare

La procedura General Loglinear Analysis visualizza le statistiche di modello e le statistiche di bontà. Inoltre, è possibile scegliere uno o più dei seguenti:

**Visualizza.** Sono disponibili diverse statistiche per la visualizzazione - frequenze cellulari osservate e previste; residui di materie prime, regolate e devianze; una matrice di progettazione del modello; e le stime dei parametri per il modello.

**Grafico.** I tracciati, disponibili solo per i modelli personalizzati, includono due matrici di tracciamento (residui regolati o residui di devianza contro i conteggi delle cellule osservate e previste). È inoltre possibile visualizzare la probabilità normale e grafici normali detrendizzati dei residui corretti o della devianza.

**Intervallo di confidenza.** L'intervallo di confidenza per le stime dei parametri può essere regolato.

**Criteri.** Il metodo Newton-Raphson viene utilizzato per ottenere le stime di parametri di massima verosimiglianza. È possibile specificare nuovi valori per il numero massimo di interazioni, il criterio di convergenza e il delta (una costante aggiunta a tutte le celle per le approssimazioni iniziali). Delta rimane nelle celle per i modelli saturi.

## Analisi generale di analisi Loglineare

Selezionare i valori che si desidera salvare come nuove variabili nel dataset attivo. Il suffisso  $n$  nei nuovi nomi delle variabili incrementa per rendere un nome univoco per ogni variabile salvata.

I valori salvati si riferiscono ai dati aggregati (celle nella tabella di contingenza), anche se i dati sono registrati nelle singole osservazioni nell'editor dei dati. Se si salvano i residui o i valori previsti per i dati non aggregati, il valore salvato per una cella nella tabella di contingenza viene inserito nell'Editor dei dati per ogni caso in quella cella. Per dare un senso ai valori salvati, è necessario aggregare i dati per ottenere la conta delle cellule.

Quattro tipi di residui possono essere salvati: crudi, standardizzati, regolati e devianti. I valori previsti possono anche essere salvati.

- *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 relativo errore standard. I residui standardizzati sono anche noti come residui di Pearson.
- *Residui corretti*. Il residuo standardizzato diviso per il suo errore standard stimato. Poiché i residui adattati sono asintoticamente normali standard quando il modello selezionato è corretto, sono preferiti ai residui standardizzati per la verifica della normalità.
- *Residui di devianza*. La radice quadrata con segno di un contributo individuale alla statistica chi - quadrato del rapporto di verosimiglianza (G quadrato), dove il segno è il segno del residuo (conteggio osservato meno conteggio atteso). I residui di devianza hanno una distribuzione della normale standard asintotica.

## Funzioni Aggiuntive Di Comando GENLOG

Il linguaggio della sintassi dei comandi consente inoltre di:

- Calcolare combinazioni lineari di frequenze cellulari osservate e frequenze delle celle previste e residui di stampa, residui standardizzati e residui regolati di quella combinazione (utilizzando il comando GERESID).
- Modificare il valore di soglia predefinito per la verifica della ridondanza (utilizzando il comando CRITERIA).
- Visualizzare i residui standardizzati (utilizzando il sottocomando PRINT).

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

## Analisi loglineare logit

---

La Procedura Logit Loglinear Analysis analizza il rapporto tra variabili dipendenti (o risposta) e variabili indipendenti (o esplicative). Le variabili dipendenti sono sempre categoriali, mentre le variabili indipendenti possono essere categoriali (fattori). Altre variabili indipendenti (covariate cellulari) possono essere continue, ma non vengono applicate caso per caso. La media covariata ponderata per una cella viene applicata a quella cella. Il logaritmo della probabilità a favore delle variabili dipendenti viene espresso come combinazione lineare di parametri. Si assume automaticamente una distribuzione multinomiale; questi modelli sono talvolta chiamati modelli logit multinazionali. 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 fattore combinate. Una variabile struttura cellulare consente di definire zeri strutturali per tabelle incomplete, includere un termine offset nel modello, adattare un modello di velocità di registrazione o implementare il metodo di regolazione delle tabelle marginali. Le variabili di contrasto consentono il calcolo dei rapporti di log-odds generalizzati (GLOR). I valori della variabile di contrasto sono i coefficienti per la combinazione lineare dei tronchi della conta delle cellule previste.

Le informazioni sul modello e le statistiche di bontà vengono visualizzate automaticamente. È anche possibile visualizzare una varietà di statistiche e complotti o salvare i residui e i valori previsti nel dataset attivo.

**Esempio.** Uno studio in Florida ha incluso 219 alligatori. In che modo il tipo alimentare degli alligatori varia con le loro dimensioni e i quattro laghi in cui vivono? Lo studio ha scoperto che le probabilità che un alligatore più piccolo preferisca i rettili ai pesci sono 0.70 volte inferiori rispetto agli alligatori più grandi; inoltre, le probabilità di selezionare principalmente i rettili invece dei pesci erano più alte nel lago 3.

**Statistiche.** Frequenze osservate e attese; residui di materie prime, regolate e devianze; matrice di design; stime dei parametri; rapporto di registrazione generalizzato; statistica Wald; e intervalli di confidenza. Tracciati: residui regolati, residui di devianza e complotti di probabilità normali.

Logit Loglinear Analysis Data Considerazioni

**Dati.** Le variabili dipendenti sono categoriali. I fattori sono categoriali. Le covariate di cella possono essere continue, ma quando una covariata è nel modello, il valore medio covariato per i casi in una cella viene applicato a quella cella. Le variabili di contrasto sono continue. Sono utilizzati per calcolare i rapporti di scostamento generalizzati (GLOR). I valori della variabile di contrasto sono i coefficienti per la combinazione lineare dei tronchi della conta delle cellule previste.

Una variabile struttura cellulare assegna pesi. Ad esempio, se alcune delle celle sono zeri strutturali, la variabile della struttura cellulare ha un valore di 0 o 1. Non utilizzare una struttura a struttura cellulare variabile per il peso dei dati aggregati. Utilizzare invece i casi di Peso sul menu Dati.

**Ipotesi.** Si ipotizza che i conteggi all'interno di ciascuna combinazione di categorie di variabili esplicative abbiano una distribuzione multinomiale. Sotto l'assunzione di distribuzione multinomiale:

- La dimensione totale del campione è fissa, oppure l'analisi è condizionata alla dimensione totale del campione.
- I conteggi delle cellule non sono statisticamente indipendenti.

**Procedure correlate.** Utilizzare la procedura Crosstabs per visualizzare le tabelle di contingenza. Utilizzare la procedura di Analisi Loglineare generale quando si desidera analizzare il rapporto tra un conteggio osservato e una serie di variabili esplicative.

Ottenimento di un'analisi Loglineare Logit

1. Dai menu, scegliere:

**Analizzare > Loglinear > Logit ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Nella finestra di dialogo Analisi Loglineare Logit selezionare una o più variabili dipendenti.  
3. Selezionare una o più variabili fattore.

Il numero totale di variabili dipendente e fattore deve essere inferiore o uguale a 10.

Facoltativamente, è possibile:

- Selezionare covariate di celle.
- Selezionare una variabile di struttura cellulare per definire zeri strutturali o includere un termine di offset.
- Selezionare una o più variabili di contrasto.

## Modello di analisi Loglineare logit

**Specifica modello.** Un modello saturo contiene tutti gli effetti principali e le interazioni che coinvolgono variabili di fattore. Non contiene termini covariati. Selezionare **Personalizzato** per specificare un solo sottoinsieme di interazioni o interazioni dei fattori e covariate.

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

**Termini in 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.

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.

I termini vengono aggiunti al design prendendo tutte le combinazioni possibili dei termini dipendenti e corrispondenti ad ogni combinazione con ogni termine nella lista dei modelli. Se **Includi costante per dipendente** viene selezionato, c'è anche un termine di unità (1) aggiunto alla lista dei modelli.

Ad esempio, supporre le variabili  $D1$  e  $D2$  sono le variabili dipendenti. Un elenco di termini dipendenti è creato dalla procedura di Analisi Loglineare di Logit ( $D1$ ,  $D2$ ,  $D1*D2$ ). Se le Termini in elenco Modello contengono  $M1$  e  $M2$  e una costante è inclusa, la lista dei modelli contiene 1,  $M1$  e  $M2$ . Il design risultante include combinazioni di ogni termine modello con ogni termine dipendente:

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

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

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

**Includi costante per dipendente.** Include una costante per la variabile dipendente in un modello personalizzato.

## **Creazione di termini e termini personalizzati**

### **Crea termini**

Utilizzare questa opzione quando si desidera includere termini non nidificati di un certo tipo (ad esempio, effetti principali) per tutte le combinazioni di una serie selezionata di fattori e covariate.

### **Crea termini personalizzati**

Utilizzare questa opzione quando si desidera includere termini nidificati o creare esplicitamente una variabile termine dalla variabile. La creazione di un termine nidificato richiede le seguenti operazioni:

## **Opzioni di analisi Loglineare logit**

La Procedura Logit Loglinear Analysis visualizza le informazioni sul modello e le statistiche di bontà. Inoltre, è possibile scegliere una o più delle seguenti opzioni:

**Visualizza.** Sono disponibili diverse statistiche per la visualizzazione: frequenze delle celle osservate e previste; residui di materie prime, regolate e devianze; una matrice di progettazione del modello; e le 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 delle celle osservati e previsti). È inoltre possibile visualizzare la probabilità normale e grafici normali detrendizzati dei residui corretti o della devianza.

**Intervallo di confidenza.** L'intervallo di confidenza per le stime dei parametri può essere regolato.

**Criteri.** Il metodo Newton-Raphson viene utilizzato per ottenere le stime di parametri di massima verosimiglianza. È possibile specificare nuovi valori per il numero massimo di interazioni, il criterio di convergenza e il delta (una costante aggiunta a tutte le celle per le approssimazioni iniziali). Delta rimane nelle celle per i modelli saturi.

## Analisi Logit Loglineare Salva

Selezionare i valori che si desidera salvare come nuove variabili nel dataset attivo. Il suffisso  $n$  nei nuovi nomi delle variabili incrementa per rendere un nome univoco per ogni variabile salvata.

I valori salvati si riferiscono ai dati aggregati (alle celle nella tabella di contingenza), anche se i dati sono registrati nelle singole osservazioni nell'editor dei dati. Se si salvano i residui o i valori previsti per i dati non aggregati, il valore salvato per una cella nella tabella di contingenza viene inserito nell'Editor dei dati per ogni caso in quella cella. Per dare un senso ai valori salvati, è necessario aggregare i dati per ottenere la conta delle cellule.

Quattro tipi di residui possono essere salvati: crudi, standardizzati, regolati e deviati. I valori previsti possono anche essere salvati.

- *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 relativo errore standard. I residui standardizzati sono anche noti come residui di Pearson.
- *Residui corretti.* Il residuo standardizzato diviso per il suo errore standard stimato. Poiché i residui adattati sono asintoticamente normali standard quando il modello selezionato è corretto, sono preferiti ai residui standardizzati per la verifica della normalità.
- *Residui di devianza.* La radice quadrata con segno di un contributo individuale alla statistica  $\chi^2$  - quadrato del rapporto di verosimiglianza (G quadrato), dove il segno è il segno del residuo (conteggio osservato meno conteggio atteso). I residui di devianza hanno una distribuzione della normale standard asintotica.

## Funzioni Aggiuntive Di Comando GENLOG

Il linguaggio della sintassi dei comandi consente inoltre di:

- Calcolare combinazioni lineari di frequenze cellulari osservate e frequenze delle celle previste, e residui di stampa, residui standardizzati e residui regolati di quella combinazione (utilizzando il comando GERESID).
- Modificare il valore di soglia predefinito per la verifica della ridondanza (utilizzando il comando CRITERIA).
- Visualizzare i residui standardizzati (utilizzando il sottocomando PRINT).

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

## Tablelle vita

---

Ci sono molte situazioni in cui si vorrebbe esaminare la distribuzione dei tempi tra due eventi, come la durata dell'occupazione (tempo tra essere assunti e lasciare l'azienda). Tuttavia, questo tipo di dati di solito include alcuni casi per i quali il secondo evento non viene registrato (ad esempio, le persone lavorano ancora per l'azienda alla fine dello studio). Questo può accadere per diversi motivi: per alcuni casi, l'evento semplicemente non si verifica prima della fine dello studio; per altri casi, perdiamo traccia del loro status qualche volta prima della fine dello studio; ancora altri casi potrebbero non essere in grado di continuare per motivi non correlati allo studio (come un dipendente che si ammala e si prende un congedo di assenza). Collettivamente, tali casi sono noti come **casi censurati**, e rendono questo tipo di studio inappropriato per le tecniche tradizionali come i test  $t$  o la regressione lineare.

Una tecnica statistica utile per questo tipo di dati è chiamata follow-up **Tabella vita**. L'idea di base della tavola vita è quella di suddividere il periodo di osservazione in intervalli di tempo più piccoli. Per ogni intervallo, tutte le persone che sono state osservate almeno così a lungo vengono utilizzate per calcolare la probabilità che si verifichi un evento terminale in quell' intervallo. Le probabilità stimate da ciascuno degli intervalli vengono poi utilizzate per stimare la probabilità complessiva dell'evento che si verifica in diversi punti temporali.

**Esempio.** Una nuova terapia della patch nicotina è migliore rispetto alla tradizionale terapia delle patch nell'aiutare le persone a smettere di fumare? Si potrebbe condurre uno studio utilizzando due gruppi di fumatori, uno dei quali ha ricevuto la tradizionale terapia e l'altro dei quali ha ricevuto la terapia sperimentale. Costruire tabelle di vita dai dati permetterebbe di confrontare i tassi di astinenza complessiva tra i due gruppi per determinare se il trattamento sperimentale è un miglioramento rispetto alla terapia tradizionale. È inoltre possibile tracciare le funzioni di sopravvivenza o di pericolo e confrontarle visivamente per informazioni più dettagliate.

**Statistiche.** Numero ingresso, numero in uscita, numero esposto a rischio, numero di eventi terminali, proporzione terminazione, proporzione sopravvissuta, proporzione cumulativa sopravvissuta (e errore standard), densità di probabilità (e errore standard), e tasso di pericolo (e errore standard) per ogni intervallo di tempo per ciascun gruppo; tempo di sopravvivenza mediano per ciascun gruppo; e Wilcoxon (Gehan) test per confrontare le distribuzioni di sopravvivenza tra gruppi. Tracciati: complotti di funzioni per la sopravvivenza, la sopravvivenza log, la densità, la velocità di pericolo e una sola meno sopravvivenza.

Considerazioni Dati Tabelle di vita

**Dati.** La tua variabile di tempo dovrebbe essere quantitativa. La tua variabile di stato dovrebbe essere dicotomica o categoriale, codificata come numeri interi, con gli eventi codificati come un valore unico o una gamma di valori consecutivi. Le variabili del fattore dovrebbero essere categoriali, codificate come numeri interi.

**Ipotesi.** Le probabilità per l'evento di interesse dovrebbero dipendere solo dal tempo dopo l'evento iniziale - si ipotizza che siano stabili rispetto al tempo assoluto. Ovvero, i casi che entrano nello studio in momenti diversi (ad esempio, i pazienti che iniziano il trattamento in momenti diversi) dovrebbero comportarsi analogamente. Non ci dovrebbero essere anche differenze sistematiche tra casi censurati e non censurati. Se, ad esempio, molti dei casi censurati sono pazienti con condizioni più gravi, i vostri risultati possono essere biased.

**Procedure correlate.** La procedura Life Tables utilizza un approccio attuariale a questo tipo di analisi (nota generalmente come Survival Analysis). La Procedura Kaplan - Meier Survival Analysis utilizza un metodo leggermente diverso di calcolo delle tabelle di vita che non si affida al partizionamento del periodo di osservazione in intervalli di tempo minori. Questo metodo è consigliato se si dispone di un numero ridotto di osservazioni, tale che ci sarebbe solo un piccolo numero di osservazioni in ogni intervallo di tempo di sopravvivenza. Se si dispone di variabili che si sospetta siano correlate al tempo di sopravvivenza o di variabili che si desidera controllare (covariate), utilizzare la procedura Cox Regression. Se le covariate possono avere valori diversi in momenti diversi dello stesso caso, utilizzare Cox Regression with Time-Dependent Covariates.

Creazione delle tabelle di vita

1. Dai menu, scegliere:

**Analizzare > Sopravvivenza > Tavole di vita ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di sopravvivenza *numerica* .

3. Specificare gli intervalli di tempo da esaminare.

4. Selezionare una variabile di stato per definire i casi per i quali si è verificato l'evento terminale.

5. Clicca su **Definisci evento** per specificare il valore della variabile di stato che indica che si è verificato un evento.

Facoltativamente, è possibile selezionare una variabile del fattore di primo ordine. Le tabelle attuariali per la variabile di sopravvivenza sono generate per ogni categoria della variabile fattore.

È inoltre possibile selezionare una variabile di secondo ordine *per fattore*. Le tabelle attuariali per la variabile di sopravvivenza sono generate per ogni combinazione delle variabili fattore di primo e secondo ordine.

## Tabelle di vita Definisci intervallo

I casi con valori per la variabile fattore nell'intervallo che si specificano verranno inclusi nell'analisi e verranno generate tabelle separate (e grafici, se richiesto) per ogni valore univoco nell'intervallo.

## Opzioni tabelle di vita

È possibile controllare vari aspetti dell'analisi di Life Tables.

**Tabella di vita.** Per sopprimere la visualizzazione delle tabelle di vita nell'output, deselezionare **Life table (s)**.

**Grafico.** Consente di richiedere complotti delle funzioni di sopravvivenza. Se si hanno definiti variabili del fattore, i tracciati vengono generati per ogni sottogruppo definito dalla variabile fattore (s). I tracciati disponibili sono la sopravvivenza, la sopravvivenza log, il pericolo, la densità e uno meno la sopravvivenza.

- *Sopravvivenza.* Visualizza la funzione di sopravvivenza cumulata in scala lineare.
- *sopravvivenza del log.* Visualizza la funzione di sopravvivenza cumulata in scala logaritmica.
- *Rischio.* Visualizza la funzione di rischio cumulato in scala lineare.
- *Densità.* Visualizza la funzione di densità.
- *Uno meno sopravvivenza.* Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.

**Confronta Livelli di Primo Fattore.** Se hai una variabile di controllo di primo ordine, puoi selezionare una delle alternative in questo gruppo per eseguire il test di Wilcoxon (Gehan), che confronta la sopravvivenza dei sottogruppi. I test vengono eseguiti sul fattore di primo ordine. Se hai definito un fattore di secondo ordine, vengono eseguiti dei test per ogni livello della variabile di secondo ordine.

## SURVIVAL Comando Funzioni Aggiuntive

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare più di una variabile dipendente.
- Specificare intervalli non altrettanto spaziati.
- Specificare più di una variabile di stato.
- Specificare i confronti che non includono tutto il fattore e tutte le variabili di controllo.
- Calcolare approssimativi, piuttosto che esatti, confronti.

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

## Modelli Temporal Accelerati Parametrici accelerati

A Parametric Accelerated Failure Time (AFT) L'analisi del modello richiama la procedura dei modelli di sopravvivenza parametrica con dati di vita non ricorrenti. I modelli di sopravvivenza parametrica ipotizza che il tempo di sopravvivenza segua una distribuzione conosciuta e questa analisi si adatta a modelli di tempo di guasto accelerato con i loro effetti modello proporzionale rispetto al tempo di sopravvivenza.

### Ottenimento di un'analisi Parametric Accelerated Failure Time Models

1. Dai menu, scegliere:

**Analizzare > Sopravvivenza > Modelli di guasto accelerato parametrico (aFT)**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di origine.

#### **Ora**

##### **Sopravvivenza**

Variabile numerica singola che denota la durata del tempo di sopravvivenza.

##### **Inizio/Fine**

Variabili numeriche che denotano **Start Time** e **Ora di fine**.

#### **Stato**

Singola stringa opzionale o variabile numerica che determina una delle seguenti impostazioni di stato:

##### **Errore/Evento**

Associa un record a una categoria fallimento / evento. Il valore predefinito per una variabile di stato stringa è F.

##### **Censura a destra**

Associa un record ad una categoria di censura giusta. Il valore predefinito per una variabile di stato stringa è R.

##### **Censura a sinistra**

Associa un record a una categoria di censura sinistra. Il valore predefinito per una variabile di stato stringa è L.

##### **Censura intervallo**

Associa un record a una categoria di censura dell'intervallo. Solo per **Start / Fine** . Il valore predefinito per una variabile di stato stringa è I.

##### **Trattamento dei valori non mappati**

Controlla a quale categoria associare i record non mappati. Per eliminare i record che non sono riusciti ad essere mappati, selezionare **Escludi dall'analisi**.

Per **Sopravvivenza**, lo stato predefinito per tutti i casi è **Failure / Event**. Per **Start / Fine**, lo stato predefinito è **Interval Censoring**. Clicca sul pulsante **Definisci evento** per definire un evento per la variabile di stato.

#### **Covariate**

Una o più variabili numeriche facoltative da trattare come covariate. Si noti che una variabile non può essere specificata sia da **Covariate (s)** che **Fixed Factor (s)**.

#### **Fattore(i) fisso(i)**

Una o più variabili facoltative da trattare come fattori. Una variabile non può essere specificata sia da **Fattore fisso (s)** che **Covariati**.

#### **Troncamento a sinistra**

Singola variabile numerica opzionale per il tronco di sinistra solo per **Sopravvivenza** .

## **Modelli parametrici di tempo di guasto accelerato: Criteri**

### **Criteri**

Un pannello opzionale per specificare i criteri generali.

#### **Intervallo di confidenza**

Una percentuale opzionale per specificare il livello per gli intervalli di confidenza dei parametri di regressione. Deve essere un valore numerico unico compreso tra 0 e 100. Il valore predefinito è 95.

#### **Valori mancanti**

Un'opzione per controllare come vengono trattati i valori mancanti dell'utente:

##### **Escludi valori mancanti di sistema e definiti dall'utente**

Cura i valori mancanti dell'utente come valori validi. È l'impostazione predefinita.

**I valori mancanti definiti dall'utente vengono considerati come validi**

Ignora le designazioni di valore mancanti dell'utente e li tratta come valori validi.

**Trattamento stato**

Solo per **Start / Fine** . Un'opzione per controllare come affrontare i record con campi di stato errati:

**Elimina record conflittuali**

Elimina i record conflittuali. Questa è l'impostazione predefinita.

**Ottieni le informazioni sul tempo in base allo stato**

Ottiene le informazioni del tempo in base allo stato.

**Deriva lo stato in base alle informazioni sul tempo**

Cambia lo stato in base alle informazioni sul tempo.

## Parametric Accelerated Failure Time Models: Modello

**Modello**

Un pannello opzionale per specificare le opzioni e le impostazioni del modello.

**Distribuzione del tempo di sopravvivenza**

Un'opzione per specificare la distribuzione del tempo di sopravvivenza.

**Weibull**

Specifica la distribuzione di Weibull. Questa è l'impostazione predefinita.

**Esponenziale**

Specifica la distribuzione esponenziale.

**Logaritmica normale**

Specifica la distribuzione log - normale.

**Logaritmica logistica**

Specifica la distribuzione log - logistica.

**Impostazioni covariata**

Specificare le variabili covariate.

**Impostazioni fattore**

Specificare le variabili del fattore.

**Valore iniziale di intercept**

Un'opzione per specificare il valore iniziale del termine di intercettazione. Se specificato, deve essere un valore numerico unico e non può essere 0.

**Valore iniziale del parametro di scala**

Un'opzione per controllare l'impostazione del parametro di scala.

**Errore standard della regressione OLS corrispondente**

Utilizza l'errore standard della regressione ordinaria dei minimi quadrati come valore iniziale.

**Inverte errore standard della regressione OLS corrispondente**

Utilizza il reciproco dell'errore standard.

**Valore fornito dall'utente**

Se viene specificato un solo valore numerico, il valore viene utilizzato come valore iniziale. Se specificato, deve essere maggiore di 0.

## Parametric Accelerated Failure Time Models: Preventivo

**Stima**

Un pannello opzionale per specificare le impostazioni per controllare la stima dei modelli di tempo di guasto accelerato e il processo di selezione delle funzioni opzionali.

## Alternating Direction Method o Multipinze (ADMM)

### Rapido

Applica il metodo di direzione alternata veloce dei moltiplicatori (ADMM). È l'impostazione predefinita.

### Tradizionale

Applica l'algoritmo tradizionale ADMM.

## Applica regolarizzazione L-1

Conduce il processo per controllare la selezione delle funzioni. Il campo **Penalty Parametro** specifica il parametro di penalità che controlla il processo di regolarizzazione. Deve essere un valore unico superiore a 0. L'impostazione predefinita è 0.001.

## Criteri di convergenza del modello

### Convergenza parametri

Specifica i criteri di convergenza per il parametro. Deve essere un valore numerico unico appartenente a  $[0, 1)$ . L'impostazione predefinita è 0.000001. Per **Tipo** è possibile selezionare **ABSOLUTE** per applicare la convergenza assoluta all'ottimizzazione interna o **RELATIVE** per applicare la relativa convergenza all'ottimizzazione interna. Il campo **Valore** opzionale specifica una parola chiave.

### Convergenza funzione obiettivo

Specifica i criteri di convergenza per la funzione oggettiva. Deve essere un valore numerico unico appartenente a  $[0, 1)$ . L'impostazione predefinita è 0, che non applica i criteri di convergenza. Per **Tipo** è possibile selezionare **ABSOLUTE** per applicare la convergenza assoluta all'ottimizzazione interna o **RELATIVE** per applicare la relativa convergenza all'ottimizzazione interna. Il campo **Valore** opzionale specifica una parola chiave.

### Convergenza hessiana

Specifica i criteri di convergenza per la matrice hessiana. Deve essere un valore numerico unico appartenente a  $[0, 1)$ . L'impostazione predefinita è 0, che non applica i criteri di convergenza. Per **Tipo** è possibile selezionare **ABSOLUTE** per applicare la convergenza assoluta all'ottimizzazione interna o **RELATIVE** per applicare la relativa convergenza all'ottimizzazione interna. Il campo **Valore** opzionale specifica una parola chiave.

## Criteri di convergenza residui

Un'opzione per controllare il processo di ottimizzazione.

### Residuo primitivo e duale

Applica il criterio di convergenza sia primitivo che duplice. Questa è l'impostazione predefinita.

### Solo residuo primitivo

Applica il criterio di convergenza residuo primitivo.

### Solo residuo duale

Applica il duplice criterio di convergenza residuo.

## Metodo

Un parametro opzionale per specificare il metodo di stima.

### Automatico

Sceglie automaticamente il metodo in base al dataset di esempio. È l'impostazione predefinita. Il campo **Numero di predittori** specifica la soglia del numero di predittori e deve essere un intero intero maggiore di 1. Il valore predefinito è 1000.

### Newton-Raphson

Applica il metodo di Newton - Raphson.

### L-BFGS

Applica l'algoritmo BFGS a memoria limitata. Il campo **Aggiornamento** specifica il numero di aggiornamenti passati mantenuto dall'algoritmo BFGS a memoria limitata e deve essere un intero intero maggiore o uguale a 1. Il valore predefinito è 5.

## Iterazione

### **Numero massimo di iterazioni**

Specifica il numero massimo di iterazioni. Deve essere un intero appartenente a [ 1, 100]. L'impostazione predefinita è 20.

### **Numero massimo di dimezzamenti**

Specifica il numero massimo di dimezzamento. Deve essere un intero appartenente a [ 1, 20]. L'impostazione predefinita è 5.

### **Numero massimo di ricerche di linea**

Specifica il numero massimo di ricerche di linea. Deve essere un intero appartenente a [ 1, 100]. L'impostazione predefinita è 20.

### **Convergenza assoluta per il processo di iterazione**

Specifica la convergenza assoluta per il processo di iterazione esterna. Deve essere un valore numerico unico appartenente a (0, 1). L'impostazione predefinita è 0.0001.

### **Convergenza relativa per il processo di iterazione**

Specifica la relativa convergenza per il processo di iterazione esterna. Deve essere un valore numerico unico appartenente a (0, 1). L'impostazione predefinita è 0.01.

## Parametric Accelerated Failure Time Models: Stampa

### Stampa

Un pannello opzionale per controllare le uscite della tabella.

#### **Dettagli di codifica del fattore**

Se selezionato, visualizza e stampa i dettagli di codifica dei fattori. Il processo viene ignorato se non ci sono fattori in vigore.

#### **Valori iniziali assegnati ai parametri di regressione**

Se selezionato, visualizza i valori iniziali utilizzati nel processo di stima.

#### **Cronologia iterazioni modello**

Se selezionato, visualizza la cronologia di iterazione dell'analisi di sopravvivenza. Nel campo **Numero di passi** specificare il numero di passi compresi tra 1 e 99999999. L'impostazione predefinita è 1.

#### **Risultati selezione contenenti**

Controlla la visualizzazione dei dettagli della selezione della funzione.

#### **Sia variabili selezionate che non selezionate**

Visualizza nella tabella entrambe le variabili selezionate e non selezionate.

#### **Solo variabili selezionate**

Solo visualizzare le variabili selezionate.

#### **Solo variabili non selezionate**

Solo visualizzare le variabili non selezionate. Il campo **Massimo variabili da visualizzare** specifica il numero massimo delle variabili stampate nella tabella. L'impostazione predefinita è 30.

## Parametric Accelerated Failure Time Models: Prevedere

### Previsione

Un pannello opzionale per segnare e salvare le statistiche previste al dataset attivo.

#### **Valori temporali per il calcolo del punteggio**

##### **Valori temporali definiti da variabili dipendenti**

Punteggi le **Previsioni** in base alla variabile temporale specificata per il modello di sopravvivenza parametrica.

### **Intervalli regolari**

Punteggi le **Previsioni** in base ai valori temporali futuri. Il campo **Intervallo di tempo** specifica l'intervallo di tempo, e deve essere un valore numerico singolo maggiore di 0. Il campo **Numero di periodi temporali** specifica il numero dei periodi di tempo, e deve essere un intero numero intero numerico compreso tra 2 e 100.

### **Durata**

Punta le **Previsioni** in base alla durata del tempo per definire i valori temporali futuri. Deve essere una singola variabile numerica.

## **Previsioni**

### **Sopravvivenza**

Punteggi e salva le statistiche di sopravvivenza previste al dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredSurvival`.

### **Rischio**

Punteggi e salva i pericoli previsti per il dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredHazard`.

### **Rischio cumulativo**

Punteggi e salva i rischi cumulativi previsti per il dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredCumHazard`.

### **Sopravvivenza condizionale**

Punteggi e salva le statistiche di sopravvivenza condizionata previste al dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredConditionalSurvival`. Il processo verrà ignorato se `PASTTIME` non è specificato. È richiesto un valore **Tempo di sopravvivenza passato** e specifica i valori di tempo passato per il calcolo del punteggio. Deve essere una singola variabile numerica.

## **Modelli parametrici di tempo di guasto accelerato: Tracciato**

### **Grafico**

#### **Grafico dei residui di Cox-Snell**

Selezionare **Visualizza il grafico** per creare un grafico dei residui Cox -Snell. Nel campo **Numero di punti di taglio del binning** specificare un numero da 1 a 10000. L'impostazione predefinita è 100.

#### **Grafici di funzione**

Un'opzione per controllare i tracciati delle funzioni.

#### **Tipo**

##### **Sopravvivenza**

Crea la trama per le funzioni di sopravvivenza.

##### **Rischio**

Crea la trama per le funzioni di pericolo.

##### **Densità**

Crea una trama per le funzioni di densità.

#### **Numero di punti da visualizzare**

Specifica il numero di punti funzione tra 1 e 200. L'impostazione predefinita è 100.

#### **Valori di covariata del grafico**

Un optional per specificare i valori forniti dall'utente e assegnarli ai predittori. Per impostazione predefinita verranno creati i complotti designati al **Mean** di ciascuna covariata in vigore e la frequenza di categoria di ogni fattore in vigore. Se specificato, verranno creati i complotti designati in base all'impostazione del pattern. In presenza di eventuali variabili duplicate, quello specificato per primo verrebbe riconosciuto e il resto verrebbe ignorato. Una variabile valida deve essere contenuta in un effetto modello. Per una covariata, il valore fornito dall'utente deve essere numerico. L'omissione di una variabile in vigore indica che la frequenza di categoria e il **Mean**

sarebbero utilizzati per impostazione predefinita per il fattore e la covariata, rispettivamente. Se un valore non valido viene assegnato ad una variabile, il modello richiesto non verrà tracciato.

### Valori dei fattori del grafico

Un optional per specificare i valori forniti dall'utente e assegnarli ai predittori. In presenza di eventuali variabili duplicate, quello specificato per primo verrebbe riconosciuto e il resto verrebbe ignorato. Una variabile valida deve essere contenuta in un effetto modello. L'omissione di una variabile in vigore indica che la frequenza di categoria e la media sarebbero utilizzate per default rispettivamente per il fattore e la covariata. Se un valore non valido viene assegnato ad una variabile, il modello richiesto non verrà tracciato.

### Separa linee per

Un'opzione per specificare una variabile categoriale tramite la quale verranno tracciati i tracciati di linea.

### Numero massimo di righe in un grafico

Specifica il numero massimo delle righe in un grafico se si specifica **Linee Separate per**. L'impostazione predefinita è 10.

## Modelli parametrici di tempo di guasto accelerato: Esportazione

### Esporta

Selezionare **Esporta informazioni modello su file XML** per scrivere il modello e le informazioni dei parametri ad un file PMML per il punteggio. È necessario specificare la directory e il nome file del file PMML da salvare.

## Eventi di definizione AFT di sopravvivenza per variabili di stato

Le ricorrenze del valore o dei valori selezionati per la variabile di stato indicano che l'evento terminale si è verificato per tali casi. Tutti gli altri casi sono considerati censurati. Immettere un valore singolo o un intervallo di valori che identifichino l'evento di interesse.

## Modelli di tempo di errore accelerati parametrici: Seleziona categoria

L'impostazione Seleziona categoria fornisce le opzioni per scegliere un valore che denota la categoria da modellare come baseline per il confronto.

Selezione della categoria

Fare clic su 'Ultima categoria' per aprire la finestra di dialogo 'Seleziona categoria'.

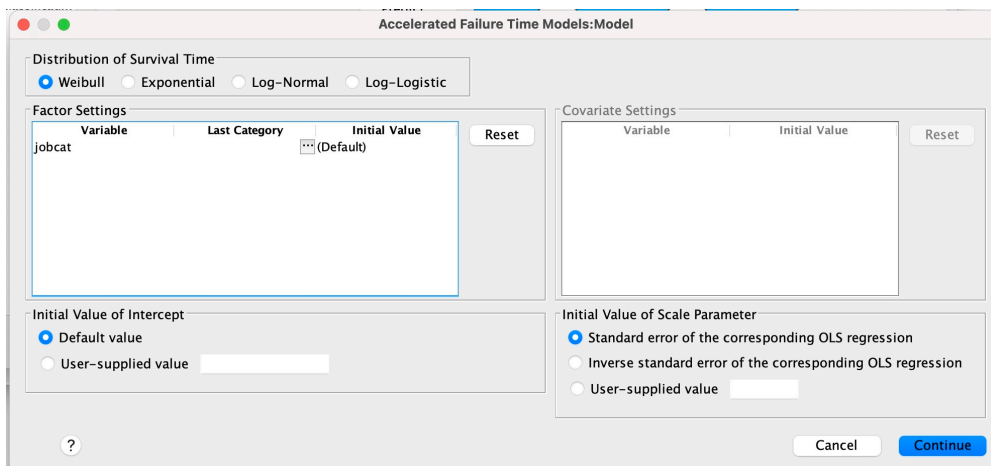


Figura 3. Modelli di durata accelerata - Finestra di dialogo - Categoria

Per designare una categoria come baseline, selezionare un valore dalla casella di dialogo 'Seleziona categoria'.

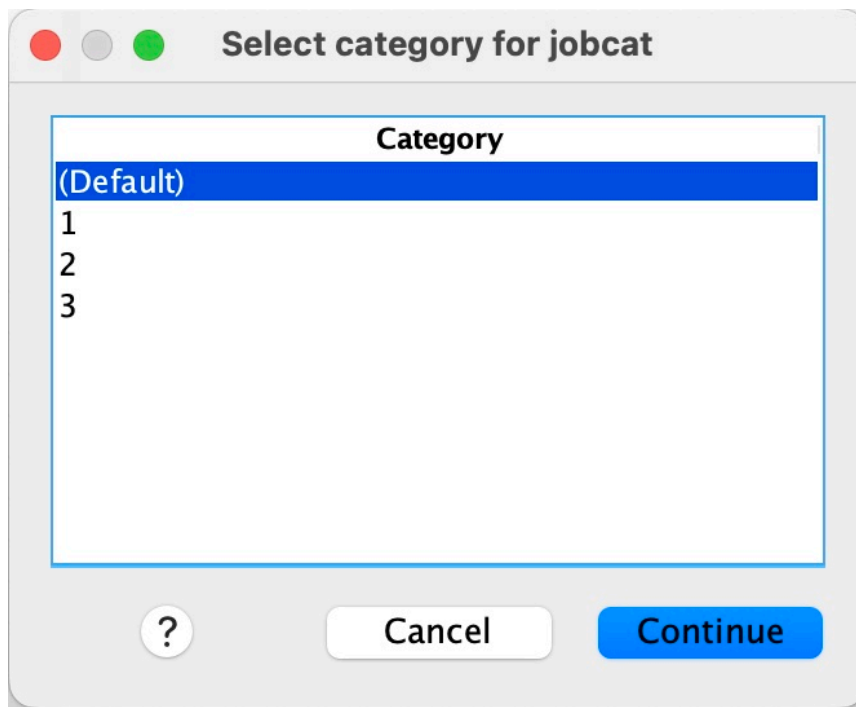


Figura 4. Modelli di durata accelerata - Finestra di dialogo - Seleziona categoria

Fare clic su Continua.

## Modelli di fragilità condivisa parametrici

Un'analisi di sopravvivenza dei modelli di fragilità condivisa parametrica avvia la procedura dei modelli di sopravvivenza parametrica con input di dati di durata ricorrente. I modelli di sopravvivenza parametrica presuppongono che il tempo di sopravvivenza segua una distribuzione nota, e questa analisi incorpora un termine di fragilità in un modello di sopravvivenza parametrica. Viene trattato come una componente casuale per tenere conto di un effetto non osservato dovuto alla variabilità individuale o di gruppo.

### Come ottenere un'analisi dei modelli di Frailty condivisi parametrici

1. Dal menu, scegliere:

**Analisi > Sopravvivenza > Modelli di fragilità condivisa parametrica**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile origine.

#### Ora

##### Sopravvivenza

Il tempo di sopravvivenza è rappresentato da una variabile per indicare l'ora di fine. L'ora di avvio è impostata su 0.

##### Inizio / Fine

Variabili numeriche che denotano **Ora di inizio** e **Ora di fine**.

#### Oggetto

Richiesto per eseguire la procedura. Specifica una singola variabile per l'ID soggetto.

#### Intervallo

Specifica una variabile singola e numerica per il numero di intervallo utilizzato per identificare i diversi record ricorrenti che condividono lo stesso ID oggetto.

## **Stato**

Singola stringa facoltativa o variabile numerica che determina una delle seguenti impostazioni di stato:

### **Errore/Evento**

Associa un record a una categoria di errore / evento. Il valore predefinito per una variabile di stato stringa è F.

### **Censura a destra**

Associa un record a una categoria di censura di destra. Il valore predefinito per una variabile di stato stringa è R.

### **Trattamento dei valori non mappati**

Controlla la categoria a cui associare i record non associati. Per eliminare i record che non sono stati associati, selezionare **Escludi dall'analisi**.

Fare clic sul pulsante **Definisci evento** per definire un evento per la variabile stato.

## **Covariate**

Una o più variabili numeriche facoltative da considerare come covariate. Si noti che una variabile non può essere specificata sia da **Covariate** che da **Fattori fissi**.

## **Fattore(i) fisso(i)**

Una o più variabili facoltative da considerare come fattori. Una variabile non può essere specificata sia da **Fattori fissi** che da **Covariate**.

# **Modelli di fragilità condivisa parametrica: criteri**

## **Criteri**

Un pannello facoltativo per specificare i criteri generali.

### **Intervallo di confidenza**

Una percentuale facoltativa per specificare il livello degli intervalli di confidenza dei parametri di regressione. Deve essere un valore numerico singolo compreso tra 0 e 100. Il valore predefinito è 95.

### **Livello di significatività**

Un'opzione per specificare il livello di significatività del test del rapporto di verosimiglianza per il componente di fragilità. Deve essere un valore numerico singolo compreso tra 0 e 1. L'impostazione predefinita è 0.05.

### **Valori mancanti**

Un'opzione per controllare come vengono trattati i valori mancanti definiti dall'utente:

#### **Escludi valori mancanti di sistema e definiti dall'utente**

Considera i valori mancanti definiti dall'utente come valori validi. È l'impostazione predefinita.

#### **I valori mancanti definiti dall'utente vengono considerati come validi**

Ignora le designazioni dei valori mancanti definiti dall'utente e li considera come valori validi.

### **Trattamento intervallo**

Un'opzione per controllare come gestire i record il cui intervallo è in conflitto con l'ora di inizio e di fine. Ha effetto se ci sono due variabili di tempo con una variabile di intervallo specificata nella finestra di dialogo principale.

#### **Elimina i record in conflitto**

Elimina tutti i record seriali dell'oggetto se il valore dell'intervallo è in conflitto con l'ora di inizio e di fine. Questa è l'impostazione predefinita.

#### **Rileva i valori dell'intervallo in base all'ora di inizio e fine**

Rileva il valore intervallo dall'ora di inizio e di fine.

## Modelli di fragilità condivisa parametrica: Modello

### Modello

Un pannello facoltativo per specificare le opzioni e impostazioni del modello.

#### Distribuzione del tempo di sopravvivenza

Un'opzione per specificare la distribuzione del tempo di sopravvivenza.

##### Weibull

Specifica la distribuzione Weibull. Questa è l'impostazione predefinita.

##### Esponenziale

Specifica la distribuzione esponenziale.

##### Logaritmica normale

Specifica la distribuzione Log - normal.

##### Logaritmica logistica

Specifica la distribuzione log - logistica.

#### Impostazioni covariata

Specificare le variabili covariate.

#### Impostazioni fattore

Specificare le variabili fattore.

#### Valore iniziale di intercept

Un'opzione per specificare il valore iniziale del termine di intercettazione. Se specificato, deve essere un valore numerico singolo e non può essere 0.

#### Valore iniziale del parametro di scala

Un'opzione per controllare l'impostazione del parametro di scala.

#### Errore standard della regressione OLS corrispondente

Utilizza l'errore standard della corrispondente regressione dei minimi quadrati ordinari come valore iniziale.

#### Inverte errore standard della regressione OLS corrispondente

Utilizza il reciproco dell'errore standard.

#### Valore fornito dall'utente

Se viene specificato un singolo valore numerico, il valore viene utilizzato come valore iniziale. Se specificato, deve essere maggiore di 0.

#### Componente di fragilità

Un parametro facoltativo per specificare la **Distribuzione** del componente di fragilità.

##### Gamma

Specifica la distribuzione gamma. Questa è l'impostazione predefinita.

##### Gaussiana inversa

Specifica la distribuzione gaussiana inversa.

##### Valore iniziale della varianza

Specifica il valore iniziale della varianza della componente di fragilità. Deve essere un valore numerico singolo maggiore di 0. Il valore predefinito è 1.0 per la distribuzione gamma e 0.1 per la distribuzione gaussiana inversa.

## Modelli di fragilità condivisa parametrica: Stima

### Stima

Un pannello facoltativo per specificare le impostazioni per controllare la stima dei modelli di fragilità condivisi e il processo di selezione delle funzioni facoltativo.

## Metodo di direzione alternata o Multipinza (ADMM)

### Rapido

Applica il metodo di direzione alternata rapida dei moltiplicatori (ADMM). È l'impostazione predefinita.

### Tradizionale

Applica l'algoritmo ADMM tradizionale.

## Applica regolarizzazione L-1

Conduce il procedimento per controllare la selezione delle funzioni. Il campo **Parametro penale** specifica il parametro di penalità che controlla il processo di regolarizzazione. Deve essere un valore singolo maggiore di 0. L'impostazione predefinita è 0.001.

## Criteri di convergenza del modello

### Convergenza parametri

Specifica i criteri di convergenza per il parametro. Deve essere un singolo valore numerico appartenente a  $[0, 1)$ . L'impostazione predefinita è 0.000001. Per **Tipo**, è possibile selezionare **ASSOLUTO** per applicare la convergenza assoluta all'ottimizzazione interna oppure **RELATIVE** per applicare la convergenza relativa all'ottimizzazione interna. Il **Valore** facoltativo specifica una soglia numerica per il tipo di convergenza.

### Convergenza funzione obiettivo

Specifica i criteri di convergenza per la funzione obiettivo. Deve essere un valore numerico singolo che appartiene a  $[0, 1)$ . L'impostazione predefinita è 0, che non applica i criteri di convergenza. Per **Tipo**, è possibile selezionare **ASSOLUTO** per applicare la convergenza assoluta all'ottimizzazione interna oppure **RELATIVE** per applicare la convergenza relativa all'ottimizzazione interna. Il **Valore** facoltativo specifica una soglia numerica per il tipo di convergenza.

### Convergenza hessiana

Specifica i criteri di convergenza per la matrice hessiana. Deve essere un valore numerico singolo che appartiene a  $[0, 1)$ . L'impostazione predefinita è 0, che non applica i criteri di convergenza. Per **Tipo**, è possibile selezionare **ASSOLUTO** per applicare la convergenza assoluta all'ottimizzazione interna oppure **RELATIVE** per applicare la convergenza relativa all'ottimizzazione interna. Il **Valore** facoltativo specifica una soglia numerica per il tipo di convergenza.

## Criteri di convergenza residui

Un'opzione per controllare il processo di ottimizzazione.

### Residuo primitivo e duale

Applica entrambi i criteri di convergenza dei residui primari e doppi. Questa impostazione è predefinita.

### Solo residuo primitivo

Applica il criterio di convergenza dei residui primari.

### Solo residuo duale

Applica il criterio di convergenza dei residui duali.

## Metodo

Un parametro facoltativo per specificare il metodo di stima.

### Automatico

Sceglie automaticamente il metodo in base al dataset di esempio. Questo metodo è selezionato per impostazione predefinita. Il campo **Numero soglia di predittori** specifica la soglia del numero di predittori e deve essere un singolo numero intero maggiore di 1. Il valore predefinito è 1000.

### Newton-Raphson

Applica il metodo di Newton - Raphson.

### L-BFGS

Applica l'algoritmo BFGS a memoria limitata. Il campo **Aggiorna** specifica il numero di aggiornamenti passati conservati dall'algoritmo BFGS a memoria limitata e deve essere un singolo numero intero maggiore o uguale a 1. Il valore predefinito è 5.

## Iterazione

### Numero massimo di iterazioni

Specifica il numero massimo di iterazioni. Deve essere un singolo numero intero che appartiene a [ 1, 300]. L'impostazione predefinita è 20.

### Numero massimo di dimezzamenti

Specifica il numero massimo di dimezzamenti. Deve essere un singolo numero intero appartenente a [ 1, 200]. L'impostazione predefinita è 5.

### Numero massimo di ricerche di riga

Specifica il numero massimo di ricerche di riga. Deve essere un singolo numero intero che appartiene a [ 1, 300]. L'impostazione predefinita è 20.

### Convergenza assoluta per il processo di iterazione

Specifica la convergenza assoluta per il processo di iterazione esterno. Deve essere un valore numerico singolo che appartiene a (0, 1). L'impostazione predefinita è 0.0001.

### Convergenza relativa per il processo di iterazione

Specifica la convergenza relativa per il processo di iterazione esterno. Deve essere un valore numerico singolo che appartiene a (0, 1). L'impostazione predefinita è 0.01.

## Modelli di fragilità condivisa parametrica: Stampa

### Stampa

Un pannello facoltativo che controlla gli output della tabella.

### Dettagli di codifica del fattore

Se selezionato, visualizza e stampa i dettagli di codifica dei fattori. Il processo viene ignorato se non vi sono fattori attivi.

### Valori iniziali assegnati ai parametri di regressione

Se selezionato, visualizza i valori iniziali utilizzati nel processo di stima.

### Cronologia iterazioni modello

Se selezionato, visualizza la cronologia delle interazioni dell'analisi di sopravvivenza. Nel campo **Numero di passi**, specificare il numero di passi compreso tra 1 e 99999999. L'impostazione predefinita è 1.

## Modelli di fragilità condivisa parametrici: Prevedi

### Previsione

Un pannello facoltativo per calcolare il punteggio e salvare le statistiche previste nel dataset attivo.

### Valori temporali per il calcolo del punteggio

#### Valori temporali definiti da variabili dipendenti

Calcola il punteggio delle **Previsioni** in base alla variabile tempo specificata per il modello di sopravvivenza parametrico.

#### Intervalli regolari

Calcola il punteggio delle **Previsioni** in base ai valori temporali futuri. Il campo **Intervallo di tempo** specifica l'intervallo di tempo e deve essere un singolo valore numerico maggiore di 0. Il campo **Numero di periodi di tempo** specifica il numero di periodi di tempo e deve essere un singolo numero intero compreso tra 2 e 100.

#### Durata

Calcola il punteggio delle **previsioni** in base al periodo di tempo per definire i valori temporali futuri. Deve essere una singola variabile numerica.

## Previsioni

### Sopravvivenza

Calcola il punteggio e salva le statistiche di sopravvivenza previste nel dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredSurvival`.

### Rischio

Calcola il punteggio e salva i rischi previsti per il dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredHazard`.

### Rischio cumulativo

Calcola il punteggio e salva i rischi cumulativi previsti per il dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredCumHazard`.

### Sopravvivenza condizionale

Calcola il punteggio e salva le statistiche di sopravvivenza condizionale previste nel dataset attivo. Il nome della variabile personalizzata predefinito (o il nome root) è `PredConditionalSurvival`. Il processo viene ignorato se `PASTTIME` non è specificato. È richiesto un valore **Tempo di sopravvivenza passato** e specifica i valori del tempo passato per il calcolo del punteggio. Deve essere una singola variabile numerica.

### Sopravvivenza non condizionale

Calcola il punteggio e salva le statistiche di sopravvivenza non condizionale previste nel dataset attivo. La parola chiave viene eliminata per impostazione predefinita. Se specificato, potrebbe essere seguito da un nome variabile fornito dall'utente facoltativo (o nome root) specificato tra parentesi. Il nome predefinito è `PredUnCondSurvival`.

### Rischio non condizionale

Calcola il punteggio e salva le statistiche dei rischi non condizionali previsti nel dataset attivo. La parola chiave viene eliminata per impostazione predefinita. Se specificato, potrebbe essere seguito da un nome variabile fornito dall'utente facoltativo (o nome root) specificato tra parentesi. Il nome predefinito è `PredUncondHazard`.

### Rischio cum non condizionale

Calcola il punteggio e salva le statistiche di rischio cumulativo non condizionale previste nel dataset attivo. La parola chiave viene eliminata per impostazione predefinita. Se specificato, potrebbe essere seguito da un nome di variabile fornito dall'utente facoltativo (o da un nome root specificato tra parentesi. Il nome predefinito è `PredUncondCumHazard`.

## Modelli di fragilità parametrici condivisi: grafico

### Grafico

#### Grafici di funzione

Un'opzione per controllare i grafici della funzione.

#### Tipo

##### Sopravvivenza

Crea il grafico per le funzioni di sopravvivenza non condizionale.

##### Rischio

Crea il grafico per le funzioni di rischio non condizionale.

##### Densità

Crea un grafico per le funzioni di densità.

#### Numero di punti da visualizzare

Specifica il numero di punti funzione tra 1 e 200. L'impostazione predefinita è 100.

#### Valori di covariata del grafico

Un'opzione per specificare i valori forniti dall'utente e assegnarli ai predittori. Per impostazione predefinita, i grafici designati verranno creati alla media di ciascuna covariata in vigore. Se specificato, i grafici designati verranno creati in base all'impostazione del pattern. In presenza di qualsiasi variabile duplicata, quella specificata per prima verrà riconosciuta e il resto verrà

ignorato. Una variabile valida deve essere contenuta in un effetto del modello. Per una covariata, il valore fornito dall'utente deve essere numerico. L'omissione di una variabile in vigore indica che la media verrà utilizzata per impostazione predefinita per la covariata. Se a una variabile viene assegnato un valore non valido, il pattern richiesto non verrà tracciato.

### **Valori dei fattori del grafico**

Un'opzione per specificare i valori forniti dall'utente e assegnarli ai predittori. Per impostazione predefinita, i grafici designati verranno creati alla frequenza di categoria di ciascun fattore effettivo. Se specificato, i grafici designati verranno creati in base all'impostazione del pattern. In presenza di qualsiasi variabile duplicata, quella specificata per prima verrà riconosciuta e il resto verrà ignorato. Una variabile valida deve essere contenuta in un effetto del modello. L'omissione di una variabile in vigore indica che la frequenza della categoria verrà utilizzata per impostazione predefinita per il fattore. Se a una variabile viene assegnato un valore non valido, il pattern richiesto non verrà tracciato.

### **Separa linee per**

Un'opzione per specificare una variabile categoriale in base alla quale verranno tracciati i grafici a linee.

### **Numero massimo di linee in un grafico**

Specifica il numero massimo di linee in un grafico se è specificato **Separa linee per** . L'impostazione predefinita è 10.

## **Modelli di fragilità condivisa parametrici: Esporta**

### **Esporta**

Selezionare **Esporta informazioni sul modello in file XML** per scrivere le informazioni sul modello e sul parametro in un file PMML per il calcolo del punteggio. È necessario specificare la directory e il nome file del file PMML da salvare.

## **Modelli di fragilità parametrica condivisa: Definisci eventi**

Un'opzione per definire lo stato. Se la variabile di stato viene omessa, l'errore o l'evento diventa lo stato predefinito per tutti i casi.

1. Dal menu scegliere,

**Analisi > Sopravvivenza > Modelli di fragilità condivisa parametrica ...**

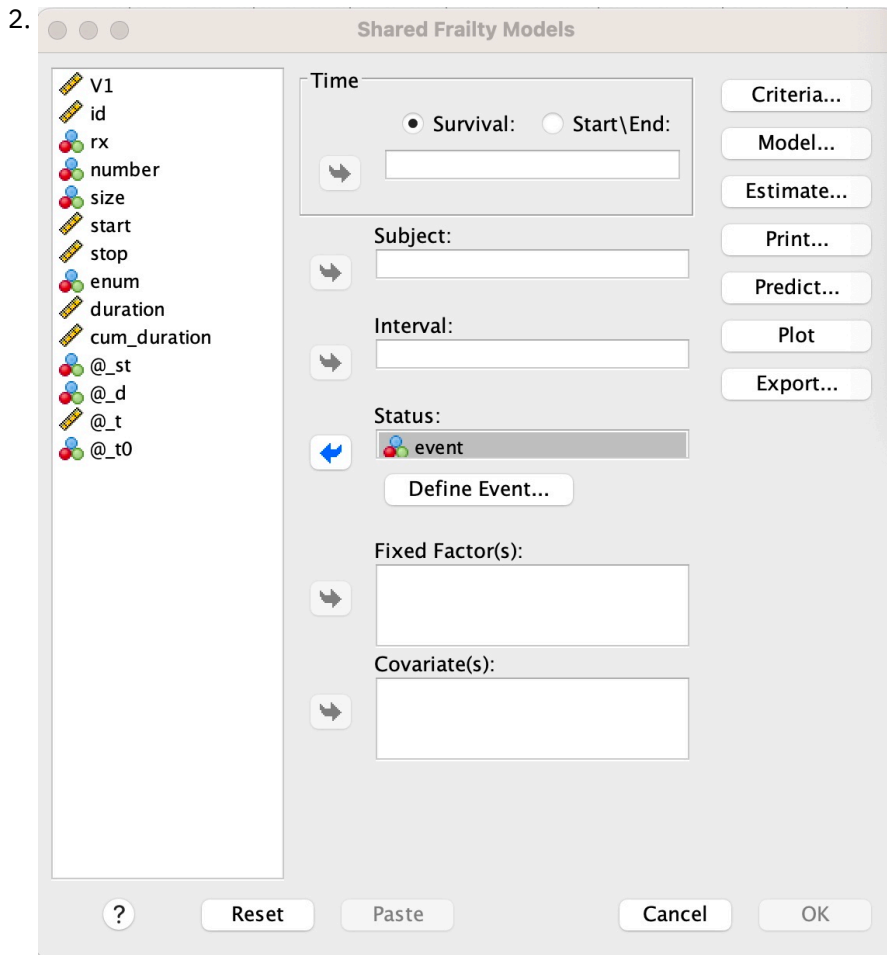


Figura 5. Modelli di fragilità condivisi - finestra di dialogo - Stato

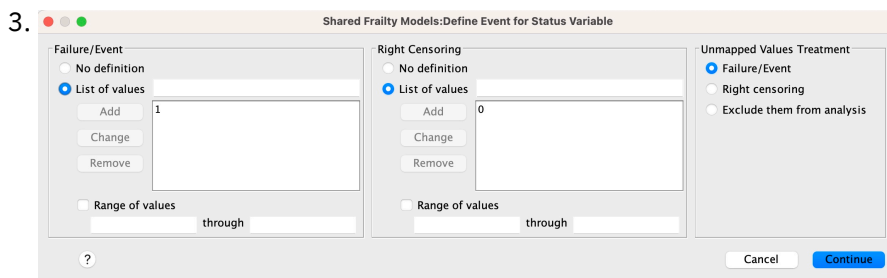


Figura 6. Modelli di fragilità condivisi - finestra di dialogo - Stato - Definisci evento

## Modelli di fragilità condivisa parametrici - Esempi

### Esempio 1

SURVREG RECURRENT y CON: x1 BY x2

/MODELLO SOGGETTO = id FRAILTY=GAMMA DISTRIBUTION=WEIBULL.

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2.

Il tempo di sopravvivenza è rappresentato da una singola variabile y.

I soggetti sono identificati dall'ID variabile.

Si presume che il tempo di sopravvivenza segua una distribuzione di Weibull.

Si presume che la varianza della fragilità segua una distribuzione gamma.

Tutti i record validi vengono utilizzati nell'analisi di sopravvivenza.

### **Esempio 2**

```
SURVREG RECURRENT y CON: x1 BY x2
```

```
/MODEL SUBJECT = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_NORMAL INTERVAL=z.
```

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2.

Il tempo di sopravvivenza è rappresentato da due variabili y1 e y2 che indicano l'ora di avvio e di fine.

I soggetti sono identificati dall'ID variabile.

Si presuppone che il tempo di sopravvivenza segua una distribuzione log - normale.

Si presume che la varianza della fragilità segua una distribuzione gaussiana inversa.

Gli intervalli di tempo sono definiti dalla variabile z. Per ogni oggetto, la procedura utilizza solo i record non in conflitto ed esclude dall'analisi tutti i record dopo il primo stato di errore.

### **Esempio 3**

```
SURVREG RECURRENT y1 y2 CON x1 BY x2(1)
```

```
/MODELLO OGGETTO = id FRAILITY=INV_GAUSSIAN DISTRIBUTION=LOG_LOGISTIC
```

```
/STATUS VARIABLE=evento FAILURE=1 RIGHT=0.
```

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2. Sopravvivenza

il tempo è rappresentato da due variabili y1 e y2 che indicano rispettivamente l'ora di inizio e di fine. Per il fattore x2, la categoria "1" è designata come baseline da modellare.

I soggetti sono identificati dall'ID variabile.

Si presuppone che il tempo di sopravvivenza segua una distribuzione log - logistica.

Si presume che la varianza della fragilità segua una distribuzione gaussiana inversa.

L'evento della variabile è specificato per definire lo stato con 1 e 0 che indicano rispettivamente l'errore e la censura a destra.

### **Esempio 4**

```
SURVREG RECURRENT y CON: x1 BY x2
```

```
/MODELLO OGGETTO = ID
```

```
/STATUS VARIABLE=evento FAILURE=1 RIGHT=0
```

```
/PREDIRE UNCONDSURVIVAL UNCONDHAZARD UNCONDCUMHAZARD
```

```
/FUNZIONPLOT SURVIVAL HAZARD DENSITY PLOTBY (x2).
```

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2. Il tempo di sopravvivenza è rappresentato da una singola variabile y.

I soggetti sono identificati dall'ID variabile.

La sopravvivenza non condizionale o basata sul popolamento, il rischio e il rischio cumulativo vengono assegnati e salvati nel dataset attivo.

Le curve di sopravvivenza e di rischio non condizionali o basate sul popolamento vengono tracciate separate dalle categorie in x2.

### Esempio 5

```
SURVREG RECURRENT y CON: x1 BY x2
```

```
/MODELLO OGGETTO = id FRAILITY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABILE=evento FAILURE=1 RIGHT=0
```

```
/STIMA HCONVERGE=1e-12(RELATIVE) PCONVERGE=0 FCONVERGE=0SELECTFEATURES=TRUE  
PENALTY=0.01.
```

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2. Il tempo di sopravvivenza è rappresentato da una singola variabile y.

I soggetti sono identificati dall'ID variabile.

Si presume che il tempo di sopravvivenza segua una distribuzione di Weibull.

Si presume che la varianza della fragilità segua una distribuzione gamma.

I criteri di convergenza si basano sulla matrice hessiana. Utilizza 1e-12 come convergenza relativa.

Il modello include un termine di penalità per controllare il processo di regolarizzazione. Il parametro di penalità è impostato su 0.01.

### Esempio 6

```
SURVREG RECURRENT y CON: x1 BY x2
```

```
/MODELLO OGGETTO = ID
```

```
/STATUS VARIABILE=infetto FAILURE=1 RIGHT=0
```

```
/STIMA MAXLINESEARCH=100 MAXITER=50 MAXSTEPHALVING=20.
```

Un modello di sopravvivenza di fragilità condivisa parametrico è adattato di y su una covariata x1 e fattore x2. Il tempo di sopravvivenza è rappresentato da una singola variabile y.

I soggetti sono identificati dall'ID variabile.

La procedura specifica che il numero massimo della ricerca di riga è 100, il numero massimo di iterazioni è 50 e il numero massimo di dimezzamenti è 20.

## Modelli di fragilità parametrici condivisi - Un caso di studio per i dati più di recente

Modelli di fragilità parametrici condivisi - Un caso di studio per i dati più di recente

Nome caso di uso - Effetto laterale di trattamento.

Attori - Investigatore e professionista della sanità pubblica.

Precondizioni - Un dataset ripulito disponibile in base al tempo di sopravvivenza, allo stato degli effetti collaterali e ai predittori da regolare.

Descrizione - Patrick, un investigatore della sanità pubblica, sta indagando su un campione di dati che include 20 partecipanti. Questi partecipanti sono reclutati in uno studio su un effetto collaterale lieve che è potenzialmente causato da un nuovo trattamento. Il progettista del trattamento sostiene che non ci sarebbero differenze tra maschi e femmine, per quanto riguarda l'effetto collaterale. Patrick vorrebbe valutare tale ipotesi. Le variabili incluse nel campione di dati sono elencate di seguito:

- patID: Numero ID per identificare un partecipante unico.

- endTime: Tempo di sopravvivenza (in giorni) dell'effetto collaterale, a seguito di un trattamento, misurato dall'inizio del trattamento alla segnalazione di un effetto collaterale o alla censura entro 60 giorni.

- sideEffect: Stato dell'effetto collaterale, stato = 0 se censurato e stato = 1 se viene riportato l'effetto collaterale lieve.

- età: età del partecipante al periodo di ricerca.

- femmina: femmina = 0 se maschio e femmina = 1 se femmina.

Potrebbero essere applicati più trattamenti, il che risulta nei record multipli dei tempi di ricorrenza misurati per un determinato partecipante. L'ora di inizio è sempre 0 per ogni record, che viene omesso nel campione di dati. Patrick è interessato a visualizzare le funzioni di sopravvivenza e di pericolo per tracciare un confronto tra un maschio e una femmina controllando la loro età e fragilità. Egli è consapevole che i trattamenti che vengono somministrati allo stesso partecipante sono più correlati. Supponendo che il tempo di sopravvivenza segua una distribuzione di Weibull, Patrick decide di creare un modello di sopravvivenza parametrico a frammentazione condivisa in SPSS Statistics per tenere conto della dipendenza dal trattamento per lo stesso partecipante.

Sintassi -

```
DATA LIST FREE
/patID(F5.0) endTime(F5.0) sideEffect(F2.0) age(F5.2) female(F2.0) .
BEGIN DATA .
1 45 0 38.00 0
2 26 1 20.00 1
3 58 0 53.00 0
4 31 1 37.00 1
4 24 0 37.00 1
4 50 0 37.00 1
5 20 1 51.00 0
5 38 1 51.00 0
6 30 0 35.00 1
7 22 1 58.00 1
8 53 1 29.00 1
8 49 1 29.00 1
9 25 0 45.00 0
9 25 0 45.00 0
10 27 0 33.00 1
11 34 1 21.00 1
11 40 0 21.00 1
11 49 0 21.00 1
12 42 1 26.00 0
13 25 0 40.00 0
14 21 1 52.00 0
14 32 1 52.00 0
15 56 0 28.00 1
15 34 0 28.00 1
16 30 0 41.00 0
16 29 0 41.00 0
17 25 1 27.00 0
18 26 1 54.00 1
18 36 1 54.00 1
19 27 0 39.00 0
20 58 1 22.00 1
20 54 0 22.00 1
20 43 1 22.00 1
END DATA.
SURVREG RECURRENT endTime WITH age BY female
/MODEL SUBJECT=patID FRAILTY=GAMMA DISTRIBUTION=WEIBULL
/ESTIMATION HCONVERGE=1e-12 PCONVERGE=0 FCONVERGE=0
/STATUS VARIABLE=sideEffect FAILURE=1 RIGHT=0
/FUNCTIONPLOT SURVIVAL HAZARD PLOTBY(female) .
```

### Sintesi:

La sintassi specificata da Patrick designa endTime come una singola variabile temporale dipendente. La procedura presuppone automaticamente che l'ora di avvio sia 0 per ogni record. Le variabili età e femmina sono modellate rispettivamente come covariata e come fattore. Si presume che i tempi di sopravvivenza della ricorrenza seguano una distribuzione di Weibull. Si presume che il termine di fragilità non osservato segua una distribuzione gamma e che la sua componente di varianza sia modellata. Per quanto riguarda

gli output, la tabella Riepilogo del modello fornisce le informazioni sulla procedura e sul modello. La tabella Riepilogo elaborazione casi fornisce un elenco completo dello stato di errore / censura e anche dei casi esclusi dall'analisi.

Nel campione di dati di Patrick, tutti i record sono validi e inclusi nell'analisi. Confrontando il logaritmo della verosimiglianza con quello del modello corrispondente senza il componente di fragilità, il modello di fragilità condivisa non riesce a raggiungere un livello significativo (valore  $p = 0.168$ ). Patrick si chiede se sia necessario includere un termine di fragilità condivisa nel modello. Il fattore di accelerazione stimato di un partecipante maschio è 1.017, ottenuto calcolando l'esponente del coefficiente di regressione stimato 0.017 di [ femmina = 0.0]. L'intervallo di confidenza del 95% associato è (.688, 1.504). Questi risultati suggeriscono che un individuo maschio ha quasi lo stesso fattore di accelerazione di un individuo femmina con la stessa età e fragilità. A livello di popolazione, Patrick traccia separatamente le curve di sopravvivenza incondizionata e di rischio per i maschi e le femmine valutati alla media campione dell'età (37.45 anni).

Patrick conferma che, per qualsiasi valore fisso del tempo di sopravvivenza, un maschio e una femmina in media dovrebbero avere la stessa probabilità di sopravvivenza. È interessante notare che, nonostante una forma unimodale mostrata nella tabella dei pericoli incondizionati, Patrick scopre che entro un periodo di 60 giorni il rischio per la popolazione sta effettivamente aumentando. Questo comportamento potrebbe implicare l'esistenza dell'effetto di fragilità. Per indagare ulteriormente l'effetto collaterale causato dai trattamenti, Patrick può continuare con un modello senza la componente di fragilità e confrontare il comportamento di maschi e femmine. Inoltre, può considerare di seguire i partecipanti per un periodo superiore a 60 giorni per raccogliere più dati.

## Kaplan - Meier Survival Analysis

---

Ci sono molte situazioni in cui si vorrebbe esaminare la distribuzione dei tempi tra due eventi, come la durata dell'occupazione (tempo tra essere assunti e lasciare l'azienda). Tuttavia, questo tipo di dati di solito include alcuni casi censurati. I casi censurati sono casi per i quali il secondo evento non è registrato (ad esempio, le persone lavorano ancora per l'azienda alla fine dello studio). La procedura Kaplan - Meier è un metodo di stima dei modelli a tempo determinato in presenza di casi censurati. Il modello Kaplan - Meier si basa sulla stima delle probabilità condizionate in ogni momento in cui si verifica un evento e si prende il limite del prodotto di tali probabilità per stimare il tasso di sopravvivenza in ogni momento.

**Esempio.** Un nuovo trattamento per l'AIDS ha qualsiasi beneficio terapeutico nell'allungare la vita? Potresti condurre uno studio usando due gruppi di malati di AIDS, uno che riceve la terapia tradizionale e l'altro che riceve il trattamento sperimentale. Costruire un modello Kaplan - Meier dai dati permetterebbe di confrontare i tassi di sopravvivenza globale tra i due gruppi per determinare se il trattamento sperimentale sia un miglioramento rispetto alla terapia tradizionale. È inoltre possibile tracciare le funzioni di sopravvivenza o di pericolo e confrontarle visivamente per informazioni più dettagliate.

**Statistiche.** Tabella di sopravvivenza, incluso il tempo, lo stato, la sopravvivenza cumulativa e l'errore standard, gli eventi cumulativi e il numero rimanente; e il tempo di sopravvivenza medio e mediano, con errore standard e intervallo di confidenza del 95%. Tracciati: sopravvivenza, pericolo, sopravvivenza log e una sola meno sopravvivenza.

Kaplan - Meier Data Considerazioni

**Dati.** La variabile del tempo dovrebbe essere continua, la variabile di stato può essere categoriale o continua e le variabili fattore e strati dovrebbero essere categoriali.

**Ipotesi.** Le probabilità per l'evento di interesse dovrebbero dipendere solo dal tempo dopo l'evento iniziale - si ipotizza che siano stabili rispetto al tempo assoluto. Ovvero, i casi che entrano nello studio in momenti diversi (ad esempio, i pazienti che iniziano il trattamento in momenti diversi) dovrebbero comportarsi analogamente. Non ci dovrebbero essere anche differenze sistematiche tra casi censurati e non censurati. Se, ad esempio, molti dei casi censurati sono pazienti con condizioni più gravi, i vostri risultati possono essere biased.

**Procedure correlate.** La procedura Kaplan - Meier utilizza un metodo di calcolo delle tabelle di vita che stima la funzione di sopravvivenza o di pericolo al momento di ogni evento. La procedura Life Tables utilizza un approccio attuariale all'analisi di sopravvivenza che si affida al partizionamento del periodo

di osservazione in intervalli di tempo più piccoli e può essere utile per affrontare campioni di grandi dimensioni. Se si dispone di variabili che si sospetta siano correlate al tempo di sopravvivenza o di variabili che si desidera controllare (covariate), utilizzare la procedura Cox Regression. Se le covariate possono avere valori diversi in momenti diversi dello stesso caso, utilizzare Cox Regression with Time-Dependent Covariates.

Ottenimento di un Kaplan - Meier Survival Analysis

1. Dai menu, scegliere:

**Analizzare > Sopravvivenza > Kaplan-Meier ...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di tempo.

3. Selezionare una variabile di stato per identificare i casi per i quali si è verificato l'evento terminale.

Questa variabile può essere numerica o *stringa breve*. Quindi fare clic su **Definisci evento**.

Facoltativamente, è possibile selezionare una variabile fattore per esaminare le differenze di gruppo. È inoltre possibile selezionare una variabile strati, che produrrà analisi separate per ogni livello (strato) della variabile.

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

Inserire il valore o i valori che indicano che l'evento terminale si è verificato. È possibile inserire un valore unico, una serie di valori o un elenco di valori. L'opzione Intervallo di Valori è disponibile solo se la tua variabile di stato è numerica.

## Kaplan - Meier Confronta Factor Livelli

È possibile richiedere statistiche per testare l'uguaglianza delle distribuzioni di sopravvivenza per i diversi livelli del fattore. Le statistiche disponibili sono log rank, Breslow e Tarone-Ware. Seleziona una delle alternative per specificare i confronti da fare: pooled over strati, per ogni strato, pairwise su strati, o pairwise per ogni strato.

**Trend lineare per i livelli dei fattori.** Consente di testare un andamento lineare su livelli del fattore. Questa opzione è disponibile solo per i confronti complessivi (piuttosto che per pairwise) dei livelli dei fattori.

- *rank di registrazione.* Un test per confrontare l'uguaglianza delle distribuzioni di sopravvivenza. In questo test tutti i punti temporali vengono pesati ugualmente.
- *Breslow.* Un test per confrontare l'uguaglianza delle distribuzioni di sopravvivenza. I punti temporali vengono pesati in base al numero di casi a rischio in ogni punto temporale.
- *Tarone - Ware.* Un test per confrontare l'uguaglianza delle distribuzioni di sopravvivenza. I punti temporali sono pesati per la radice quadrata del numero dei casi a rischio in ogni punto temporale.
- *Pooled over strati.* Confronta tutti i livelli dei fattori in un singolo test per verificare l'uguaglianza delle curve di sopravvivenza.
- *Pairwise over strati.* Confronta ogni coppia distinta di livelli fattore. I test delle tendenze pairwise non sono disponibili.
- *Per ogni strato.* Esegue un test separato di uguaglianza di tutti i livelli dei fattori per ogni strato. In mancanza di una variabile di stratificazione i test non verranno eseguiti.
- *Pairwise per ogni strato.* Confronta ogni coppia distinta di livelli fattore per ogni strato. Non disponibile per i test della tendenza pairwise. In mancanza di una variabile di stratificazione i test non verranno eseguiti.

## Kaplan - Meier Save New Variables

È possibile salvare le informazioni dalla tabella Kaplan - Meier come nuove variabili, che possono poi essere utilizzate nelle successive analisi per verificare ipotesi o verificare ipotesi. È possibile salvare la sopravvivenza, errore standard di sopravvivenza, pericolo e eventi cumulativi come nuove variabili.

- *Sopravvivenza*. Stima della probabilità di sopravvivenza cumulativa. Il nome predefinito delle variabili è costituito da un prefisso *sur\_* e da un numero sequenziale. Ad esempio, se *sur\_1* già esiste, Kaplan-Meier assegna il nome di variabile *sur\_2*.
- *Errore standard di sopravvivenza*. Errore standard della stima di sopravvivenza cumulativa. Il nome delle variabili predefinito è composto da un prefisso *cum\_* e da un numero progressivo. Ad esempio, se *se\_1* già esiste, Kaplan-Meier assegna il nome di variabile *se\_2*.
- *Rischio*. Stima della funzione di pericolo cumulativo. Le variabili hanno il nome composto da un prefisso *haz\_* e da un numero progressivo. Ad esempio, se *haz\_1* già esiste, Kaplan-Meier assegna il nome di variabile *haz\_2*.
- *Eventi cumulativi*. Frequenza cumulativa degli eventi quando i casi sono ordinati in base ai loro tempi di sopravvivenza e ai loro codici di stato. Le variabili hanno il nome composto da un prefisso *cum\_* e da un numero progressivo. Ad esempio, se *cum\_1* già esiste, Kaplan-Meier assegna il nome di variabile *cum\_2*.

## Kaplan - Opzioni Meier

È possibile richiedere vari tipi di output dall'analisi Kaplan - Meier.

**Statistiche.** È possibile selezionare le statistiche visualizzate per le funzioni di sopravvivenza calcolate, comprese le tabelle di sopravvivenza, la sopravvivenza media e mediana e le quartili. Se sono state incluse variabili fattore, per ciascun gruppo verranno generate statistiche distinte.

**Grafici.** I tracciati consentono di esaminare visivamente le funzioni di sopravvivenza, di - minus - sopravvivenza, di pericolo e di sopravvivenza log - sopravvivenza. Se sono state incluse variabili fattore, le funzioni verranno tracciate per ciascun gruppo.

- *Sopravvivenza*. Visualizza la funzione di sopravvivenza cumulata in scala lineare.
- *Uno meno sopravvivenza*. Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.
- *Rischio*. Visualizza la funzione di rischio cumulato in scala lineare.
- *sopravvivenza del log*. Visualizza la funzione di sopravvivenza cumulata in scala logaritmica.

## Funzioni Aggiuntive di comando KM

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere tabelle di frequenza che considerano i casi persi a follow-up come categoria separata dai casi censurati.
- Specificare una spaziatura non uguale per il test per la tendenza lineare.
- Ottenere percentili diversi da quartili per la variabile tempo di sopravvivenza.

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

## Cox Analisi di regressione

La regressione di Cox consente di creare 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 per i predittori vengono stimati in base ai soggetti osservati; il modello può quindi essere applicato a nuovi casi che hanno misurazioni per le variabili predittore. Si noti che le informazioni provenienti da soggetti censurati, ovvero quelli che non sono interessati dall'evento nel tempo di osservazione, contribuiscono notevolmente alla stima del modello.

**Esempio.** Gli uomini e le donne hanno diversi rischi di sviluppare un tumore al polmone basato sul fumo di sigaretta? Costruendo un modello di regressione Cox, con l'uso di sigarette (sigarette fumate al giorno) e il sesso inseriti come covariate, è possibile testare le ipotesi relative agli effetti del sesso e dell'uso di sigarette sul tempo di insorgenza del cancro ai polmoni.

**Statistiche.** Per ogni modello:  $-2LL$ , la statistica di probabilità - rapporto e il chi - quadrato complessivo. Per le variabili nel modello: stime dei parametri, errori standard e statistiche Wald. Per le variabili non nel modello: statistiche del punteggio e del chi - quadrato residuo.

Cox Considerazioni sui dati di regressione

**Dati.** La tua variabile di tempo dovrebbe essere quantitativa, ma la tua variabile di stato può essere categoriale o continua. Le variabili indipendenti (covariate) possono essere continue o categoriali; se categoriali, dovrebbero essere dummy - o codificate con indicatore (c'è un'opzione nella procedura per codificare automaticamente le variabili categoriali). Le variabili strati dovrebbero essere categoriali, codificate come numeri interi o brevi.

**Ipotesi.** Le osservazioni dovrebbero essere indipendenti, e il rapporto di pericolo dovrebbe essere costante nel tempo; cioè la proporzionalità dei pericoli da un caso all'altro non deve variare nel tempo. Quest' ultima assunzione è nota come l' **assunzione di rischi proporzionali**.

**Procedure correlate.** Se l'ipotesi di rischio proporzionale non è valida (vedi sopra), potrebbe essere necessario utilizzare la procedura Cox con covariate dipendenti dal tempo. Se non hai covariate o se hai solo una covariata categoriale, puoi utilizzare la procedura Life Tables o Kaplan - Meier per esaminare le funzioni di sopravvivenza o di pericolo per il tuo campione (s). Se nel tuo campione non hai dati censurati (cioè ad ogni caso sperimentato l'evento terminale), è possibile utilizzare la procedura di Regressione Lineare per modellare il rapporto tra predittori e temporali.

Ottenere un'analisi di regressione Cox

1. Dai menu, scegliere:

**Analizzare > Sopravvivenza > Cox Regressione...**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di tempo. I casi i cui valori temporali sono negativi non vengono analizzati.

3. Selezionare una variabile di stato, quindi fare clic su **Definisci evento**.

4. 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>**.

Facoltativamente, è possibile calcolare modelli separati per diversi gruppi definendo una variabile strati.

## Cox Regressione Definire le variabili categoriali

È possibile specificare il modo in cui la procedura Cox Regression gestirà le variabili categoriche.

**Covariati.** Elenca tutte le covariate specificate nella finestra di dialogo principale, sia da sole che come parte di un'interazione, in qualsiasi strato. Se alcune di queste sono variabili di stringa o sono categoriali, è possibile utilizzarle solo come covariate categoriali.

**Covariati categoriali.** Elenca le variabili identificate come categoriali. Ogni variabile include una notazione tra parentesi che indica la codifica di contrasto da utilizzare. Le variabili di stringa (denotate dal simbolo < seguendo i loro nomi) sono già presenti nell'elenco Categorie Categorie. Selezionare qualsiasi altra covariata categoriale dalla lista dei Covariati e spostarle nella lista Categorie Categorie.

**Contrasto di modifica.** 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 fila di zeri.
- **Semplice.** Ogni categoria della variabile predittivo tranne la categoria di riferimento è paragonata alla categoria di riferimento.

- **Differenza.** Ogni categoria della variabile predittivo tranne la prima categoria viene paragonata all'effetto medio delle categorie precedenti. Sono noti anche come contrasti inversi di Helmert.
- **Helmert.** Ogni categoria della variabile predittivo tranne l'ultima categoria viene paragonata all'effetto medio delle categorie successive.
- **Ripetuto.** Ogni categoria della variabile predittivo tranne la prima categoria viene paragonata alla categoria che lo precede.
- **Polinomiale.** Contrasti polinomiali ortogonali. Si presume che le categorie siano equamente distanziate. I contrasti polinomiali sono disponibili solo per variabili numeriche.
- **Deviazione.** Ogni categoria della variabile predittivo tranne la categoria di riferimento è paragonata all'effetto complessivo.

Se si seleziona **Deviazione**, **Semplice Indicatore**, selezionare **Primo** o **Ultimo** come categoria di riferimento. Da notare che il metodo non viene effettivamente modificato fino a quando non si fa clic su **Modifica**.

Le covariate di stringa devono essere covariate categoriali. Per rimuovere una variabile stringa dall'elenco Categorie Categorie, è necessario rimuovere tutti i termini contenenti la variabile dall'elenco Covariati nella finestra di dialogo principale.

## Cox Grafici di regressione

I grafici agevolano la valutazione del modello stimato e l'interpretazione dei risultati. È possibile rappresentare graficamente funzioni di sopravvivenza, pericolo, logaritmo meno logaritmo e sopravvivenza meno uno.

### Tipo di grafico

#### Sopravvivenza

Visualizza la funzione di sopravvivenza cumulata in scala lineare.

#### Rischio

Visualizza la funzione di rischio cumulato in scala lineare.

#### Log meno log

La stima di sopravvivenza cumulativa dopo che la trasformazione  $\ln(-\ln)$  viene applicata alla stima.

#### Sopravvivenza meno uno

Visualizza il complemento a uno della funzione di sopravvivenza su una scala lineare.

### Valori di covariata tracciati in

È possibile tracciare una riga separata per ogni valore di una covariata categoriale spostando quella covariata nella casella di testo **Linee separate per**. Questa opzione è disponibile solo per le covariate categoriali, che sono denotate da **(Cat)** dopo i loro nomi nella lista **Covariate Values Plotted at**.

### Modifica valore

Poiché queste funzioni dipendono dai valori delle covariate, è necessario utilizzare valori costanti per le covariate per tracciare le funzioni rispetto al tempo. Il default è quello di utilizzare la media di ciascuna covariata come valore costante, ma è possibile inserire i propri valori per la trama utilizzando il gruppo di controllo **Modifica valore**.

## Cox Regressione Salva nuove variabili

Puoi salvare vari risultati della tua analisi come nuove variabili. Queste variabili possono quindi essere utilizzate nelle successive analisi per testare le ipotesi o per verificare le assunzioni.

### Salva variabili modello

Consente di salvare la funzione di sopravvivenza e il suo errore standard, le stime log-minus-log, la funzione di rischio, i residui parziali,  $Df\beta(s)$  per la regressione e il predittore lineare  $X*\beta$  come nuove variabili.

### Funzione di sopravvivenza

Il valore della funzione di sopravvivenza cumulativa per un determinato periodo di tempo. Equivale alla probabilità di sopravvivenza a tale periodo di tempo.

**Errore standard della funzione di sopravvivenza**

Errore standard della stima cumulativa di sopravvivenza.

**Funzione di sopravvivenza log meno log**

La stima di sopravvivenza cumulativa 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**

È possibile tracciare i residui parziali rispetto al tempo di sopravvivenza per verificare l'ipotesi di rischi proporzionali. Viene salvata una variabile per ogni covariata. I residui parziali sono disponibili solo per i modelli che contengono almeno una covariata.

**DfBeta(s)**

Variatione stimata in un coefficiente se viene rimosso un caso. Viene salvata una variabile per ogni covariata. Le DfBeta (differenze nei beta) 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 del parametro per ogni caso.

**Nota:** Solo DfBeta(s) viene salvato se si esegue Cox con una covariata dipendente dal tempo.

**Esporta informazioni modello in file XML**

Le stime dei parametri vengono esportate nel file specificato in formato XML. È possibile utilizzare questo file modello per applicare le informazioni del modello ad altri file di dati per il calcolo del punteggio.

## Cox Opzioni di regressione

È possibile controllare vari aspetti della propria analisi e produzione.

**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 hai selezionato un metodo stepwise, è possibile specificare la probabilità per l'ingresso o la rimozione dal modello. Viene inserita una variabile se il livello di significatività del suo  $F$ -to - enter è inferiore al valore di Voce e viene rimossa una variabile se il livello di significatività è maggiore del valore di rimozione. Il valore Voce deve essere inferiore al valore rimovibile.

**Numero massimo 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. Questo display non è disponibile se si hanno specificate covariate a tempo determinato.

## Cox Regressione Definire l'evento per la variabile di stato

Inserire il valore o i valori che indicano che l'evento terminale si è verificato. È possibile inserire un valore unico, una serie di valori o un elenco di valori. L'opzione Intervallo di Valori è disponibile solo se la tua variabile di stato è numerica.

## Funzioni Aggiuntive Comando COXREG

Il linguaggio della sintassi dei comandi consente inoltre di:

- Ottenere tabelle di frequenza che considerano i casi persi a follow-up come categoria separata dai casi censurati.

- Selezionare una categoria di riferimento, diversa da prima o ultima, per la deviazione, i metodi di contrasto semplici e indicatori.
- Specificare una spaziatura non uguale di categorie per il metodo di contrasto polinomiale.
- Specificare ulteriori criteri di iterazione.
- Controllare il trattamento dei valori mancanti.
- Specificare i nomi per le variabili salvate.
- Scrivere l'output in un file di dati IBM SPSS Statistica esterno.
- Tieni i dati per ogni gruppo di file suddiviso in un file di scratch esterno durante l'elaborazione. Questo può aiutare a conservare le risorse di memoria quando si eseguono analisi con i grandi datasets. Questo non è disponibile con covariate dipendenti dal tempo.

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

## Tempo di calcolo - Covariati dipendenti

In alcune situazioni, si vorrebbe calcolare un modello di regressione Cox ma l'ipotesi di rischio proporzionale non è valida. Ovvero i rapporti di pericolo cambiano nel tempo; i valori di uno (o più) delle vostre covariate sono diversi in momenti diversi. In questi casi, è necessario utilizzare un modello di regressione esteso Cox, in cui è possibile specificare le **covariate dipendenti dal tempo**.

Per analizzare un modello di questo tipo, è necessario prima definire la covariata dipendente dal tempo. Per facilitare questo, è disponibile una 'variabile di sistema' che rappresenta il tempo. Questa variabile si chiama  $T_*$ . È possibile utilizzare questa variabile per definire covariate tempo - dipendenti in due modi generali:

- Per verificare l'ipotesi di rischio proporzionale rispetto a una particolare covariata o per stimare un modello di regressione Cox esteso che ammette rischi non proporzionali, è necessario definire la covariata dipendente dal tempo come funzione della variabile temporale  $T_*$  e della covariata in questione. Un esempio comune sarebbe il semplice prodotto della variabile del tempo e della covariata, ma possono essere specificate anche funzioni più complesse. La verifica della significatività del coefficiente della covariata dipendente dal tempo indica se l'ipotesi dei rischi proporzionali è ragionevole.
- Alcune variabili possono avere valori differenti in periodi di tempo differenti ma non sono sistematicamente correlate al tempo. In tali casi, è necessario definire una **covariata dipendente dal tempo segmentata**, che può essere eseguita utilizzando le **espressioni logiche**. Le espressioni logiche assumono il valore 1 se vero e 0 se false. Utilizzando una serie di espressioni logiche, è possibile creare la covariata dipendente dal tempo da una serie di misurazioni. Ad esempio, se si ha la pressione arteriosa misurata una volta alla settimana per le quattro settimane dello studio (identificata come da  $BP1$  a  $BP4$ ), è possibile definire la covariata dipendente dal tempo come  $(T_* < 1) * BP1 + (T_* \geq 1 \& T_* < 2) * BP2 + (T_* \geq 2 \& T_* < 3) * BP3 + (T_* \geq 3 \& T_* < 4) * BP4$ . Si noti che esattamente uno dei termini tra parentesi è uguale a 1 per ogni caso specifico e il resto sarà uguale a 0. Per riassumere, questa funzione indica che se il tempo è inferiore a una settimana, utilizzare  $BP1$ ; se è superiore a una settimana ma inferiore a due settimane, utilizzare  $BP2$  e così via.

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

Nella finestra di dialogo Calcola covariata dipendente dal tempo, è possibile utilizzare i controlli di creazione della funzione per creare l'espressione per una covariata dipendente dal tempo oppure è possibile immetterla direttamente nell'area di testo Nome. Notare che le costanti di stringa devono essere racchiuse tra virgolette o apostrofi e le costanti numeriche devono essere digitate in formato americano, con il punto come delimitatore decimale. Qualsiasi covariata risultante dipendente dal tempo deve essere inclusa come covariata nel modello di regressione Cox.

## Computing a tempo - Patto dipendente

1. Dal menu, scegliere

### Analizzare la > sopravvivenza > Cox con il tempo di cov...

2. Inserire un'espressione per la covariata dipendente dal tempo.
3. Fare clic su **Modello** per procedere con la regressione Cox.

Nota - Assicurarsi di includere le nuove variabili aggiunte come covariate nel modello di regressione Cox.

## Schemi Di Codifica Variabile Categoriale

In molte procedure è possibile richiedere la sostituzione automatica di una variabile indipendente categoriale con una serie di variabili di contrasto, che verranno poi inserite o rimosse da un'equazione come blocco. È possibile specificare come deve essere codificata la serie di variabili di contrasto, di solito nel comando CONTRAST . Questa appendice spiega e illustra come diversi tipi di contrasto richiesti su CONTRAST funzionano effettivamente.

## Deviazione

**Deviazione dal grand mean.** In termini di matrice, questi contrasti hanno la forma:

```
mean ( 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 il numero della categoria omessa tra parentesi dopo la parola chiave DEVIATION . Ad esempio, il seguente comando ottiene le deviazioni per la prima e la terza categoria e omette la seconda:

```
/CONTRAST (FACTOR)=DEVIATION (2)
```

Supponga che *fattore* abbia tre categorie. La matrice di contrasto risultante sarà

```
( 1/3 1/3 1/3)
( 2/3 -1/3 -1/3)
(-1/3 -1/3 2/3)
```

## Semplice

**contrast semplici.** Confronta ogni livello di un fattore fino all'ultimo. Il modulo di matrice generale è

```
mean (1/k 1/k ... 1/k 1/k)
df(1) ( 1 0 ... 0 -1)
df(2) ( 0 1 ... 0 -1)
.
.
df(k-1) ( 0 0 ... 1 -1)
```

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:

```
(1/4 1/4 1/4 1/4)
( 1 0 0 -1)
( 0 1 0 -1)
( 0 0 1 -1)
```

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 è

necessariamente il valore associato a tale categoria. Ad esempio, il seguente comando CONTRAST ottiene una matrice di contrasto che omette la seconda categoria:

```
/CONTRAST(FACTOR) = SIMPLE(2)
```

Supponga che *fattore* abbia quattro categorie. La matrice di contrasto risultante sarà

```
(1/4  1/4  1/4  1/4)
(  1  -1   0   0)
(  0  -1   1   0)
(  0  -1   0   1)
```

## Helmert

**contrasti Helmert.** Confronta le categorie di una variabile indipendente con la media delle categorie successive. Il modulo di matrice generale è

```
mean (1/k  1/k  ...  1/k  1/k  1/k)
df(1) (  1 -1/(k-1) ... -1/(k-1) -1/(k-1) -1/(k-1))
df(2) (  0  1 ... -1/(k-2) -1/(k-2) -1/(k-2))
.
.
df(k-2) (  0  0 ...  1 -1/2 -1/2)
df(k-1) (  0  0 ...  0  1 -1)
```

dove  $k$  è il numero di categorie della variabile indipendente. Ad esempio, una variabile indipendente con quattro categorie ha una matrice di contrasto Helmert del seguente modulo:

```
(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 reverse Helmert contrasti.** Confronta le categorie di una variabile indipendente con la media delle categorie precedenti della variabile. Il modulo di matrice generale è

```
mean (  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 di 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 contiene l'effetto lineare in tutte le categorie; il secondo grado di libertà, l'effetto quadratico; il terzo grado di libertà, il cubico; e così via, per gli effetti di ordine superiore.

È possibile specificare la spaziatura tra i livelli del trattamento misurati dalla variabile categoriale data. La spaziatura uguale, che è il default se si ometta la metrica, può essere specificata come numeri interi consecutivi da 1 a  $k$ , dove  $k$  è il numero di categorie. Se la variabile *farmaco* ha tre categorie, il sottocomando

```
/CONTRAST(DRUG)=POLYNOMIAL
```

è lo stesso che

```
/CONTRAST(DRUG)=POLYNOMIAL(1,2,3)
```

La parità di spazio non è sempre necessaria, però. Ad esempio, supponiamo che *farmaco* rappresenti dosaggi diversi di un farmaco somministrato a tre gruppi. Se il dosaggio somministrato al secondo gruppo è due volte quello che dato al primo gruppo e il dosaggio somministrato al terzo gruppo è tre volte quello dato al primo gruppo, le categorie di trattamento sono ugualmente distanziate, e una metrica appropriata per questa situazione è costituita da numeri interi consecutivi:

```
/CONTRAST (DRUG)=POLYNOMIAL (1,2,3)
```

Se però il dosaggio somministrato al secondo gruppo è quattro volte quello dato al primo gruppo, e il dosaggio somministrato al terzo gruppo è sette volte quello dato al primo gruppo, una metrica appropriata è

```
/CONTRAST (DRUG)=POLYNOMIAL (1,4,7)
```

In entrambi i casi, il risultato della specifica di contrasto è che il primo grado di libertà per *farmaco* contiene l'effetto lineare dei livelli di dosaggio e il secondo grado di libertà contiene l'effetto quadratico.

I contrasti polinomiali sono particolarmente utili nei test delle tendenze e per indagare sulla natura delle superfici di risposta. È inoltre possibile utilizzare contrasti polinomiali per eseguire l'allestimento di curve non lineari, come la regressione curvilinea.

## Ripetuto

**Confronta i livelli adiacenti di una variabile indipendente.** Il modulo di matrice generale è

```
mean (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:

```
(1/4 1/4 1/4 1/4)
( 1 -1 0 0)
( 0 1 -1 0)
( 0 0 1 -1)
```

Questi contrasti sono utili nell'analisi del profilo e ovunque siano necessari punteggi di differenza.

## Speciale

**Un contrasto definito dall'utente.** Consente l'ingresso di contrasti speciali sotto forma di matrici quadrate con altrettante righe e colonne in quanto ci sono categorie della determinata variabile indipendente. Per MANOVA e LOGLINEAR, la prima riga inserita è sempre la media, o costante, l'effetto e rappresenta la serie di pesi che indicano come media altre variabili indipendenti, se presenti, sulla variabile data. Generalmente questo contrasto è un vettore di quelli.

Le righe rimanenti della matrice contengono i contrasti speciali che indicano i confronti tra categorie della variabile. Di solito, i contrasti ortogonali sono i più utili. I contrasti ortogonali sono statisticamente indipendenti e non sono ridondanti. I contrasti sono ortogonali se:

- Per ogni riga, i coefficienti di contrasto si somma a 0.
- I prodotti dei corrispondenti coefficienti per tutte le coppie di righe disgiunte si somma anche a 0.

Ad esempio, supponga che il trattamento abbia quattro livelli e che si voglia confrontare i vari livelli di trattamento tra loro. Un adeguato contrasto speciale è

```
(1 1 1 1) weights for mean calculation
(3 -1 -1 -1) compare 1st with 2nd through 4th
(0 2 -1 -1) compare 2nd with 3rd and 4th
(0 0 1 -1) compare 3rd with 4th
```

che si specifica per mezzo del seguente comando CONTRAST per MANOVA, LOGISTIC REGRESSIONE 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 )
```

Ogni riga tranne la riga significa somma a 0. Prodotti di ogni coppia di righe disgiunte somma anche a 0:

```
Rows 2 and 3: (3)(0) + (-1)(2) + (-1)(-1) + (-1)(-1) = 0
Rows 2 and 4: (3)(0) + (-1)(0) + (-1)(1) + (-1)(-1) = 0
Rows 3 and 4: (0)(0) + (2)(0) + (-1)(1) + (-1)(-1) = 0
```

I contrasti speciali non devono essere ortogonali. Tuttavia, non devono essere combinazioni lineari tra loro. Se lo sono, la procedura riporta la dipendenza lineare e cessa l'elaborazione. Helmert, differenza e contrasti polinomiali sono tutti contrasti ortogonali.

## Indicatore

**Codifica variabile indicatore.** Noto anche come codifica fittizio, questo non è disponibile in LOGLINEAR o MANOVA. Il numero di nuove variabili codificate è  $k-1$ . I casi nella categoria di riferimento sono codificati 0 per tutte le variabili  $k-1$ . Un caso nella categoria  $i^{\text{th}}$  è codificato 0 per tutte le variabili indicatore tranne  $i^{\text{th}}$ , codificati 1.

## Strutture covarianza

Questa sezione fornisce ulteriori informazioni sulle strutture di covarianza.

**Ante - Dipendenza: Primo - Ordine.** Questa struttura di covarianza ha varianze eterogenee e correlazioni eterogenee tra elementi adiacenti. La correlazione tra due elementi non adiacenti è il prodotto delle correlazioni tra gli elementi che risiedono tra gli elementi di interesse.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_1 & \sigma_3 \sigma_1 \rho_1 \rho_2 & \sigma_4 \sigma_1 \rho_1 \rho_2 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_3 \sigma_2 \rho_2 & \sigma_4 \sigma_2 \rho_2 \rho_3 \\ \sigma_3 \sigma_1 \rho_1 \rho_2 & \sigma_3 \sigma_2 \rho_2 & \sigma_3^2 & \sigma_4 \sigma_3 \rho_3 \\ \sigma_4 \sigma_1 \rho_1 \rho_2 \rho_3 & \sigma_4 \sigma_2 \rho_2 \rho_3 & \sigma_4 \sigma_3 \rho_3 & \sigma_4^2 \end{pmatrix}$$

**AR (1).** Si tratta di una struttura autoregressiva di primo ordine con varianze omogenee. La correlazione tra due elementi è uguale a rho per gli elementi adiacenti, rho<sup>2</sup> per gli elementi separati da un terzo e così via. è vincolato in modo che  $-1 < \rho < 1$ .

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 & \sigma^2 \rho^3 \\ \sigma^2 \rho & \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 \\ \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 & \sigma^2 \rho \\ \sigma^2 \rho^3 & \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 \end{pmatrix}$$

**AR (1): Eeterogenesi.** Si tratta di una struttura autoregressiva di primo ordine con varianze eterogenee. La correlazione tra due elementi è uguale a r per gli elementi adiacenti, a r<sup>2</sup> per due elementi separati da un terzo e così via. è vincolato a mentire tra il -1 e il 1.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho & \sigma_3 \sigma_1 \rho^2 & \sigma_4 \sigma_1 \rho^3 \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_3 \sigma_2 \rho & \sigma_4 \sigma_2 \rho^2 \end{pmatrix}$$

$$\begin{pmatrix} \sigma_3 \sigma_1 \rho^2 & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_4 \sigma_3 \rho \\ \sigma_4 \sigma_1 \rho^3 & \sigma_4 \sigma_2 \rho^2 & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{pmatrix}$$

**ARMA (1, 1).** Si tratta di una struttura media mobile autoregressiva di primo ordine, che ha varianze omogenee. La correlazione tra due elementi è uguale a  $\rho$  per elementi adiacenti,  $\rho^2$  per elementi separati da un terzo, e così via. e sono i parametri di media autoregressiva e di spostamento, rispettivamente, e i loro valori sono vincolati a mentire tra il -1 e il 1, inclusivo.

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 & \sigma^2 \rho^3 \\ \sigma^2 \rho & \sigma^2 & \sigma^2 \rho & \sigma^2 \rho^2 \\ \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 & \sigma^2 \rho \\ \sigma^2 \rho^3 & \sigma^2 \rho^2 & \sigma^2 \rho & \sigma^2 \end{pmatrix}$$

**Symmetry Compound.** Questa struttura ha varianza costante e covarianza costante.

$$\begin{pmatrix} \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1^2 \end{pmatrix}$$

**Symmetry Compound: Metric Metric.** Questa struttura di covarianza ha varianze omogenee e correlazioni omogenee tra elementi.

$$\begin{pmatrix} \sigma^2 & \sigma_2 \rho & \sigma_2 \rho & \sigma_2 \rho \\ \sigma_2 \rho & \sigma^2 & \sigma_2 \rho & \sigma_2 \rho \\ \sigma_2 \rho & \sigma_2 \rho & \sigma^2 & \sigma_2 \rho \\ \sigma_2 \rho & \sigma_2 \rho & \sigma_2 \rho & \sigma^2 \end{pmatrix}$$

**Symmetry Composto: Eeterogenico.** Questa struttura di covarianza ha varianze eterogenee e correlazione costante tra elementi.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho & \sigma_3 \sigma_1 \rho & \sigma_4 \sigma_1 \rho \\ \sigma_2 \sigma_1 \rho & \sigma_2^2 & \sigma_3 \sigma_2 \rho & \sigma_4 \sigma_2 \rho \\ \sigma_3 \sigma_1 \rho & \sigma_3 \sigma_2 \rho & \sigma_3^2 & \sigma_4 \sigma_3 \rho \\ \sigma_4 \sigma_1 \rho & \sigma_4 \sigma_2 \rho & \sigma_4 \sigma_3 \rho & \sigma_4^2 \end{pmatrix}$$

**Diagonale.** Questa struttura di covarianza ha varianze eterogenee e zero correlazione tra elementi.

$$\begin{pmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_3^2 & 0 \\ 0 & 0 & 0 & \sigma_4^2 \end{pmatrix}$$

**Prodotto diretto AR1 (UN\_AR1).** Specifica il prodotto Kronecker di una matrice non strutturata e l'altra matrice di covarianza auto - regressione di primo ordine. La prima matrice non strutturata modella l'osservazione multivariata e la seconda struttura di covarianza auto - regressione di primo ordine modella la covarianza dei dati nel tempo o in un altro fattore.

**Prodotto diretto non strutturato (UN\_UN).** Specifica il prodotto Kronecker di due matrici non strutturate, con la prima modellazione dell'osservazione multivariata, e la seconda la modellazione della covarianza dei dati nel tempo o un altro fattore.

**La simmetria composta del prodotto diretto (UN\_CS).** Specifica il prodotto Kronecker di una matrice non strutturata e l'altra matrice di covarianza composta - simmetria con varianza costante e covarianza. La prima matrice non strutturata modella l'osservazione multivariata e la seconda struttura di covarianza di simmetria composta modella la covarianza dei dati nel tempo o in un altro fattore.

**Fattore analitico: Primo - Ordine.** Questa struttura di covarianza ha varianze eterogenee che sono composte da un termine che è eterogeneo tra gli elementi e un termine che è omogeneo tra gli elementi. La covarianza tra ogni due elementi è la radice quadrata del prodotto della loro varianza eterogenea.

$$\begin{pmatrix} \lambda_1^2 + d & \lambda_2 \lambda_1 & \lambda_3 \lambda_1 & \lambda_4 \lambda_1 \\ \lambda_2 \lambda_1 & \lambda_2^2 + d & \lambda_3 \lambda_2 & \lambda_4 \lambda_2 \\ \lambda_3 \lambda_1 & \lambda_3 \lambda_2 & \lambda_3^2 + d & \lambda_4 \lambda_3 \\ \lambda_4 \lambda_1 & \lambda_4 \lambda_2 & \lambda_4 \lambda_3 & \lambda_4^2 + d \end{pmatrix}$$

**Fattore analitico: Primo - Ordine, Eeterogenico.** Questa struttura di covarianza ha varianze eterogenee che sono composte da due termini che sono eterogenei tra gli elementi. La covarianza tra ogni due elementi è la radice quadrata del prodotto del primo dei loro termini di varianza eterogenei.

$$\begin{pmatrix} \lambda_1^2 + d_1 & \lambda_2 \lambda_1 & \lambda_3 \lambda_1 & \lambda_4 \lambda_1 \\ \lambda_2 \lambda_1 & \lambda_2^2 + d_2 & \lambda_3 \lambda_2 & \lambda_4 \lambda_2 \\ \lambda_3 \lambda_1 & \lambda_3 \lambda_2 & \lambda_3^2 + d_3 & \lambda_4 \lambda_3 \\ \lambda_4 \lambda_1 & \lambda_4 \lambda_2 & \lambda_4 \lambda_3 & \lambda_4^2 + d_4 \end{pmatrix}$$

**Huynh-Feldt.** Si tratta di una matrice "circolare" in cui la covarianza tra due elementi è uguale alla media delle loro varianze meno una costante. Né le varianze né le covarianze sono costanti.

$$\begin{pmatrix} \sigma_1^2 & [\sigma_1^2 + \sigma_2^2]/2 - \lambda & [\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_1^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_2^2]/2 - \lambda & \sigma_2^2 & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_3^2]/2 - \lambda & [\sigma_2^2 + \sigma_3^2]/2 - \lambda & \sigma_3^2 & [\sigma_3^2 + \sigma_4^2]/2 - \lambda \\ ([\sigma_1^2 + \sigma_4^2]/2 - \lambda & [\sigma_2^2 + \sigma_4^2]/2 - \lambda & [\sigma_3^2 + \sigma_4^2]/2 - \lambda & \sigma_4^2 \end{pmatrix}$$

**Identità scalata.** Questa struttura ha una varianza costante. Si ipotizza che non ci sia alcuna correlazione tra gli elementi.

$$\begin{pmatrix} \sigma^2 & 0 & 0 & 0 \\ 0 & \sigma^2 & 0 & 0 \\ 0 & 0 & \sigma^2 & 0 \\ 0 & 0 & 0 & \sigma^2 \end{pmatrix}$$

**Spaziale: Power.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenei tra elementi.  $d_{ij}$  è la distanza Euclidea stimata tra la misurazione dei  $i^{\text{th}}$  e  $j^{\text{th}}$ .

$$\begin{pmatrix} \sigma^2 & \sigma^2 \rho d^{12} & \sigma^2 \rho d^{13} & \sigma^2 \rho d^{14} \\ \sigma^2 \rho d^{12} & \sigma^2 & \sigma^2 \rho d^{24} & \sigma^2 \rho d^{24} \\ \sigma^2 \rho d^{13} & \sigma^2 \rho d^{24} & \sigma^2 & \sigma^2 \rho d^{34} \\ \sigma^2 \rho d^{14} & \sigma^2 \rho d^{24} & \sigma^2 \rho d^{34} & \sigma^2 \end{pmatrix}$$

**Spaziale: Esponenziale.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenei tra elementi.  $d_{ij}$  è la distanza Euclidea stimata tra la misurazione dei  $i^{\text{th}}$  e  $j^{\text{th}}$ .

$$\begin{pmatrix} \sigma^2 & \sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \exp\{-d_{14}/\theta\} \end{pmatrix}$$

$$\begin{pmatrix}
\sigma^2 \exp\{-d_{12}/\theta\} & \sigma^2 & \sigma^2 \text{escp}\{-d_{23}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} \\
\sigma^2 \exp\{-d_{13}/\theta\} & \sigma^2 \text{escp}\{-d_{23}/\theta\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\theta\} \\
\sigma^2 \exp\{-d_{14}/\theta\} & \sigma^2 \exp\{-d_{24}/\theta\} & \sigma^1 \exp\{-d_{34}/\theta\} & \sigma^2
\end{pmatrix}$$

**Spaziale: Gaussiano.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenee tra elementi.  $d_{ij}$  è la distanza Euclidea stimata tra la misurazione dei  $i^{\text{th}}$  e  $j^{\text{th}}$ .

$$\begin{pmatrix}
\sigma^2 & \sigma^2 \text{escp}\{-d_{12}/\rho^2\} & \sigma^2 \text{esp}\{-d_{13}/\rho^2\} & \sigma^2 \exp\{-d_{14}/\rho^2\} \\
\sigma^2 \text{esp}\{-d_{12}/\rho^2\} & \sigma^2 & \sigma^2 \text{esp}\{-d_{23}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} \\
\sigma^2 \exp\{-d_{13}/\rho^2\} & \sigma^2 \text{esp}\{-d_{23}/\rho^2\} & \sigma^2 & \sigma^2 \exp\{-d_{34}/\rho^2\} \\
\sigma^2 \exp\{-d_{14}/\rho^2\} & \sigma^2 \exp\{-d_{24}/\rho^2\} & \sigma^2 \text{esp}\{-d_{34}/\rho^2\} & \sigma^2
\end{pmatrix}$$

**Spaziale: Linear.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenee tra elementi.  $d_{ij}$  è la distanza euclidea stimata tra la misurazione  $i^{\text{th}}$  e  $j^{\text{th}}$  e  $1_{ij}$  è una funzione indicatore che è 1 se  $\rho d_{ij} \leq 0$  e 0 altrimenti.

$$\begin{pmatrix}
\sigma^2 & \sigma^2 (1 - \rho d_{12}) 1_{12} & \sigma^2 (1 - \rho d_{13}) 1_{13} & \sigma^2 (1 - \rho d_{14}) 1_{14} \\
\sigma^2 (1 - \rho d_{12}) 1_{12} & \sigma^2 & \sigma^2 (1 - \rho d_{23}) 1_{23} & \sigma^2 (1 - \rho d_{24}) 1_{24} \\
\sigma^2 (1 - \rho d_{13}) 1_{13} & \sigma^2 (1 - \rho d_{23}) 1_{23} & \sigma^2 & \sigma^2 (1 - \rho d_{34}) 1_{34} \\
\sigma^2 (1 - \rho d_{14}) 1_{14} & \sigma^2 (1 - \rho d_{24}) 1_{24} & \sigma^2 (1 - \rho d_{34}) 1_{34} & \sigma^2
\end{pmatrix}$$

**Spaziale: Linear - log.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenee tra elementi.  $d_{ij}$  è la distanza euclidea stimata tra la misurazione  $i^{\text{th}}$  e  $j^{\text{th}}$  e  $1_{ij}$  è una funzione indicatore che è 1 se  $\rho \log(d_{ij}) \leq 0$  e 0 altrimenti.

$$\begin{pmatrix}
\sigma^2 & \sigma^2 (1 - \rho \log(d_{12})) 1_{12} & \sigma^2 (1 - \rho \log(d_{13})) 1_{13} & \sigma^2 (1 - \rho \log(d_{14})) 1_{14} \\
\sigma^2 (1 - \rho \log(d_{12})) 1_{12} & \sigma^2 & \sigma^2 (1 - \rho \log(d_{23})) 1_{23} & \sigma^2 (1 - \rho \log(d_{24})) 1_{24} \\
\sigma^2 (1 - \rho \log(d_{13})) 1_{13} & \sigma^2 (1 - \rho \log(d_{23})) 1_{23} & \sigma^2 & \sigma^2 (1 - \rho \log(d_{34})) 1_{34} \\
\sigma^2 (1 - \rho \log(d_{14})) 1_{14} & \sigma^2 (1 - \rho \log(d_{24})) 1_{24} & \sigma^2 (1 - \rho \log(d_{34})) 1_{34} & \sigma^2
\end{pmatrix}$$

**Spaziale: Sferico.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenee tra elementi.  $r_{ij} = d_{ij}/\rho$ , dove  $d_{ij}$  è la distanza Euclidea stimata tra la misurazione dei  $i^{\text{th}}$  e  $j^{\text{th}}$ .  $1_{ij}$  è una funzione indicatore che è 1 se  $d_{ij} \leq \rho$  e 0 in caso contrario.

$$\begin{pmatrix}
\sigma^2 & \sigma^2 (1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2 (1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2 (1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} \\
\sigma^2 (1 - 3/2r_{12} + 1/2r_{12}^3) 1_{12} & \sigma^2 & \sigma^2 (1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2 (1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} \\
\sigma^2 (1 - 3/2r_{13} + 1/2r_{13}^3) 1_{13} & \sigma^2 (1 - 3/2r_{23} + 1/2r_{23}^3) 1_{23} & \sigma^2 & \sigma^2 (1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} \\
\sigma^2 (1 - 3/2r_{14} + 1/2r_{14}^3) 1_{14} & \sigma^2 (1 - 3/2r_{24} + 1/2r_{24}^3) 1_{24} & \sigma^2 (1 - 3/2r_{34} + 1/2r_{34}^3) 1_{34} & \sigma^2
\end{pmatrix}$$

**Toeplitz.** Questa struttura di covarianza ha varianze omogenee e correlazioni eterogenee tra elementi. 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.

$$\begin{pmatrix}
\sigma^2 & \sigma^2 \rho_1 & \sigma^2 \rho_2 & \sigma^2 \rho_3 \\
\sigma^2 \rho_1 & \sigma^2 & \sigma^2 \rho_1 & \sigma^2 \rho_2 \\
\sigma^2 \rho_2 & \sigma^2 \rho_1 & \sigma^2 & \sigma^2 \rho_1
\end{pmatrix}$$

$$(\sigma^2 \rho_3 \qquad \qquad \qquad \sigma^2 \rho_2 \qquad \qquad \qquad \sigma^2 \rho_1 \qquad \qquad \qquad \sigma^2)$$

**Toeplitz: Eterogenico.** Questa struttura di covarianza ha varianze eterogenee e correlazioni eterogenee tra elementi. 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.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_1 & \sigma_3 \sigma_1 \rho_2 & \sigma_4 \sigma_1 \rho_3 \\ \sigma_2 \sigma_1 \rho_1 & \sigma_2^2 & \sigma_3 \sigma_2 \rho_1 & \sigma_4 \sigma_2 \rho_2 \\ \sigma_3 \sigma_1 \rho_2 & \sigma_3 \sigma_2 \rho_1 & \sigma_3^2 & \sigma_4 \sigma_3 \rho_1 \\ \sigma_4 \sigma_1 \rho_3 & \sigma_4 \sigma_2 \rho_2 & \sigma_4 \sigma_3 \rho_1 & \sigma_4^2 \end{pmatrix}$$

**Non strutturato.** Si tratta di una matrice di covarianza completamente generale.

$$\begin{pmatrix} \sigma_1^2 & \sigma_{21} & \sigma_{31} & \sigma_{41} \\ \sigma_{21} & \sigma_2^2 & \sigma_{32} & \sigma_{42} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \sigma_{43} \\ \sigma_{41} & \sigma_{42} & \sigma_{43} & \sigma_4^2 \end{pmatrix}$$

**Non Strutturato: Metrica di correlazione.** Questa struttura di covarianza ha varianze eterogenee e correlazioni eterogenee.

$$\begin{pmatrix} \sigma_1^2 & \sigma_2 \sigma_1 \rho_{21} & \sigma_3 \sigma_1 \rho_{31} & \sigma_4 \sigma_1 \rho_{41} \\ \sigma_2 \sigma_1 \rho_{21} & \sigma_2^2 & \sigma_3 \sigma_2 \rho_{32} & \sigma_4 \sigma_2 \rho_{42} \\ \sigma_3 \sigma_1 \rho_{31} & \sigma_3 \sigma_2 \rho_{32} & \sigma_3^2 & \sigma_4 \sigma_3 \rho_{43} \\ \sigma_4 \sigma_1 \rho_{41} & \sigma_4 \sigma_2 \rho_{42} & \sigma_4 \sigma_3 \rho_{43} & \sigma_4^2 \end{pmatrix}$$

**Componenti di varianza.** Questa struttura assegna una struttura di identità (ID) scalata a ciascuno degli effetti casuali specificati.

## Statistiche bayesiane

IBM SPSS Statistica fornisce supporto per le seguenti statistiche Bayesiane.

### Un esempio e Pandate di esempio T - test

La procedura di Inferenza di esempio Bayesiana One fornisce opzioni per rendere l'inferenza bayesiana su uno - campione e a due campioni accoppiati a t - test caratterizzando le distribuzioni posteriori. Quando si hanno dati normali, è possibile utilizzare un normale precedente per ottenere un normale posteriore.

### Test Binomiale Proporzione

L'Inferenza di esempio Bayesiana: la procedura binomiale fornisce le opzioni per l'esecuzione dell'inferenza bayesiana a campione sulla distribuzione binomiale. Il parametro di interesse è  $\pi$ , che indica la probabilità di successo in un numero fisso di prove che potrebbe portare a un esito positivo o negativo. Da notare che ogni processo è indipendente l'uno dall'altro, e la probabilità  $\pi$  rimane la stessa in ogni processo. Una variabile casuale binomiale può essere visualizzata come la somma di un numero fisso di prove Bernoulli indipendenti.

### Analisi distribuzione di Poisson

L'Inferenza di esempio Bayesiana: la procedura di Poisson fornisce opzioni per l'esecuzione dell'inferenza bayesiana su distribuzione di Poisson. La distribuzione di Poisson, un modello utile per gli eventi rari, presuppone che entro piccoli intervalli di tempo, la probabilità che un evento si verifichi sia proporzionale alla durata dell'attesa. Una probabilità a priori coniugata all'interno di una famiglia di distribuzione Gamma viene utilizzata quando si traccia l'inferenza statistica bayesiana su una distribuzione Poisson.

### **Campioni correlati**

Il design di inferenza del campione correlato Bayesiana è abbastanza simile all'inferenza bayesiana a campione in termini di gestione dei campioni accoppiati. È possibile specificare nomi variabile in coppie ed eseguire l'analisi bayesiana sulla differenza media.

### **Campioni Indipendenti T - test**

La procedura di inferenza del campione indipendente bayesiana fornisce opzioni per l'utilizzo di una variabile di gruppo per definire due gruppi indipendenti, e rendere l'inferenza bayesiana sulla differenza dei due mezzi di gruppo. È possibile stimare i fattori di Bayes utilizzando approcci diversi, e caratterizzare anche la distribuzione posteriore desiderata o ipotizzando che le varianze siano note o sconosciute.

### **Pairwise Correlazione (Pearson)**

L'inferenza bayesiana circa il coefficiente di correlazione di Pearson misura la relazione lineare tra due variabili di scala seguendo congiuntamente una distribuzione normale bivariata. L'inferenza statistica convenzionale relativa al coefficiente di correlazione è stata ampiamente discussa, e la sua pratica è stata a lungo offerta in IBM SPSS Statistics. Il disegno dell'inferenza bayesiana circa il coefficiente di correlazione di Pearson consente di disegnare l'inferenza bayesiana stimando i fattori di Bayes e caratterizzando le distribuzioni posteriori.

### **Regressione lineare**

L'inferenza bayesiana su Linear Regression è un metodo statistico ampiamente utilizzato nella modellazione quantitativa. La regressione lineare è un approccio di base e standard in cui i ricercatori utilizzano i valori di diverse variabili per spiegare o prevedere valori di un risultato su scala. La regressione lineare univariata bayesiana è un approccio alla regressione lineare dove l'analisi statistica è intrapresa nel contesto dell'inferenza bayesiana.

### **ANOVA a una via**

La procedura Bayesiano One-Way ANOVA produce un'analisi unidirezionale della varianza per una variabile dipendente quantitativa da una variabile singolo fattore (indipendente). L'analisi della varianza consente di verificare l'ipotesi di uguaglianza di più medie. SPSS Statistics supporta i fattori di Bayes, i priori coniugati e i priori non informativi.

### **Tronco - Modelli di regressione lineare**

Il design per testare l'indipendenza di due fattori richiede due variabili categoriali per la costruzione di un tavolo di contingenza, e rende l'inferenza bayesiana sull'associazione a colonne di fila. È possibile stimare i fattori di Bayes assumendo diversi modelli, e caratterizzare la distribuzione posteriore desiderata simulando l'intervallo credibile simultaneo per i termini di interazione.

### **Misure ripetute ANOVA a una via**

Il Bayesiano One-way Ripetuto Misure ANOVA procedura misura un fattore dallo stesso soggetto ad ogni singolo punto temporale o condizione, e consente di incrociare i soggetti all'interno dei livelli. Si suppone che ogni soggetto abbia un'osservazione unica per ogni punto o condizione (in quanto tale, l'interazione oggetto - trattamento non è contabilizzata).

## **Inferenza a un campione bayesiana: Normale**

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

La procedura di inferenza bayesiana su un campione: La procedura normale fornisce opzioni per fare inferenza bayesiana su t-test a un campione e a due campioni, caratterizzando le distribuzioni posteriori. Quando si dispone di dati normali, è possibile utilizzare un antecedente normale per ottenere un posteriore normale.

1. Dai menu, scegliere:

**Analizzare le > statistiche bayesiane > Un campione normale**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare le **variabili di test** appropriate dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile.

**Nota:** L'elenco delle variabili disponibili contiene tutte le variabili, ad eccezione delle variabili Data e Stringa.

3. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 2. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

1  
2

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

4. Selezionare e/o inserire le impostazioni appropriate per la **varianza dei dati e i valori delle ipotesi**. La tabella riflette le variabili attualmente presenti nell'elenco **Variabili di test**. Quando si aggiungono o si rimuovono variabili dall'elenco delle **variabili di test**, la tabella aggiunge o rimuove automaticamente le stesse variabili dalle sue colonne di variabili.

- Quando una o più variabili sono presenti nell'elenco **Variabili del test**, le colonne **Variabile nota** e **Valore della varianza** sono abilitate.

**Varianza nota**

Selezionare questa opzione per ogni variabile quando la varianza è nota.

**Valore varianza**

Un parametro opzionale che specifica il valore della varianza, se noto, per i dati osservati.

- Quando una o più variabili sono presenti nell'elenco **Variabili test** e non è selezionata la voce **Caratterizza distribuzione posteriore**, le colonne **Valore test nullo** e **Valore g** sono abilitate.

<sup>1</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>2</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Valore test null

Parametro obbligatorio che specifica il valore nullo nella stima del fattore di Bayes. È consentito un solo valore e 0 è il valore predefinito.

### g Valore

Specifica il valore per definire  $\psi^2 = g\sigma^2_x$  nella stima del fattore di Bayes. Quando viene specificato il **valore di varianza**, il **valore g** è impostato di default su 1. Se il **valore della varianza** non è specificato, è possibile specificare un *g* fisso oppure omettere il valore per integrarlo.

5. È possibile fare clic su **Criteri** per specificare le impostazioni di “Bayesiana One Sample Inference: Criteri” a pagina 109 (percentuale dell'intervallo credibile, opzioni per i valori mancanti e impostazioni del metodo numerico), oppure fare clic su **Priori** per specificare le impostazioni di “Bayesiana One Sample Inference: Priori normali” a pagina 110 (tipo di priori, come parametri di inferenza, media data varianza o precisione).

## Bayesiana One Sample Inference: Criteri

È possibile specificare i seguenti criteri di analisi per la tua Inferenza Bayesiano One - Sample:

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

### Valori mancanti

Specificare il metodo in cui controllare i valori mancanti.

#### Escludi casi pairwise

Questa è l'impostazione predefinita ed esclude record con valori mancanti su un'analisi per analisi. I record che includono valori mancanti, per un campo che viene utilizzato per un test specifico, vengono omessi dal test.

#### Escludi casi listwise

Questa impostazione esclude record che includono i valori mancanti listwise. I record che includono valori mancanti per qualsiasi campo che viene denominato in qualsiasi comando sono esclusi da tutte le analisi.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

### Metodo numerico

Specificare il metodo numerico utilizzato per stimare l'integrale.

#### Quadratura Gauss-Lobatto adattiva

Questa è l'impostazione predefinita e richiama l'approccio Adaptive Gauss - Lobatta Quadratura.

#### Tolleranza

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001. L'opzione è disponibile solo quando viene selezionata l'impostazione **Adaptive Gauss - Lobatta Quadratura**.

#### Numero massimo di iterazioni

Specificare il numero massimo di iterazioni di metodo Adaptive Gauss - Lobatte Quadrature. Il valore deve essere un intero positivo. L'impostazione predefinita è 2000. L'opzione è disponibile solo quando viene selezionata l'impostazione **Adaptive Gauss - Lobatta Quadratura**.

### Approssimazione di Monte Carlo

Questa opzione richiama l'approccio di approssimazione Monte Carlo.

#### Imposta seme personalizzato

Quando selezionato, è possibile specificare un valore di seed personalizzato nel campo **Seed**.

#### Seme

Specificare un seme casuale fissato per il metodo Monte Carlo Approfondimento. Il valore deve essere un intero positivo. Per impostazione predefinita viene assegnato un valore seed casuale.

### Numero di campioni Monte Carlo

Specificare il numero di punti che vengono campionati per l'approssimazione Monte Carlo. Il valore deve essere un intero positivo. Il valore predefinito è 1000000. L'opzione è disponibile solo quando viene selezionata l'impostazione **Monte Carlo Approfondimento**.

## Bayesiana One Sample Inference: Priori normali

È possibile specificare i seguenti criteri di distribuzione preventiva per la tua Inferenza Bayesiano One - Sample:

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

### Probabilità a priori su varianza/precisione

Fornisce opzioni per definire i valori di varianza e precisione.

#### Varianza

Selezionare per specificare la distribuzione precedente per il parametro varianza. Quando questa opzione viene selezionata, l'elenco **Primi Distribuzione** fornisce le seguenti opzioni:

**Nota:** Quando la varianza dei dati è già specificata per alcune variabili, le seguenti impostazioni vengono ignorate per quelle variabili.

- **Diffuso** - L'impostazione predefinita. Specifica il precedente diffuso.
- **Chi - quadrato inverso** - Specifica la distribuzione e i parametri per l'inverso -  $\chi^2(\nu_0, \sigma^2_0)$ , dove  $\nu_0 > 0$  è il grado di libertà e  $\sigma^2_0 > 0$  è il parametro di scala.
- **Inverse Gamma** - Specifica la distribuzione e i parametri per l'inverso - Gamma ( $\alpha_0, \beta_0$ ), dove  $\alpha_0 > 0$  è il parametro di forma, e  $\beta_0 > 0$  è il parametro di scala.
- **Jeffreys S2** - Specifica il preventivo non informativo  $1/\sigma^2_0$ .
- **Jeffreys S4** - Specifica il preventivo non informativo  $1/\sigma^4_0$ .

#### Precisione

Selezionare per specificare la distribuzione precedente per il parametro di precisione. Quando questa opzione viene selezionata, l'elenco **Primi Distribuzione** fornisce le seguenti opzioni:

- **Gamma** - Specifica la distribuzione e i parametri per Gamma ( $\alpha_0, \beta_0$ ), dove  $\alpha_0 > 0$  è il parametro di forma, e  $\beta_0 > 0$  è il parametro di scala.
- **Chi - Square** - Specifica la distribuzione e i parametri per  $\chi^2(\nu_0)$ , dove  $\nu_0 > 0$  è il grado di libertà.

#### Parametri forma

Specificare il parametro shape  $a_0$  per la distribuzione Inverse - Gamma. È necessario inserire un valore unico superiore a 0.

#### Parametro di scala

Specificare il parametro di scala  $b_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0. Più grande è il parametro di scala, più si diffonde la distribuzione.

### Probabilità a priori su varianza data della media/precisione

Specificare la distribuzione precedente per il parametro medio che è condizionato alla varianza o al parametro di precisione.

#### Normale

Specificare la distribuzione e i parametri per Normale ( $\mu_0, R^{-1}_0\sigma^2_0$ ) sulla varianza o Normale ( $\mu_0, R_0/\sigma^2_0$ ) sulla precisione, dove  $\mu_0 \in (-\infty, \infty)$  e  $\sigma^2 > 0$ .

**Parametro ubicazione**

Inserire un valore numerico che specifica il parametro di posizione per la distribuzione.

**Parametro di scala**

Specificare il parametro di scala  $b_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0.

**Kappa**

Specificare il valore di  $K_0$  in Normale ( $\mu_0, K^{-1}_0\sigma^2_0$ ) o Normale ( $\mu_0, K_0/\sigma^2_0$ ). È necessario inserire un valore unico superiore a 0 (1 è il valore predefinito).

**Diffuso**

L'impostazione predefinita che specifica il diffuso precedente 1.

**Inferenza a un campione bayesiana: Binomiale**

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

L'Inferenza di esempio Bayesiana: la procedura binomiale fornisce le opzioni per l'esecuzione dell'inferenza bayesiana a campione sulla distribuzione binomiale. Il parametro di interesse è  $\pi$ , che indica la probabilità di successo in un numero fisso di prove che potrebbe portare a un esito positivo o negativo. Da notare che ogni processo è indipendente l'uno dall'altro, e la probabilità  $\pi$  rimane la stessa in ogni processo. Una variabile casuale binomiale può essere visualizzata come la somma di un numero fisso di prove Bernoulli indipendenti.

Anche se non è necessario, un preventivo della famiglia di distribuzione Beta viene normalmente scelto quando si stima un parametro binomiale. La Famiglia Beta è coniugata per la famiglia binomiale, e come tale porta alla distribuzione a posteriori con una forma chiusa ancora nella famiglia di distribuzione Beta.

1. Dai menu, scegliere:

**Analisi > Statistiche Bayesiane > Un esempio Binomiale**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare le **variabili di test** appropriate dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile.

**Nota:** L'elenco delle variabili disponibili contiene tutte le variabili, ad eccezione delle variabili Data e Stringa.

3. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

<i>Tabella 3. Soglie comunemente utilizzate per definire la significatività dell'evidenza</i>					
<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1

*Tabella 3. Soglie comunemente utilizzate per definire la significatività dell'evidenza (Continua)*

<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

3

4

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.
4. Selezionare e / o inserire le impostazioni appropriate **Success Categories and Ipotesi Values** . La tabella riflette le variabili che si trovano attualmente nella lista **Variabili di test** . Poiché le variabili vengono aggiunte o eliminate dalle **Variabili di test**, la tabella aggiunge automaticamente o rimuove le stesse variabili dalle sue colonne di coppia variabili.
- Quando **Caratterizzare La Distribuzione Posteriore** viene selezionata come **Analisi Bayesiana**, la colonna **Categorie di successo** è abilitata.
  - Quando **Stima Bayes Factor** o **Utilizza entrambi i metodi** sono selezionati come **Analisi Bayesiana**, tutte le colonne modificabili sono abilitate.

#### **Punto nullo**

Abilita e disabilita l'opzione **Null Proporzione** . Quando l'impostazione è abilitata, sia le opzioni **Null Precedente Shape** che **Null Preventivo Scale** sono disabilitate.

#### **Forma probabilità a priori null**

Specifica il parametro di forma  $\alpha_0$  nell'ipotesi null dell'inferenza binomiale.

#### **Scala probabilità a priori null**

Specifica il parametro di scala  $b_0$  nell'ipotesi null di inferenza binomiale.

#### **Proporzione nulla**

Specifica il parametro di forma  $\alpha_0$  e il parametro di scala  $b_0$  sotto l'ipotesi null per una distribuzione a priori coniugata (per contenere le probabilità a priori di Beta e di Haldane). La gamma valida è valori numerici compresi tra 0 e 1.

#### **Forma probabilità a priori alternativa**

Un parametro obbligatorio per specificare  $\alpha_0$  nell'ipotesi alternativa di inferenza binomiale se il fattore di Bayes deve essere stimato.

#### **Scala probabilità a priori alternativa**

Un parametro obbligatorio per specificare  $b_0$  nell'ipotesi alternativa di inferenza binomiale se il fattore di Bayes deve essere stimato.

#### **Categorie riuscite**

Fornisce opzioni per definire le distribuzioni precedenti coniugate. Le opzioni fornite specificano come è definito il successo, per le variabili numeriche e di stringa, quando i valori di dati vengono testati contro il valore di prova.

<sup>3</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>4</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Ultima categoria**

L'impostazione predefinita che esegue il test binomiale utilizzando l'ultimo valore numerico rilevato nella categoria dopo che è ordinato in ordine crescente.

**Prima categoria**

Esegue il test binomiale utilizzando il primo valore numerico trovato nella categoria dopo che viene ordinato in ordine crescente.

**Punto centrale**

Utilizza i valori numerici  $\geq$  il midpoint come casi. Un valore di midpoint è la media dei dati di esempio minimi e massimi.

**Punto di divisione**

Utilizza i valori numerici  $\geq$  un valore di cutoff specificato come casi. L'impostazione deve essere un valore numerico unico.

**Livello**

I valori stringa specificati dall'utente (possono essere più di 1) come casi. Utilizzare le virgole per separare i diversi valori.

5. È possibile opzionalmente fare clic su **Criteri** per specificare le impostazioni “[Bayesiana One Sample Inference: Criteri](#)” a pagina 109 (percentuale di intervallo credibile, opzioni di valori mancanti e impostazioni del metodo numerico) oppure fare clic su **Priori** per specificare le impostazioni “[Bayesiana One Sample Inference: Binomiale / Poisson Priors](#)” a pagina 113 (coniugate o distribuzioni precedenti personalizzate).

## Bayesiana One Sample Inference: Binomiale / Poisson Priors

È possibile specificare i seguenti criteri di distribuzione preventiva per la tua Inferenza Bayesiana One - Sample:

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

**Parametri forma**

Per le probabilità a priori binomiali, specificare il parametro shape  $a_0$  per la distribuzione beta.

Per le probabilità a priori di Poisson, specificare il parametro shape  $a_0$  per la distribuzione gamma.

È necessario inserire un valore unico superiore a 0.

**Parametro di scala**

Per le probabilità a priori binomiali, specificare il parametro di scala  $b_0$  per la distribuzione beta.

Per le probabilità a priori di Poisson, specificare il parametro di scala  $b_0$  per la distribuzione gamma.

È necessario inserire un valore unico superiore a 0.

## Inferenza a un campione bayesiana: Poisson

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

L'Inferenza di esempio Bayesiana: la procedura di Poisson fornisce opzioni per l'esecuzione dell'inferenza bayesiana su distribuzione di Poisson. La distribuzione di Poisson, un modello utile per gli eventi rari, presuppone che entro piccoli intervalli di tempo, la probabilità che un evento si verifichi sia proporzionale alla durata dell'attesa. Una probabilità a priori coniugata all'interno di una famiglia di distribuzione Gamma viene utilizzata quando si traccia l'inferenza statistica bayesiana su una distribuzione Poisson.

1. Dai menu, scegliere:

**Analisi > Statistiche Bayesiane > One Sample Poisson**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare le **variabili di test** appropriate dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile.

**Nota:** L'elenco delle variabili disponibili contiene tutte le variabili, ad eccezione delle variabili Data e Stringa.

3. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 4. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

5

6

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.
4. Selezionare e / o inserire le impostazioni appropriate **Ipotesi Valori** . La tabella riflette le variabili che si trovano attualmente nella lista **Variabili di test** . Poiché le variabili vengono aggiunte o eliminate dalle **Variabili di test** , la tabella aggiunge automaticamente o rimuove le stesse variabili dalle sue colonne di coppia variabili.
- Quando **Caratterizzare La Distribuzione Posteriore** viene selezionata come **Analisi Bayesiana**, nessuna delle colonne è abilitata.
  - Quando **Stima Bayes Factor** o **Utilizza entrambi i metodi** sono selezionati come **Analisi Bayesiana**, tutte le colonne modificabili sono abilitate.

<sup>5</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>6</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Punto nullo**

Abilita e disabilita l'opzione **Null Rate**. Quando l'impostazione è abilitata, sia le opzioni **Null Precedente Shape** che **Null Preventivo Scale** sono disabilitate.

**Forma probabilità a priori null**

Specifica il valore del parametro di forma  $a_0$  nell'ipotesi null dell'inferenza di Poisson.

**Scala probabilità a priori null**

Specifica il parametro di scala  $b_0$  nell'ipotesi null dell'inferenza di Poisson.

**Tasso nullo**

Specifica il parametro di forma  $a_0$  e il parametro di scala  $b_0$  nell'ipotesi null per una distribuzione a priori coniugata (per adattare la relazione Poisson - Gamma). Il valore minimo deve essere un grater di valore numerico di 0; il valore massimo deve essere un valore max doppio.

**Forma probabilità a priori alternativa**

Un parametro obbligatorio per specificare  $a_1$  nell'ipotesi alternativa dell'inferenza di Poisson se il fattore di Bayes deve essere stimato.

**Scala probabilità a priori alternativa**

Un parametro obbligatorio per specificare  $b_1$  sotto l'ipotesi alternativa dell'inferenza di Poisson se il fattore di Bayes deve essere stimato.

5. È possibile opzionalmente fare clic su **Criteri** per specificare le impostazioni "Bayesiana One Sample Inference: Criteri" a pagina 109 (percentuale di intervallo credibile, opzioni di valori mancanti e impostazioni del metodo numerico) oppure fare clic su **Priori** per specificare le impostazioni "Bayesiana One Sample Inference: Binomiale / Poisson Priors" a pagina 113 (coniugate o distribuzioni precedenti personalizzate).

## Inferenza campione correlato bayesiana: Normale

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

L'Inferenza di Campionamento Correlato Bayesiana: la procedura normale fornisce opzioni di inferenza bayesiana a campione per campioni accoppiati. È possibile specificare nomi variabile in coppie ed eseguire l'analisi bayesiana sulla differenza media.

1. Dai menu, scegliere:

**Analizzare > Statistiche Bayesiane > I campioni correlati normali**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare l'apposito **Variabili accoppiate** dall'elenco **Variabili disponibili**. Deve essere selezionata almeno una coppia di variabili di origine e non possono essere selezionate più di due variabili di origine per qualsiasi coppia data set.

**Nota:** L'elenco delle variabili disponibili fornisce tutte le variabili tranne le variabili String.

3. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 5. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>	<b>Fattore di Bayes</b>	<b>Categoria di prova</b>
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

7

8

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

4. Selezionare e / o inserire le impostazioni appropriate **Data Variance e Ipotesi Valori** . La tabella riflette le coppie variabili che sono attualmente nella lista **Variabili accoppiate** . Poiché le coppie variabili vengono aggiunte o rimosse dall'elenco **Variabili accoppiate** , la tabella aggiunge automaticamente o rimuove le stesse coppie di variabili dalle sue colonne di coppia variabili.

- Quando una o più coppie variabili si trovano nell'elenco **Variabili accoppiate** , sono abilitate le colonne **Variance Knowne Variance Value** .

#### **Varianza nota**

Selezionare questa opzione per ogni variabile quando la varianza è nota.

#### **Valore varianza**

Un parametro opzionale che specifica il valore della varianza, se noto, per i dati osservati.

- Quando una o più coppie variabili si trovano nell'elenco **Variabili accoppiate** e **Caratterizzate Distribuzione Posterior** non è selezionata, le colonne **Valore del test Null** e **g Value** sono abilitate.

#### **Valore test null**

Un parametro obbligatorio che specifica il valore nullo nella stima del fattore Bayes. È consentito un solo valore e 0 è il valore predefinito.

#### **g Valore**

Specifica il valore per definire  $\psi^1 = g\sigma^2_x$  nella stima del fattore di Bayes. Quando viene specificato il **Valore Varianza** , il **valore g** viene impostato per default su 1. Quando il **Valore Varianza** non viene specificato, è possibile specificare un **g** fisso o omettere il valore per integrarlo.

5. È possibile opzionalmente fare clic su **Criteri** per specificare le impostazioni “Bayesiana One Sample Inference: Criteri” a pagina 109 (percentuale di intervallo credibile, opzioni di valori mancanti e impostazioni del metodo numerico) oppure fare clic su **Priori** per specificare le impostazioni

<sup>7</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>8</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Bayesiano Independent - Inferenza di esempio

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

La procedura di Inferenza di esempio di Bayesiano Independent fornisce opzioni per l'utilizzo di una variabile di gruppo per definire due gruppi indipendenti, e rendere l'inferenza bayesiana sulla differenza dei due mezzi di gruppo. È possibile stimare i fattori di Bayes utilizzando approcci diversi, e caratterizzare anche la distribuzione posteriore desiderata o ipotizzando che le varianze siano note o sconosciute.

1. Dai menu, scegliere:

**Analizzare > Statistiche Bayesiano > Campioni Indipendenti Normali**

2. Selezionare l'apposito **Variabili di test** dall'elenco delle variabili di origine. Deve essere selezionata almeno una variabile di origine.

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

3. Selezionare l'apposita **Variabile di raggruppamento** dall'elenco **Variabili disponibili**. Una variabile di raggruppamento definisce due gruppi per il non accoppiato *t*- test. La variabile di raggruppamento selezionata può essere una variabile numerica o stringa.

4. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 6. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

9

10

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.
5. Utilizzare le opzioni **Definisci gruppi** per definire due gruppi per il test *t* specificando due valori (per le variabili stringa), o due valori, un midpoint o un punto di taglio (per variabili numeriche).
- Nota:** I valori specificati devono esistere nella variabile, altrimenti un messaggio di errore visualizza per indicare che almeno uno dei gruppi è vuoto.
- Per variabili numeriche:
- **Usa i valori specificati.** Inserire un valore per il Gruppo 1 e un altro valore per il Gruppo 2. I casi con altri valori sono esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
  - **Utilizza valore punto centrale.** Quando selezionate, i gruppi sono separati in valori  $<$  e  $\geq$  midpoint.
  - **Utilizza punto di divisione.**
    - **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori al punto di divisione da un gruppo e i casi con valori maggiori o uguali al punto di divisione dall'altro gruppo.
- Per le variabili di raggruppamento delle stringhe, inserire una stringa per il Gruppo 1 e un altro valore per il Gruppo 2, come *si* e *no*. I casi con altre stringhe sono esclusi dall'analisi.
6. È possibile opzionalmente fare clic su **Criteri** per specificare le impostazioni “Inferenza campione indipendente bayesiana: Criteri” a pagina 119 (percentuale di intervallo credibile, opzioni di valori mancanti e impostazioni del metodo di quadratura adattativo), fare clic su **Priori** per specificare le impostazioni “Inferenza campione indipendente bayesiana: distribuzione probabilità a priori” a pagina 119 (varianza dati, prima della varianza e prima di condizionale sulla varianza) oppure fare clic su **Stima di Bayes Factor** per specificare le impostazioni “Bayesiano Independent - Esempio Inference: stima Bayes Factor” a pagina 120 .

### Bayesiano Independent - Esempio Inference Definire i gruppi (numerici)

Per le variabili di raggruppamento numerico, definire i due gruppi per il test *t* specificando due valori, un midpoint o un punto di taglio.

**Nota:** I valori specificati devono esistere nella variabile, altrimenti un messaggio di errore visualizza per indicare che almeno uno dei gruppi è vuoto.

- **Usa i valori specificati.** Inserire un valore per il Gruppo 1 e un altro valore per il Gruppo 2. I casi con altri valori sono esclusi dall'analisi. Non è necessario specificare numeri interi (ad esempio, 6,25 e 12,5 sono validi).
- **Utilizza valore punto centrale.** Quando selezionate, i gruppi sono separati in valori  $<$  e  $\geq$  midpoint.
- **Utilizza punto di divisione.**
  - **Punto di divisione.** Immettere un numero che suddivide i valori della variabile di raggruppamento in due insiemi Assegna i casi con valori minori al punto di divisione da un gruppo e i casi con valori maggiori o uguali al punto di divisione dall'altro gruppo.

### Bayesiano Independent - Esempio Inference Definisci gruppi (stringa)

Per le variabili di raggruppamento delle stringhe, inserire una stringa per il Gruppo 1 e un altro valore per il Gruppo 2, come *si* e *no*. I casi con altre stringhe sono esclusi dall'analisi.

**Nota:** I valori specificati devono esistere nella variabile, altrimenti un messaggio di errore visualizza per indicare che almeno uno dei gruppi è vuoto.

<sup>9</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>10</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

## Inferenza campione indipendente bayesiana: Criteri

È possibile specificare i seguenti criteri di analisi per la tua Inferenza di Campione Indipendente di Bayesiano:

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

### Valori mancanti

Specificare il metodo in cui controllare i valori mancanti.

#### Escludi casi pairwise

Questa è l'impostazione predefinita ed esclude record con valori mancanti su un'analisi per analisi. I record che includono valori mancanti, per un campo che viene utilizzato per un test specifico, vengono omessi dal test.

#### Escludi casi listwise

Questa impostazione esclude record che includono i valori mancanti listwise. I record che includono valori mancanti per qualsiasi campo che viene denominato in qualsiasi comando sono esclusi da tutte le analisi.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

### Metodo di quadratura adattiva

Specificare la tolleranza e i valori di iterazione massimi per il metodo Adaptive Quadratura Quadratura.

#### Tolleranza

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001.

#### Numero massimo di iterazioni

Specificare il numero massimo di iterazioni di metodo Adaptive Quadrature. Il valore deve essere un intero positivo. L'impostazione predefinita è 500.

## Inferenza campione indipendente bayesiana: distribuzione probabilità a priori

È possibile specificare i seguenti criteri di distribuzione preventiva per l'Inferenza Bayesiano Independent - Sample:

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

### Varianza dati

Fornisce le opzioni per la definizione delle impostazioni di varianza dei dati.

#### Varianza nota

Quando selezionato, consente di inserire due varianze di gruppo note. Entrambi i valori devono essere > 0.

#### Varianza gruppo 1

Inserire il primo valore di varianza di gruppo conosciuto.

#### Varianza gruppo 2

Inserire il secondo valore di varianza di gruppo conosciuto.

#### Presumi varianza uguale

Controlla se le due varianze di gruppo sono assunte per essere uguali. Per impostazione predefinita, si ipotizza che le varianze di gruppo siano disuguali. Questa impostazione viene ignorata quando vengono immessi valori per le due varianze di gruppo.

**Presumi varianza non uguale**

Controlla se le due varianze di gruppo sono considerate non uguali. Per impostazione predefinita, si ipotizza che le varianze di gruppo siano disuguali. Questa impostazione viene ignorata quando vengono immessi valori per le due varianze di gruppo.

**Probabilità a priori su varianza**

Specificare la distribuzione preventiva per le due varianze uguali.

**Jeffreys**

Quando viene selezionata, viene utilizzata una distribuzione preventiva non informativa (oggettiva) per uno spazio di parametro.

**Inverse-ChiSquare**

Specifica la distribuzione di probabilità continua di una variabile casuale con valore positivo e i parametri per l'inverso -  $\chi^2(v_0, \sigma^2_0)$ , dove  $v_0 > 0$  è il grado di libertà e  $\sigma^2_0 > 0$  è il parametro di scala.

**Gradi di libertà**

Specificare un valore per il numero di valori nel calcolo finale che sono liberi di variare.

**Parametro di scala**

Specificare il parametro di scala  $\sigma^2_0 > 0$  per inverso -  $\chi^2(v_0, \sigma^2_0)$ . È necessario inserire un valore unico superiore a 0. Più grande è il parametro di scala, più si diffonde la distribuzione.

**Probabilità a priori su varianza condizionale media**

Fornisce opzioni per specificare la distribuzione preventiva per i due mezzi di gruppo.

**Nota:** Le Opzioni **Diffuse** e **Normale** sono disponibili solo quando viene selezionata l'opzione **Varianza nota**.

**Diffuso**

Impostazione predefinita. Specifica il precedente diffuso.

**Normale**

Quando selezionato, è necessario specificare i parametri di posizione e di scala per i mezzi di gruppo definiti.

**Parametro ubicazione**

Inserire un valore numerico che specifica i parametri di posizione per le distribuzioni di gruppo.

**Parametro di scala**

Specificare il parametro di scala  $\sigma^2_0 > 0$  per inverso -  $\chi^2(v_0, \sigma^2_0)$ . Per ogni gruppo è necessario inserire un valore unico superiore a 0. Più grande è il parametro di scala, più si diffonde la distribuzione.

**Bayesiano Independent - Esempio Inference: stima Bayes Factor**

È possibile specificare il metodo che viene utilizzato per stimare il fattore di Bayes.

**Metodo di Rouder**

Quando selezionato, richiama l'approccio di Rouder. Questa è l'impostazione predefinita

**Metodo di Gonen**

Quando selezionato, richiama l'approccio Gonen e devi specificare le seguenti impostazioni di dimensione dell'effetto:

**Dimensione media per effetto**

Inserire un valore che specifica la differenza media tra i due gruppi.

**Dimensione varianza per effetto**

Inserire un valore che specifica la varianza per i due gruppi. Il valore deve essere  $> 0$ .

**Metodo iper-probabilità a priori**

Quando selezionato, richiama l'approccio hyper -g dove è necessario specificare un valore unico.

Inserire un valore compreso tra -1 e -0.5 nel campo **Parametro forma**. Il valore predefinito è -0.75.

## Inferenza Bayessiana su correlazione di Pearson

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

Il coefficiente di correlazione Pearson misura la relazione lineare tra due variabili di scala seguendo congiuntamente una distribuzione normale bivariata. L'inferenza statistica convenzionale sul coefficiente di correlazione è stata ampiamente discussa e la sua pratica è stata a lungo offerta in IBM SPSS Statistica. Il design dell'inferenza bayesiana circa il coefficiente di correlazione di Pearson permette agli utenti di disegnare l'inferenza bayesiana stimando i fattori di Bayes e caratterizzando le distribuzioni posteriori.

1. Dai menu, scegliere:

**Analizzare > Statistiche Bayesiane > Correlazione di Pearson**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare le **Variabili di test** appropriate da utilizzare per l'inferenza di correlazione pairwise dall'elenco **Variabili disponibili**. Devono essere selezionate almeno due variabili di origine. Quando vengono selezionate più di due variabili, l'analisi viene eseguita su tutte le combinazioni pairwise delle variabili selezionate.

3. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

11

12

<sup>11</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.
4. Specificare il **Numero massimo di complotti** da vedere nell'output. Una serie di tracciati può contenere 3 complotti sullo stesso riquadro. I tracciati vengono generati in ordine dalla prima variabile rispetto alle variabili rimanenti, quindi la seconda variabile versus le variabili rimanenti, e così via. Il valore intero definito deve essere compreso tra 0 e 50. Per impostazione predefinita, 10 serie di tracciati sono in uscita per ospitare cinque variabili. Questa opzione non è disponibile quando **Stima Bayes Factor** viene selezionato.
  5. You can optionally click **Criteri** to specify “Correlazione di Pearson bayesiana: criteri” a pagina 122 settings (credible interval percentage, missing values options, and numerical method settings), click **Priori** to specify “Correlazione di Pearson bayesiana: distribuzione a priori” a pagina 122 settings (value  $c$  for the prior  $p(\rho) \propto (1 - \rho^2)^c$ , or click **Fattore di Bayes** to specify “Bayesiano Independent - Esempio Inference: stima Bayes Factor” a pagina 120 settings.

## Correlazione di Pearson bayesiana: criteri

È possibile specificare i seguenti criteri di analisi per l'Inferenza Bayesiana Pearson Correlazione (pairwise).

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

### Valori mancanti

Specificare il metodo in cui controllare i valori mancanti.

#### Escludi casi pairwise

Questa impostazione esclude record che includono i valori mancanti pairwise.

#### Escludi casi listwise

Questa impostazione esclude record che includono i valori mancanti listwise. I record che includono valori mancanti per qualsiasi campo che viene denominato in qualsiasi comando sono esclusi da tutte le analisi.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

### Metodo numerico

Specificare il metodo numerico utilizzato per stimare l'integrale.

### Imposta seme personalizzato

Quando selezionato, è possibile specificare un valore di seed personalizzato nel campo **Seed**.

### Tolleranza

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001.

### Numero massimo di iterazioni

Specificare il numero massimo di iterazioni di metodo. Il valore deve essere un intero positivo. L'impostazione predefinita è 2000.

### Numero di campioni Monte Carlo

Specificare il numero di punti che vengono campionati per l'approssimazione Monte Carlo. Il valore deve essere un intero positivo. Il valore predefinito è 10000.

### Campioni simulati per distribuzione a posteriori

Specificare il numero di campioni che vengono utilizzati per disegnare la distribuzione posteriore desiderata. Il valore predefinito è 10000.

## Correlazione di Pearson bayesiana: distribuzione a priori

È possibile specificare il valore  $c$  per il precedente  $p(\rho)$  precedente  $(1-\rho^2)^c$ .

<sup>12</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

#### **Uniforme (c = 0)**

Quando viene selezionata la uniforme precedente viene utilizzata.

#### **Jeffreys (c = -1.5)**

Quando viene selezionata, viene utilizzata una distribuzione preventiva non informativa.

#### **Imposta valore c personalizzato**

Quando selezionato, è possibile specificare un valore **c personalizzato** personalizzato. Ogni singolo numero reale è consentito.

## **Correlazione di Pearson bayesiana: fattore di Bayes**

È possibile specificare il metodo che viene utilizzato per stimare il fattore di Bayes. Le seguenti opzioni sono disponibili solo quando viene selezionata l'opzione **Stima Bayes Factor** o **User Entrambi i Metodi**.

#### **Fattore di Bayes JZS**

Quando selezionato, richiama l'approccio di Zellner-Siow. Questa è l'impostazione predefinita.

#### **Fattore di Bayes frazionale**

Quando selezionato, è possibile specificare il fattore di Bayes frazionale e il valore di ipotesi nullo. Per il fattore di Bayes frazionale è necessario specificare un valore di ipotesi di valore (0, 1). Il valore predefinito è 0.5.

## **Inferenza Bayessiana su modelli di regressione lineari**

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

Regressione è un metodo statistico ampiamente utilizzato nella modellazione quantitativa. La regressione lineare è un approccio di base e standard in cui i ricercatori utilizzano i valori di diverse variabili per spiegare o prevedere valori di un risultato su scala. La regressione lineare bayesiana univariata è un approccio a regressione lineare dove l'analisi statistica è intrapresa nell'ambito dell'inferenza bayesiana.

È possibile invocare la procedura di regressione e definire un modello completo.

1. Dai menu, scegliere:

#### **Analizzare > Statistiche Bayesiane > Regressione Lineare**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile dipendente singola, non stringa, dalla lista **Variabili disponibili**. È necessario selezionare una variabile non stringa.

3. Selezionare una o più variabili di fattore categoriale per il modello dall'elenco **Variabili disponibili**.

4. Selezionare una o più variabili di scala non stringa, covariate dall'elenco **Variabili disponibili**.

**Nota:** Entrambi gli elenchi **Factor (s)** e **Covariate (s)** non possono essere vuoti. È necessario selezionare almeno una variabile **Factor (s)** o **Covariate (s)**.

5. Opzionalmente, selezionare una variabile singola, non stringa, da servire come peso di regressione dall'elenco **Variabili disponibili**.

6. Selezionare l'**analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore**: Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.

- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 8. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

13

14

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

Facoltativamente, è possibile:

- Clicca su **Criteri** per specificare la percentuale di intervallo credibile e le impostazioni del metodo numerico.
- Clicca su **Priori** per definire le impostazioni di distribuzione precedenti e coniugate.
- Clicca su **Fattore di Bayes** per specificare le impostazioni dei fattori di Bayes.
- Clicca su **Salva** per identificare quali elementi salvare e salvare le informazioni del modello in un file XML.
- Clicca su **Presa** per specificare i regressori per la previsione bayesiana.
- Fare clic su **Pagine** per tracciare le distribuzioni posteriori dei parametri di regressione, la varianza dei termini di errore e i valori previsti.
- Fare clic su **F - test** per confrontare modelli statistici al fine di identificare il modello che meglio si adatta alla popolazione da cui è stato campato.

## Modelli di regressione lineari bayesiani: criteri

È possibile specificare i seguenti criteri di analisi per i modelli di regressione lineare Bayesiana.

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

<sup>13</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>14</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

#### **Metodo numerico**

Specificare il metodo numerico utilizzato per stimare l'integrale.

#### **Tolleranza**

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001.

#### **Numero massimo di iterazioni**

Specificare il numero massimo di iterazioni di metodo. Il valore deve essere un intero positivo. L'impostazione predefinita è 2000.

## **Modelli di regressione lineari bayesiani: distribuzioni a priori**

È possibile specificare le seguenti impostazioni di distribuzione preventiva per i parametri di regressione e la varianza degli errori. Le seguenti opzioni sono disponibili solo quando l'opzione **Caratterizzate Distribuzione Posteriore** è selezionata per **Analisi Bayesiana**.

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

#### **Riferimento**

Quando selezionata, l'analisi di riferimento produce inferenza Bayesiana oggettiva. Le istruzioni inferenziali dipendono solo dal modello assunto e dai dati disponibili, e la distribuzione preventiva che viene utilizzata per effettuare un'inferenza è la meno informativa. Questa è l'impostazione predefinita.

#### **Coniuga**

Fornisce opzioni per definire le distribuzioni precedenti coniugate. I priori coniugati assumono la distribuzione congiunta Normale - Inverso - Gamma. Anche se i priori coniugati non sono richiesti quando si eseguono gli aggiornamenti bayesiani, questi aiutano i processi di calcolo.

**Nota:** Al fine di specificare i priori coniugati per un modello di regressione lineare, impostare la tua media di parametri di regressione nella tabella **Priori su varianza di errori**. È inoltre possibile scegliere di utilizzare le impostazioni **Varianza di matrice di covarianza** per specificare la varianza precedente - covarianza.

#### **Priori sulla varianza degli errori**

##### **Parametri forma**

Specificare il parametro shape  $a_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0.

##### **Parametro di scala**

Specificare il parametro di scala  $b_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0. Più grande è il parametro di scala, più si diffonde la distribuzione.

La tabella riporta la media dei parametri di regressione (incluso l'intercettazione), che specifica il vettore medio  $\theta_0$  per i parametri di regressione definiti. Il numero di valori deve soddisfare il numero di parametri di regressione, compreso il termine di intercettazione.

Il primo nome variabile è sempre INTERCEPT. Dalla seconda riga, la colonna **Variabili** viene automaticamente popolata con le variabili specificate da Fattore (s) e Covariate (s). La colonna **Mean** non include alcun valore predefinito.

Fare clic su **Reset** per cancellare i valori.

#### **Varianza della matrice di covarianza: $\sigma^2x$**

Specificare  $V_0$  i valori nel triangolo inferiore della matrice di varianza - covarianza per la probabilità a priori normale multivariata. Si noti che  $V_0$  deve essere definito semi - positivo. L'ultimo valore di

ogni riga deve essere positivo. La riga successiva dovrebbe avere un valore in più rispetto alla riga precedente. Non sono specificati valori per le categorie di riferimento (se presenti).

Fare clic su **Reset** per cancellare i valori.

#### **Utilizza matrice identità**

Quando viene selezionata, viene utilizzata la matrice di identità scalata. Non è possibile specificare  $V_0$  valori nel triangolo inferiore della matrice di varianza - covarianza per la probabilità a priori normale multivariata.

## **Modelli di regressione lineari bayesiani: fattore di Bayes**

È possibile specificare il modello di progettazione per l'analisi, incluso l'approccio utilizzato per stimare il fattore di Bayes per i modelli di regressione lineare Bayesiana. Le seguenti opzioni sono disponibili solo quando viene selezionata l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** Bayesiana.

#### **Modello nullo**

Quando selezionate, i fattori di Bayes stimati si basano sul modello nullo. Questa è l'impostazione predefinita.

#### **Modello completo**

Quando selezionate, i fattori di Bayes stimati si basano sul modello completo e si possono selezionare variabili da utilizzare e ulteriori fattori e covariate.

#### **Variabili**

Elenca tutte le variabili disponibili per il modello completo.

#### **Fattori aggiuntivi**

Selezionare le variabili dall'elenco **Variabili** da utilizzare come fattori aggiuntivi.

#### **Covariate aggiuntive**

Selezionare le variabili dall'elenco **Variabili** da utilizzare come covariate aggiuntive.

#### **Calcolo**

Specificare l'approccio per stimare i fattori di Bayes. Il metodo JZS è l'impostazione predefinita.

#### **Metodo JZS**

Quando selezionato, richiama l'approccio di Zellner-Siow. Questa è l'impostazione predefinita.

#### **Metodo di Zellner**

Quando selezionato, richiama l'approccio di Zellner e si richiede di specificare un singolo valore precedente  $g > 0$  (non esiste alcun valore predefinito).

#### **Metodo iper-probabilità a priori**

Quando questa opzione è selezionata, richiama l'approccio hyper - $g$  e viene richiesto di specificare un parametro di forma  $\alpha_0$  per la distribuzione Inverse - Gamma. È necessario specificare un valore singolo  $> 0$  (il valore predefinito è 3).

#### **Metodo di Rouder**

Quando questa opzione è selezionata, richiama l'approccio del Rouder e viene richiesto di specificare un parametro di scala  $b_0$  per la distribuzione Inverse - Gamma. È necessario specificare un valore singolo  $> 0$  (il valore predefinito è 1).

## **Modelli di regressione lineari bayesiani: Salva**

Questa finestra di dialogo consente di specificare quali statistiche vengono segnate per la distribuzione di previsione Bayesiana ed esportare il modello risultanti in un file XML.

#### **Statistiche predittive a posteriori**

È possibile segnare le seguenti statistiche che derivano dalle previsioni bayesiane.

#### **Medie**

Media della distribuzione predittiva posteriore.

#### **Varianze**

Varianza della distribuzione predittiva posteriore.

**Modalità**

Modalità della distribuzione predittiva posteriore.

**Limite inferiore intervallo credibile**

Limite inferiore dell'intervallo credibile della distribuzione predittiva posteriore.

**Limite superiore intervallo credibile**

Limite superiore dell'intervallo credibile della distribuzione predittiva posteriore.

**Nota:** È possibile assegnare nomi variabili corrispondenti per ogni statistica.

**Esporta informazioni modello in file XML**

Inserire un nome file XML e una posizione per esportare la matrice di varianza dei parametri segnati.

**Modelli di regressione lineari bayesiani: Prevedi**

È possibile specificare i regressori per generare distribuzioni predittive.

**Regressori per previsione bayesiana**

La tabella elenca tutti i regressori disponibili. La colonna **Regressori** è automaticamente popolata da alcune variabili Fattore e Covariate. Specificare i vettori osservati con i valori per i regressori. Ogni regressore può essere assegnato un valore o una stringa e sono autorizzati a prevedere un solo caso. Per i fattori sono consentiti sia valori che stringhe.

Tutti o nessuno dei valori di regressore deve essere specificato al fine di eseguire la previsione (cliccando su **Esegui analisi**).

Quando una variabile Fattore o Covariata viene rimossa, la riga di regressore corrispondente viene rimossa dalla tabella.

Per le covariate possono essere specificati solo valori numerici. Per i fattori sono ammessi sia valori numerici che stringhe.

**Nota:** Fare clic su **Reset** per cancellare i valori definiti.

**Modelli di regressione lineari bayesiani: Traccia**

È possibile controllare i complotti che sono in uscita.

**Covariate**

Elenca le covariate attualmente definite.

**Traccia covariate**

Selezionare le covariate da tracciare dall'elenco **Covariati** e aggiungerle all'elenco **Plot covariati**.

**Fattori**

Elenca i fattori attualmente definiti.

**Traccia fattori**

Selezionare i fattori da tracciare dalla lista **Fattori** e aggiungerli all'elenco **Plot fattori**.

**Numero massimo di categorie da tracciare**

Selezionare il numero massimo di categorie da tracciare (singolo, intero positivo). L'impostazione si applica a tutti i fattori. Per impostazione predefinita, i primi 2 livelli vengono tracciati per ogni fattore.

**Includi grafici di****Termine intercetta**

Quando viene selezionato, viene tracciato il termine di intercettazione. L'impostazione viene deselezionata per impostazione predefinita.

**Termini varianza di errore**

Quando viene selezionato, viene tracciata la varianza degli errori. L'impostazione viene deselezionata per impostazione predefinita.

### Distribuzione prevista bayesiana

Quando viene selezionato, viene tracciata la distribuzione predittiva. L'impostazione viene deselezionata per impostazione predefinita. L'impostazione può essere selezionata solo quando vengono selezionati valori di regressori validi.

## Modelli di regressione lineari bayesiani: Test F

È possibile creare una o più prove parziali. Una F - test è qualsiasi test statistico in cui la statistica di prova abbia una distribuzione F sotto l'ipotesi nulla. Le prove f sono comunemente utilizzate quando si confronta modelli statistici che sono stati montati su un dataset, al fine di identificare il modello che meglio si adatta alla popolazione da cui sono stati campionati i dati.

### Variabili disponibili

Elenca le variabili fattore e covariate che vengono selezionate dalla finestra di dialogo principale Bayesian Linear Regression. Quando le variabili di fattore e covariate vengono aggiunte o eliminate dalla finestra di dialogo principale, l'elenco viene aggiornato di conseguenza.

### Variabile(i) di test

Selezionare le variabili fattore / covariate da verificare dall'elenco **Variabili disponibili** e aggiungerle alla lista **Testing variable (s)**.

**Nota:** L'opzione **Include intercept term** deve essere selezionata quando non vengono selezionati fattori di prova o covariate.

### Valori e variabili di test

Specificare i valori da verificare. Il numero di valori deve corrispondere al numero di parametri nel modello originale. Quando vengono specificati i valori, il primo valore deve essere specificato per il termine di intercettazione (ipotizza che tutti i valori siano 0 quando non esplicitamente definiti).

### Includi termine intercetta

Quando selezionate, i termini di intercettazione sono inclusi nel test. Per impostazione predefinita, l'impostazione non è selezionata.

Quando abilitato, utilizzare il campo **Testing value** per specificare un valore.

### Etichetta test (facoltativo)

È possibile facoltativamente specificare un'etichetta per ogni test. È possibile specificare un valore di stringa con una lunghezza massima di 255 bytes. È consentita una sola etichetta per ogni F - test.

## ANOVA a una via bayesiana

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

La procedura ANOVA a una via produce un'analisi della varianza a una via per una variabile dipendente quantitativa in base a una singola variabile fattore (indipendente). L'analisi della varianza consente di verificare l'ipotesi di uguaglianza di più medie. SPSS Statistiche supporta i fattori di Bayes, i priori coniugati e i priori non informativi.

1. Dai menu, scegliere:

**Analizzare > Statistiche Bayesiane > One-way ANOVA**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile singola, numerica **Dependent** dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile.
3. Selezionare una singola variabile **Fattore** per il modello dalla lista **Variabili disponibili**. È necessario selezionare almeno una variabile **Fattore**.
4. Selezionare una variabile singola, non stringa, da servire come regressione **Peso** dall'elenco **Variabili disponibili**. Il campo variabile **Peso** può essere vuoto.
5. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 9. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

15

16

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

Facoltativamente, è possibile:

- Clicca su **Criteri** per specificare la percentuale di intervallo credibile e le impostazioni del metodo numerico.
- Clicca su **Priori** per definire le impostazioni di distribuzione precedenti e coniugate.
- Clicca su **Fattore di Bayes** per specificare le impostazioni dei fattori di Bayes.
- Fare clic su **Pagine** per controllare i complotti che sono in uscita.

## **ANOVA a una via bayesiana: criteri**

È possibile specificare i seguenti criteri di analisi per i modelli Bayesiano ANOVA.

### **Percentuale % intervallo credibile**

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

<sup>15</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>16</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

### Metodo numerico

Specificare il metodo numerico utilizzato per stimare l'integrale.

### Tolleranza

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001.

### Numero massimo di iterazioni

Specificare il numero massimo di iterazioni di metodo. Il valore deve essere un intero positivo. L'impostazione predefinita è 2000.

## ANOVA a una via bayesiana: probabilità a priori

È possibile specificare le seguenti impostazioni di distribuzione preventiva per i parametri di regressione e la varianza degli errori. Le seguenti opzioni sono disponibili solo quando l'opzione **Caratterizzate Distribuzione Posteriore** è selezionata per **Analisi Bayesiana**.

**Nota:** Molti ricercatori applicati possono mettere in discussione la necessità di specificare un precedente. I priori di riferimento minimizzano la preoccupazione in cui il preventivo è generalmente travolto man mano che i dati aumentano. Quando vengono specificate informazioni preliminari informative, i metodi Bayesiani possono utilizzare efficacemente le informazioni. Il requisito di specificare un preventivo non deve essere considerato un deterrente per utilizzare l'analisi bayesiana.

### Riferimento

Quando selezionata, l'analisi di riferimento produce inferenza Bayesiana oggettiva. Le istruzioni inferenziali dipendono solo dal modello assunto e dai dati disponibili, e la distribuzione preventiva che viene utilizzata per effettuare un'inferenza è la meno informativa. Questa è l'impostazione predefinita.

### Coniuga

Fornisce opzioni per definire le distribuzioni precedenti coniugate. I priori coniugati assumono la distribuzione congiunta Normale - Inverso - Gamma. Anche se i priori coniugati non sono richiesti quando si eseguono gli aggiornamenti bayesiani, questi aiutano i processi di calcolo.

### Priori sulla varianza degli errori

#### Parametri forma

Specificare il parametro shape  $a_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0.

#### Parametro di scala

Specificare il parametro di scala  $b_0$  per la distribuzione Inversa - Gamma. È necessario inserire un valore unico superiore a 0. Più grande è il parametro di scala, più si diffonde la distribuzione.

### Probabilità a priori su parametri di regressione

Specificare il vettore medio  $\beta_0$  per il mezzo di gruppo. Il numero di valori deve soddisfare il numero di parametri di regressione, compreso il termine di intercettazione.

La colonna **Variabili** viene popolata automaticamente con i livelli di Fattore. La colonna **Mean** non include alcun valore predefinito.

Fare clic su **Reset** per cancellare i valori.

### Varianza della matrice di covarianza: $\sigma^2x$

Specificare  $V_0$  i valori nel triangolo inferiore della matrice di varianza - covarianza per la probabilità a priori normale multivariata. Si noti che  $V_0$  deve essere definito semi - positivo. Deve essere specificato solo il triangolo inferiore della tabella.

Le righe e le colonne vengono automaticamente popolate con i livelli di Fattore. Tutti i valori diagonali sono 1; tutti i valori fuori diagonale sono 0.

Fare clic su **Reset** per cancellare i valori.

### Utilizza matrice identità

Quando viene selezionata, viene utilizzata la matrice di identità. Non è possibile specificare  $V_0$  valori nel triangolo inferiore della matrice di varianza - covarianza per la probabilità a priori normale multivariata.

## ANOVA a una via bayesiana: fattore di Bayes

È possibile specificare l'approccio che viene utilizzato per stimare il fattore di Bayes per i modelli ANOVA di tipo Bayesiano. Le seguenti opzioni sono disponibili solo quando viene selezionata l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** Bayesiana.

### Calcolo

Specificare l'approccio per stimare i fattori di Bayes. Il metodo JZS è l'impostazione predefinita.

#### Metodo JZS

Quando selezionato, richiama l'approccio di Zellner-Siow. Questa è l'impostazione predefinita.

#### Metodo di Zellner

Quando selezionato, richiama l'approccio di Zellner e si richiede di specificare un singolo valore precedente  $g > 0$  (non esiste alcun valore predefinito).

#### Metodo iper-probabilità a priori

Quando questa opzione è selezionata, richiama l'approccio hyper -g e viene richiesto di specificare un parametro di forma  $\alpha_0$  per la distribuzione Inverse - Gamma. È necessario specificare un valore singolo  $> 0$  (il valore predefinito è 3).

#### Metodo di Rouder

Quando questa opzione è selezionata, richiama l'approccio del Rouder e viene richiesto di specificare un parametro di scala  $b_0$  per la distribuzione Inverse - Gamma. È necessario specificare un valore singolo  $> 0$  (il valore predefinito è 1).

## ANOVA a una via bayesiana: grafici

È possibile controllare i complotti che sono in uscita.

### Gruppo di grafici

Specificare i sottogruppi da tracciare. Trama la probabilità, prima e posterior per i mezzi dei gruppi specificati. L'elenco **Gruppi** è un sottoinsieme delle categorie della variabile fattore, quindi il formato dovrebbe essere coerente con il tipo di dati del fattore e i valori effettivi.

### Termini varianza di errore

Quando viene selezionato, viene tracciata la varianza degli errori. L'impostazione viene deselezionata per impostazione predefinita. Questa opzione non è disponibile quando **Stima di Bayes Factor** viene selezionato come Analisi Bayesiana.

## Modelli Loglinear Bayesiani

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

Il design per testare l'indipendenza di due fattori richiede due variabili categoriali per la costruzione di un tavolo di contingenza, e rende l'inferenza bayesiana sull'associazione a colonne di fila. È possibile stimare i fattori di Bayes assumendo diversi modelli, e caratterizzare la distribuzione posteriore desiderata simulando l'intervallo credibile simultaneo per i termini di interazione.

1. Dai menu, scegliere:

#### **Analizzare > Statistiche Bayesiane > Modelli loglineari**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare una variabile di riga singola, non in scala dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile non in scala.

3. Selezionare una variabile di colonna singola, non in scala, dall'elenco **Variabili disponibili**. È necessario selezionare almeno una variabile non in scala.

4. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse,

integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.

- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 10. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

17

18

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

Facoltativamente, è possibile:

- Clicca su **Criteri** per specificare la percentuale di intervallo credibile e le impostazioni del metodo numerico.
- Clicca su **Fattore di Bayes** per specificare le impostazioni dei fattori di Bayes.
- Clicca su **Stampa** specificare come viene visualizzato il contenuto nelle tabelle di output.

## Modelli Loglineari Bayesiani: Criteri

È possibile specificare i seguenti criteri di analisi per i modelli Bayesiani Loglineari.

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

### Metodo numerico

Specificare il metodo numerico utilizzato per stimare l'integrale.

### Imposta seme personalizzato

Quando selezionato, è possibile specificare un valore di seed personalizzato nel campo **Seed** . Specificare un valore di serie di sementi casuale. Il valore deve essere un intero positivo. Per impostazione predefinita viene assegnato un valore seed casuale.

<sup>17</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>18</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

**Nota:** Le seguenti opzioni sono disponibili solo quando l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** è selezionata per **Analisi Bayesiana**.

#### **Tolleranza**

Specificare il valore di tolleranza per i metodi numerici. L'impostazione predefinita è 0.000001.

#### **Numero massimo di iterazioni**

Specificare il numero massimo di iterazioni di metodo. Il valore deve essere un intero positivo. L'impostazione predefinita è 2000.

#### **Campioni simulati a Posterior Distribuzione**

Specificare il numero di campioni che vengono utilizzati per disegnare la distribuzione posteriore desiderata. Il valore predefinito è 10000.

#### **Formato**

Selezionare se le categorie sono visualizzate in ordine **Ascending** o **Descending**. Ascendente è l'impostazione predefinita.

## **Bayesi Loglinear Models: Bayes Factor**

È possibile specificare il modello assunto per i dati osservati (Poisson, Multinomiale o Nonparametric). La distribuzione multinomiale è l'impostazione per impostazione predefinita. Le seguenti opzioni sono disponibili solo quando viene selezionata l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** Bayesiana.

#### **Modello Poisson**

Quando viene selezionato, il modello di Poisson viene assunto per i dati osservati.

#### **Modello multinomiale**

Quando viene selezionato, viene assunto il modello Multinomiale per i dati osservati. Questa è l'impostazione predefinita.

#### **Margini fissi**

Selezionare **Gran Totale**, **Sum riga** o **Sum colonna** per specificare i totali marginali fissi per la tabella di contingenza. **Gran Totale** è l'impostazione predefinita.

#### **Distribuzione probabilità a priori**

Specificare il tipo di distribuzione precedente quando si stima il fattore di Bayes.

#### **Coniuga**

Selezionare per specificare una distribuzione preventiva coniugata. Utilizzare la tabella **Parametri di forma** per specificare i parametri di forma  $a_{rs}$  per la distribuzione Gamma. È necessario specificare i parametri di forma quando **Conjugate** viene selezionato come tipo di distribuzione precedente.

Quando viene specificato un solo valore, tutti  $a_{rs}$  s sono considerati uguali a questo valore.  $r_s = 1$  è l'impostazione predefinita. Se è necessario specificare più di un valore, è possibile separare i valori con spazi vuoti.

Il numero di valori numerici che vengono specificati in ogni riga e ciascuna colonna deve corrispondere alla dimensione della tabella di contingenza. Tutti i valori specificati devono essere  $> 0$ .

Fare clic su **Reset** per cancellare i valori.

#### **Parametro di scala**

Specificare il parametro di scala  $b$  per la distribuzione Gamma. È necessario specificare un valore unico  $> 0$ .

#### **Dirichlet composto**

Selezionare per specificare una miscela Dirichlet di distribuzione.

#### **Intrinseco**

Selezionare per specificare una distribuzione preventiva intrinseca.

#### **Modello non parametrico**

Quando viene selezionato, viene assunto il modello Nonparametrico per i dati osservati.

### **Margini fissi**

Selezionare **Sum riga** o **Sum colonna** per specificare i totali marginali fissi per la tabella di contingenza. **Somma riga** è l'impostazione predefinita.

### **Distribuzione probabilità a priori**

Specificare i parametri per i priori Dirichlet. È necessario specificare i parametri **Distribuzione preventiva** quando si seleziona **Modello Nonparametrico**. Quando viene specificato un solo valore, tutti i  $\lambda_s$  sono considerati uguali a questo valore.  $\lambda_s = 1$  è l'impostazione per impostazione predefinita. Se è necessario specificare più di un valore, è possibile separare i valori con spazi vuoti. Tutti i valori specificati devono essere  $> 0$ . Il numero di valori specificati, i valori numerici devono corrispondere alla dimensione della riga o della colonna che non è fissata per la tabella di contingenza.

Fare clic su **Reset** per cancellare i valori.

## **Modelli Loglineari Bayesiani: Stampa**

È possibile specificare come viene visualizzato il contenuto nelle tabelle di output.

### **Progettazione tabella**

#### **Sopprimi tabella**

Quando viene selezionato, la tabella di contingenza non è inclusa nell'output. L'impostazione non è abilitata per impostazione predefinita.

**Nota:** Le seguenti impostazioni non hanno alcun effetto quando l'impostazione **Suppress Table** è abilitata.

### **Statistiche**

Specificare le statistiche per la verifica dell'indipendenza.

#### **Chi-quadrato**

Selezionare per calcolare la statistica di Pearson Chi - Square, gradi di libertà e significato asintotico a 2 lati. Per una tabella di contingenza da 2 da 2, questa impostazione calcola anche la statistica Yates continuity corretta, i gradi di libertà e il significato asintotico a 2 lati. Per un tavolo di contingenza da 2 da 2, con almeno un conteggio delle celle previsto  $< 5$ , questa impostazione calcola anche il significato esatto di 2 schierati e 1.

#### **Rapporto di verosimiglianza**

Selezionare per calcolare la statistica di prova del rapporto di verosimiglianza, gradi di libertà e il significato asintotico a 2 lati.

### **Conteggi**

Specificare quali tipi di conteggi sono inclusi nella tabella di contingenza.

#### **Osservati**

Selezionare per includere i conteggi delle cellule osservate nella tabella di contingenza.

#### **Previsto**

Selezionare per includere i conteggi delle celle previste nella tabella di contingenza.

### **Percentuali**

Specificare quali tipi di percentuale sono inclusi nella tabella di contingenza.

#### **Riga**

Selezionare per includere le percentuali di riga nella tabella di contingenza.

#### **Colonna**

Selezionare per includere le percentuali di colonna nella tabella di contingenza.

#### **Totale**

Selezionare per includere le percentuali totali nella tabella di contingenza.

## **Bayesi One-way Ripetuto Misure ANOVA Models**

Questa funzione richiede SPSS Statistiche Standard Edition o l'opzione Statistiche avanzate.

In modelli di analisi di varianza bayesiana di varianza (ANOVA) si ipotizza che vi sia una singola misurazione per soggetto. Tuttavia, questa assunzione non è sempre vera. Non è raro che un design di studio si prefigga di indagare su risposte medie su più punti temporali o condizioni. Il Bayesi One-way Ripetuto Misure ANOVA procedura misura un fattore dallo stesso soggetto ad ogni singolo punto temporale o condizione, e consente di incrociare i soggetti all'interno dei livelli. Si suppone che ogni soggetto abbia un'osservazione unica per ogni punto o condizione (in quanto tale, l'interazione oggetto - trattamento non è contabilizzata).

1. Dai menu, scegliere:

**Analizzare > Statistiche Bayesiane > Misure ripetute aNOVA**

**Nota:** I campi evidenziati in rosso sono obbligatori. I pulsanti **Incolla** e **OK** vengono abilitati dopo aver inserito valori validi in tutti i campi obbligatori.

2. Selezionare almeno due variabili **Misure Repeate** dalla lista **Variabili disponibili** .

3. Opzionalmente, selezionare una singola variabile da servire come regressione **Peso** dall'elenco **Variabili disponibili** . Il campo variabile **Peso** può essere vuoto.

**Nota:** L'elenco delle variabili disponibili fornisce tutte le variabili tranne le variabili String.

4. Selezionare l' **analisi bayesiana** desiderata:

- **Caratterizza distribuzione posteriore** : Quando viene selezionata, l'inferenza bayesiana viene fatta da una prospettiva che si avvicina alla caratterizzazione delle distribuzioni posteriori. È possibile studiare la distribuzione marginale posteriore del parametro (o dei parametri) di interesse, integrando gli altri parametri di disturbo, e costruire ulteriori intervalli credibili per trarre un'inferenza diretta. Questa è l'impostazione predefinita.
- **Stima del fattore di Bayes** : Se selezionata, la stima dei fattori di Bayes (una delle metodologie più importanti dell'inferenza bayesiana) costituisce un rapporto naturale per confrontare le probabilità marginali tra un'ipotesi nulla e un'ipotesi alternativa.

*Tabella 11. Soglie comunemente utilizzate per definire la significatività dell'evidenza*

Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova	Fattore di Bayes	Categoria di prova
>100	Prove estreme per H0	1-3	Prove aneddotiche per H0	30/1-10/1	Una forte evidenza per H1
30-100	Evidenza molto forte per H0	1	Nessuna prova	1/100-1/30	Evidenza molto forte per H1
10-30	Una forte evidenza per H0	1/3-1	Prove aneddotiche per H1	1/100	Prove estreme per H1
3-10	Evidenza moderata per H0	1/10-1/3	Evidenza moderata per H1		

*H0: Ipotesi nulla*

*H1: Ipotesi alternativa*

19

20

<sup>19</sup> Lee, M.D., e Wagenmakers, E.-J. 2013. *Bayesian Modeling for Cognitive Science: A Practical Course*. Cambridge University Press.

<sup>20</sup> Jeffreys, H. 1961. *Theory of probability*. Oxford University Press.

- **Usa entrambi i metodi** : Quando è selezionato, vengono utilizzati entrambi i metodi di inferenza **Caratterizza distribuzione posteriore** e **Stima fattore di Bayes**.

Facoltativamente, è possibile:

- Clicca su **Criteri** per specificare la percentuale di intervallo credibile e le impostazioni del metodo numerico.
- Clicca su **Fattore di Bayes** per specificare le impostazioni dei fattori di Bayes.
- Fare clic su **Pagine** per tracciare le distribuzioni posteriori del gruppo.

## Misure ripetute ANOVA a una via bayesiana: criteri

È possibile specificare i seguenti criteri di analisi per i modelli Bayesiano One-way Ripetute di ANOVA.

### Percentuale % intervallo credibile

Specificare il livello di significatività per il calcolo degli intervalli credibili. Il livello predefinito è 95%.

### Metodo numerico

Specificare il metodo numerico utilizzato per stimare l'integrale.

### Imposta seme personalizzato

Quando selezionato, è possibile specificare un valore di seed personalizzato nel campo **Seed** . Il valore predefinito è 2.000.000. Il valore deve essere un intero positivo compreso tra 1 e 2.147.483.647. Per impostazione predefinita viene assegnato un valore seed casuale.

### Numero di campioni Monte Carlo

Specificare il numero di punti che vengono campionati per l'approssimazione Monte Carlo. Il valore deve essere un intero positivo compreso tra  $10^3$  e  $10^6$ . Il valore predefinito è 30.000.

## Misure ripetute ANOVA a una via bayesiana: fattore di Bayes

È possibile specificare l'approccio che viene utilizzato per stimare il fattore di Bayes per i modelli Bayesi One-way Ripetute ANOVA. Le seguenti opzioni sono disponibili solo quando viene selezionata l'opzione **Stima Bayes Factor** o **Utilizza Entrambi i metodi** Bayesiana.

### Criterio di informazione bayesiano (o BIC)

Utilizza un'estensione dell'approssimazione BIC alle misure ripetute progetta per stimare i fattori di Bayes. L'impostazione deriva l'effettiva dimensione del campione contabile per le ripetute misure di correlazione, e suggerisce un termine di penalità migliorato quando stimola la BIC per la selezione tra due modelli concorrenti. Questa è l'impostazione predefinita.

### Disegno misto di Roudier

Utilizza le generalizzazioni multivariate della distribuzione Cauchy come priore per le dimensioni di effetto standardizzate, e un precedente non informativo per la varianza.

**Nota:** L'impostazione del peso a frequenza globale e il peso di regressione vengono ignorati quando questa opzione viene selezionata.

## Misure ripetute ANOVA a una via bayesiana: grafici

È possibile controllare i complotti che sono in uscita per illustrare le distribuzioni posteriori del gruppo. La tabella elenca tutte le variabili selezionate come misure ripetute sulla finestra di dialogo Variabili. Selezionare le variabili misure ripetute alla trama

## KRR (Kernel Ridge Regression)

Regressione di Kernel Ridge è una procedura di estensione che utilizza la classe Python **sklearn.kernel\_ridge.KernelRidge** per stimare i modelli di regressione di Kernel Ridge. I modelli di regressione del kernel ridge sono modelli di regressione non parametrica che sono in grado di modellare relazioni lineari e nonlineari tra variabili predittrici ed esiti. I risultati possono essere altamente sensibili alle scelte di iperparametri del modello. Kernel Ridge Regression facilita la scelta dei valori degli

iperparametri attraverso la convalida incrociata k - fold sulle griglie di valori specificate utilizzando la classe `sklearn.model_selection.GridSearchCV`.

### Esempio

#### Statistiche

Additive\_CHI2, CHI2, Cosine, Laplaciano, Linear, Polinomiale, RBF, Sigmoid, Alpha, Gamma, Coef0, Degree, crossvalidation, osservati versus previsti, residui versus previsti, coefficienti di peso duale, coefficienti di peso nello spazio del kernel.

## Considerazioni sui dati

### Dati

- È possibile specificare qualsiasi o tutte le otto diverse funzioni del kernel.
- La funzione del kernel selezionato determina quali hyperparametri sono attivi.
- Gli iperparametri includono alpha per la regolarizzazione ridge che sono comuni a tutti i kernel plus altrettanti altri tre iperparametri per ogni specifica funzione del kernel.
- Quando vengono specificati più sottocomandi del kernel, o viene specificato più di un valore per qualsiasi parametro, viene eseguita una ricerca in griglia con validazione incrociata per valutare i modelli e viene selezionato il modello di adattamento migliore che si basa su dati detenuti.
- L'estensione accetta le variabili split dalla procedura File di Split e i pesi utilizzando la procedura di Weight Cases.
- Quando i pesi sono inclusi, vengono utilizzati nella creazione di valori montati in tutte le analisi. A causa delle limitazioni nel metodo del punteggio nella classe `sklearn.model_selection.GridSearchCV`, le valutazioni di convalida incrociata utilizzate per selezionare il modello non vengono pesate.

### Ipotesi

## Ottenimento di una regressione del kernel Ridge

1. Dai menu, scegliere:

**Analizzare > Regressione > Kernel Ridge ...**

2. Selezionare una variabile **Dependent**.

3. Selezionare una o più variabili **Independent (s)**.

4. L'impostazione predefinita **Single model** viene utilizzata quando viene specificato un solo valore per ogni parametro di funzione del kernel. Quando viene selezionata l'impostazione **Single model**, non è possibile specificare ulteriori funzioni e pesi **Kernel (s)** aggiuntivi in tutta l'analisi, la valutazione e la valutazione dei risultati. È inoltre possibile utilizzare i controlli freccia su e verso il basso per riorganizzare le funzioni del kernel.

Facoltativamente, selezionare **Selezione modello** dall'elenco **Modalità**.

Quando **Modello selezione** viene selezionato dall'elenco **Modalità**, è possibile aggiungere più funzioni del kernel alla lista **Kernel (s)**.

a. Clicca sul controllo di aggiunta (+) per includere ulteriori funzioni del kernel.

b. Fare clic sulla cella vuota nella colonna **Kernel** per selezionare una funzione del kernel.

c. Fare doppio clic su qualsiasi cella della riga di funzione del kernel per specificare i valori dei parametri di funzione del kernel per la colonna corrispondente (**Alpha, Gamma, Coef0, Grado**).

Per ulteriori informazioni, consultare "[Parametri kernel](#)" a pagina 138. I parametri di ottimizzazione delle funzioni del kernel predefiniti sono elencati di seguito.

#### **Additive\_CHI2**

ALPHA=1 GAMMA=1

#### **CHI2**

ALPHA=1 GAMMA=1

**Coseno**

ALPHA=1

**Laplaciano**

ALPHA=1 GAMMA=1/p

**Lineare**

La funzione del kernel predefinito. ALPHA=1

**Polinomiale**

ALPHA=1 GAMMA=1/p COEF0=1 DEGREE=3

**RBF**

ALPHA=1 GAMMA=1/p

**Sigmoide**

ALPHA=1 GAMMA=1/p COEF0=1

**Nota:** Quando viene specificato più di un valore per qualsiasi parametro di funzione del kernel, viene eseguita una ricerca in griglia con validazione incrociata per valutare i modelli e viene selezionato il modello di adattamento migliore che si basa su dati detenuti.

5. Opzionalmente, clicca su **Opzioni** per specificare il numero di pieghe di crossvalidation, le opzioni di visualizzazione, le impostazioni della trama e gli articoli da salvare. Per ulteriori informazioni, consultare [“Regressione Ridge kernel: opzioni”](#) a pagina 138.
6. Fare clic su **OK**.

## Parametri kernel

La Finestra Di Dialogo **Kernel Parametri** fornisce le opzioni per specificare i valori dei parametri di funzione del singolo kernel e per specificare che la selezione del modello viene eseguita utilizzando una ricerca della griglia sulle combinazioni di kernel e valori di parametri della griglia specificati.

**Specifica singoli parametri**

Abilitare l'impostazione per specificare i valori per il parametro funzione del kernel selezionato.

- Inserire un valore e fare clic su **Aggiungi** per includere il valore nel parametro della funzione del kernel.
- Selezionare un valore di parametro e fare clic su **Modifica** per aggiornare il valore.
- Selezionare un valore di parametro e fare clic su **Rimuovi** per eliminare il valore.

**Specifica parametri griglia**

Abilitare l'impostazione per specificare che la selezione del modello viene eseguita utilizzando una ricerca di griglia sulle combinazioni di kernel e valori di parametri della griglia specificati.

## Regressione Ridge kernel: opzioni

La finestra di dialogo **Plotti** fornisce opzioni per specificare il numero di pieghe di crossvalidation, le opzioni di visualizzazione, le impostazioni della trama e gli articoli da salvare.

**Numero di occorrenze con convalida incrociata**

Il numero di spaccature o pieghe in crossvalidation con ricerca griglia per la selezione del modello. Inserire un valore intero superiore a 1. Il valore predefinito è 5. L'impostazione è disponibile solo quando **Modello selezione** viene scelta come **Modalità** sulla finestra di dialogo principale **Kernel Ridge Regression**.

**Visualizza**

Fornisce opzioni per specificare quale output visualizzare quando è in vigore il crossvalidation.

**Cordiali saluti**

L'impostazione predefinita visualizza solo i risultati di base per il modello migliore scelto.

**Confronta**

Visualizza i risultati di base per tutti i modelli valutati.

**Confronta modelli e occorrenze**

Visualizza i risultati completi per ogni suddivisione o piega per ogni modello valutato.

**Grafico**

Fornisce opzioni per specificare i tracciati di valori osservati o residui rispetto ai valori previsti.

**Confronto tra osservato e previsto**

Visualizza un tracciato di dispersione di valori osservati versus previsti per il modello specificato o migliore.

**Confronto tra residuo e previsto**

Visualizza una dispersione di residui rispetto ai valori previsti per il modello specificato o migliore.

**Salva**

La tabella fornisce le opzioni per specificare le variabili da salvare al dataset attivo.

**Valori previsti**

Salva i valori previsti dal modello specificato o migliore al dataset attivo. Può essere incluso un nome variabile opzionale.

**Residui**

Salva i residui dalle previsioni del modello specificate o migliori al dataset attivo. Può essere incluso un nome variabile opzionale.

**Coefficienti duali**

Salva i coefficienti di peso dello spazio dual o del kernel dal modello specificato al dataset attivo. Può essere incluso un nome variabile opzionale. L'impostazione non è disponibile quando **Selezione modello** viene scelta come **Modalità** sulla finestra di dialogo principale **Kernel Ridge Regression** .



## Informazioni particolari

---

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

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

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

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

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

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

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

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

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

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

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

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119*

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

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

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

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

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

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

#### LICENZA DI COPYRIGHT:

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

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

© Copyright IBM Corp. 2021. Le porzioni di questo codice derivano da IBM Corp. Programmi di esempio.

© Copyright IBM Corp. 1989 - 2021. Tutti i diritti riservati.

## Marchi

---

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

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

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

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

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

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

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



# Indice analitico

## A

- Analisi della covarianza
  - in GLM Multivariato [1](#)
- analisi della varianza
  - in Varianza Componenti [19](#)
  - nei modelli misti lineari generalizzati [50](#)
- analisi di sopravvivenza
  - a Kaplan - Meier [92](#)
  - in Cox Regressione [94](#)
  - in Tabelle di vita [72](#)
  - nella regressione Kernel Ridge [136](#)
  - Regressione di Cox dipendente dal tempo [98](#)
- analisi loglineare
  - Analisi loglineare generale [66](#)
  - Analisi loglineare logit [69](#)
  - nei modelli misti lineari generalizzati [50](#)
- Analisi loglineare generale
  - contrasti [66](#)
  - covariate di celle [66](#)
  - criteri [68](#)
  - Distribuzione dei conteggi delle celle [66](#)
  - fattori [66](#)
  - funzioni aggiuntive del comando [69](#)
  - grafici [68](#)
  - impostazione di modelli [68](#)
  - intervalli di confidenza [68](#)
  - opzioni di visualizzazione [68](#)
  - residui [69](#)
  - salvataggio di valori attesi [69](#)
  - salvataggio di variabili [69](#)
  - strutture cellulari [66](#)
- Analisi loglineare logit
  - contrasti [69](#)
  - covariate di celle [69](#)
  - criteri [71](#)
  - Distribuzione dei conteggi delle celle [69](#)
  - fattori [69](#)
  - grafici [71](#)
  - impostazione di modelli [70](#)
  - intervalli di confidenza [71](#)
  - opzioni di visualizzazione [71](#)
  - residui [72](#)
  - salvataggio di variabili [72](#)
  - strutture cellulari [69](#)
  - valori attesi [72](#)
- analisi probit
  - modelli misti lineari generalizzati [50](#)
- ANOVA
  - in GLM Multivariato [1](#)
  - in GLM Repeated Misure [9](#)

## B

- Bonferroni
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)

- bontà di adattamento
  - in Equazioni di stima Generalizzate [46](#)
  - in Modelli lineari generalizzati [35](#)

## C

- C di Dunnett
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- calcolo del punteggio
  - nei modelli misti lineari [26](#)
- Calcolo del punteggio di Fisher
  - nei modelli misti lineari [26](#)
- casi censurati
  - a Kaplan - Meier [92](#)
  - in Cox Regressione [94](#)
  - in Tabelle di vita [72](#)
- categoria di riferimento
  - in Equazioni di stima Generalizzate [43](#), [44](#)
  - in Modelli lineari generalizzati [32](#)
- classe di generazione
  - in Selezione del modello Analisi lineare [65](#)
- Componenti della varianza
  - funzioni aggiuntive del comando [21](#)
  - modello [19](#)
  - opzioni [19](#)
  - Salvataggio dei risultati [21](#)
- contrasti
  - in Analisi Loglineare Generale [66](#)
  - in Analisi Loglineare Logit [69](#)
  - in Cox Regressione [95](#)
- convergenza di verosimiglianza
  - in Equazioni di stima Generalizzate [45](#)
  - in Modelli lineari generalizzati [33](#)
  - nei modelli misti lineari [26](#)
- Convergenza hessiana
  - in Equazioni di stima Generalizzate [45](#)
  - in Modelli lineari generalizzati [33](#)
- convergenza parametri
  - in Equazioni di stima Generalizzate [45](#)
  - in Modelli lineari generalizzati [33](#)
  - nei modelli misti lineari [26](#)
- costruzione di termini [3](#), [12](#), [19](#), [66](#), [68](#), [71](#)
- covariate
  - in Cox Regressione [95](#)
- covariate di stringa
  - in Cox Regressione [95](#)
- covariate time - dipendenti
  - in Cox Regressione [98](#)
- cronologia delle iterazioni
  - in Equazioni di stima Generalizzate [46](#)
  - in Modelli lineari generalizzati [35](#)
  - nei modelli misti lineari [26](#)

## D

- decomposizione gerarchica

- decomposizione gerarchica (*Continua*)
  - in Varianza Componenti [20](#)
- differenza meno significativa (LSD)
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- differenza meno significativa (LSD) di Fisher
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- differenza significativa di Tukey
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- dimezzamenti
  - in Equazioni di stima Generalizzate [45](#)
  - in Modelli lineari generalizzati [33](#)
  - nei modelli misti lineari [26](#)
- Distanza di Cook
  - in GLM [6](#)
  - in GLM Repeated Misure [16](#)
  - in Modelli lineari generalizzati [37](#)
- distribuzione binomiale
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione binomiale negativa
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione di Poisson
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- Distribuzione di Tweedie
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione gamma
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione gaussiana inversa
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione multinomiale
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- distribuzione normale
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)

## E

- effetti casuali
  - nei modelli misti lineari [25](#), [28](#)
- effetti fissi
  - nei modelli misti lineari [24](#)
- eliminazione all'indietro
  - in Selezione del modello Analisi lineare [64](#)
- Equazioni di stima generalizzate
  - categoria riferimento per la risposta binaria [43](#)
  - criteri di stima [45](#)
  - Esportazione del modello [49](#)
  - impostazione di modelli [44](#)
  - medie marginali stimate [47](#)
  - opzioni per fattori categoriali [44](#)
  - predittori [43](#)
  - risposta [43](#)
  - Salva variabili nel dataset attivo [48](#)
  - Statistiche [46](#)
  - Tipo di modello [41](#)

- Equazioni di stima generalizzate (*Continua*)
  - valori iniziali [46](#)
- errore standard
  - in GLM [6](#)
  - in GLM Repeated Misure [16](#)

## F

- F multiplo di Ryan-Einot-Gabriel-Welsch
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- fattori
  - in GLM Repeated Misure [11](#)
- frequenze
  - in Selezione del modello Analisi lineare [66](#)
- funzione di collegamento
  - modelli misti lineari generalizzati [53](#)
- funzione di collegamento alimentazione
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento binomiale negativa
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento Cauchit cumulativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento di potenza
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento identità
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento log
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento log - log complementare
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento log - log complementare cumulativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento log - log negativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento log - log negativo cumulativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento logit
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento logit cumulativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento probit
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di collegamento probit cumulativo
  - in equazioni stimate generalizzate [41](#)
  - in modelli lineari generalizzati [28](#)
- funzione di sopravvivenzaa
  - in Tabelle di vita [72](#)
- funzione funzionale generale
  - in Equazioni di stima Generalizzate [46](#)

funzione link complemento di registrazione  
in equazioni stimate generalizzate [41](#)  
in modelli lineari generalizzati [28](#)  
Funzioni generali stimabili  
in Modelli lineari generalizzati [35](#)

## G

gamma multipla di Ryan-Einot-Gabriel-Welsch  
in GLM Multivariato [5](#)  
in GLM Repeated Misure [14](#)

### GLM

salvataggio di matrici [6](#)  
salvataggio di variabili [6](#)

### GLM Misure ripetute

Definisci fattori [11](#)  
funzioni aggiuntive del comando [17](#)  
grafici di profilo [14](#)  
modello [12](#)  
salvataggio di variabili [16](#)  
test Post Hoc [14](#)

### GLM multivariato [1](#)

### GLM Multivariato

covariate [1](#)  
fattori [1](#)  
grafici di profilo [4](#)  
test Post Hoc [5](#)  
Variabile dipendente [1](#)

### GLOR

in Analisi Loglineare Generale [66](#)

### grafici

in Analisi Loglineare Generale [68](#)  
in Analisi Loglineare Logit [71](#)

### grafici delle probabilità normale

in Selezione del modello Analisi lineare [66](#)

### grafici di profilo

in GLM Multivariato [4](#)  
in GLM Repeated Misure [14](#)

## H

### Hochberg (GT2)

in GLM Multivariato [5](#)  
in GLM Repeated Misure [14](#)

## I

### informazione del modello

in Equazioni di stima Generalizzate [46](#)  
in Modelli lineari generalizzati [35](#)

### informazioni a livello di fattore

nei modelli misti lineari [27](#)

### interruzione del rapporto di lavoro

in Equazioni di stima Generalizzate [45](#)  
in Modelli lineari generalizzati [33](#)

### intervalli di confidenza

in Analisi Loglineare Generale [68](#)  
in Analisi Loglineare Logit [71](#)  
nei modelli misti lineari [27](#)

### iterazioni

in Equazioni di stima Generalizzate [45](#)  
in Modelli lineari generalizzati [33](#)  
in Selezione del modello Analisi lineare [66](#)

## K

### Kaplan-Meier

confronto dei livelli dei fattori [93](#)  
definizione degli eventi [93](#)  
esempio [92](#)  
funzioni aggiuntive del comando [94](#)  
grafici [94](#)  
quartili [94](#)  
salvataggio di nuove variabili [94](#)  
Statistiche [92](#), [94](#)  
Tabelle sopravvivenza [94](#)  
tempo di sopravvivenza medio e mediano [94](#)  
Tendenza lineare per i livelli dei fattori [93](#)  
variabili stato di sopravvivenza [93](#)

### Kernel Ridge

alfa [136](#)  
coef0 [136](#)  
gamma [136](#)  
laurea [136](#)  
Modello singolo [136](#)  
selezione modello [136](#)

### KRR (Kernel Ridge Regression)

grafici [138](#)  
parametri [138](#)  
parametri della griglia [138](#)  
pieghe incrociate [138](#)  
salva [138](#)  
visualizzazione [138](#)

## M

### matrice dei coefficienti di contrasto

in Equazioni di stima Generalizzate [46](#)  
in Modelli lineari generalizzati [35](#)

### matrice di correlazione

in Equazioni di stima Generalizzate [46](#)  
in Modelli lineari generalizzati [35](#)  
nei modelli misti lineari [27](#)

### matrice di covarianza

in Equazioni di stima Generalizzate [45](#), [46](#)  
in GLM [6](#)

in Modelli lineari generalizzati [33](#), [35](#)  
nei modelli misti lineari [27](#)

### matrice di covarianza ad effetto casuale

nei modelli misti lineari [27](#)

### Matrice di covarianza dei residui

nei modelli misti lineari [27](#)

### Matrice di covarianza parametri

nei modelli misti lineari [27](#)

### matrice L

in Equazioni di stima Generalizzate [46](#)  
in Modelli lineari generalizzati [35](#)

### medie marginali stimate

in Equazioni di stima Generalizzate [47](#)  
in Modelli lineari generalizzati [36](#)  
nei modelli misti lineari [27](#)

### Metodo Newton - Raphson

in Analisi Loglineare Generale [66](#)  
in Analisi Loglineare Logit [69](#)

### MINQUE

in Varianza Componenti [19](#)

Modelli di fragilità condivisa parametrici  
analisi [81](#)

Modelli di fragilità condivisa parametrici (*Continua*)

- criteri [82](#)
- esportazione [87](#)
- modello [83](#)
- predire [85](#)
- stampare [85](#)
- stima [83](#)
- tracciare [86](#)

Modelli di fragilità parametrica

- variabili di stato di sopravvivenza [87](#)

modelli fattoriali completi

- in GLM Repeated Misure [12](#)
- in Varianza Componenti [19](#)

modelli gerarchici

- modelli misti lineari generalizzati [50](#)

Modelli lineari generalizzati

- categoria riferimento per la risposta binaria [32](#)
- criteri di stima [33](#)
- distribuzione [28](#)
- Esportazione del modello [38](#)
- funzione di collegamento [28](#)
- impostazione di modelli [33](#)
- medie marginali stimate [36](#)
- opzioni per fattori categoriali [32](#)
- predittori [32](#)
- risposta [31](#)
- Salva variabili nel dataset attivo [37](#)
- Statistiche [35](#)
- tipi di modello [28](#)
- valori iniziali [34](#)

modelli logit multinazionali [69](#)

modelli loglineari gerarchici [64](#)

modelli longitudinali

- modelli misti lineari generalizzati [50](#)

modelli misti

- lineare [21](#)
- modelli misti lineari generalizzati [50](#)

Modelli misti lineari

- costruzione di termini [24](#)
- criteri di stima [26](#)
- effetti casuali [25, 28](#)
- effetti fissi [24](#)
- funzioni aggiuntive del comando [28](#)
- medie marginali stimate [27](#)
- modello [27](#)
- Salva variabili [28](#)
- struttura della covarianza [102](#)
- termini di interazione [24](#)

modelli misti lineari generalizzati

- blocco effetti casuali [56](#)
- coefficienti fissi [62](#)
- covarianze effetti casuali [63](#)
- distribuzione obiettivo [53](#)
- effetti casuali [56](#)
- effetti fissi [55, 62](#)
- Esportazione del modello [60](#)
- funzione di collegamento [53](#)
- medie marginali stimate [59](#)
- medie stimate [63, 64](#)
- parametri di covarianza [63](#)
- peso dell'analisi [57](#)
- previsioni e osservazioni [61](#)
- riepilogo del modello [61](#)
- Salva campi [60](#)

modelli misti lineari generalizzati (*Continua*)

- scostamento [57](#)
- struttura dati [61](#)
- tabella di classificazione [61](#)
- termini personalizzati [55](#)
- vista modello [60](#)

modelli multilivello

- modelli misti lineari generalizzati [50](#)

modelli personalizzati

- in GLM Repeated Misure [12](#)
- in Selezione del modello Analisi lineare [65](#)
- in Varianza Componenti [19](#)

modelli saturi

- in Selezione del modello Analisi lineare [65](#)

Modelli Temporal Accelerati Parametrici accelerati

- analisi [74](#)
- criteri [75](#)
- esportazione [80](#)
- modello [76](#)
- predire [78](#)
- stampare [78](#)
- stima [76](#)
- tracciare [79](#)

modello di rischio proporzionale

- in Cox Regressione [94](#)

modello lineare generale

- modelli misti lineari generalizzati [50](#)

modello lineare generalizzato

- nei modelli misti lineari generalizzati [50](#)

multivariato ANOVA [1](#)

## N

Newman-Keuls

- in GLM Multivariato [5](#)
- in GLM Repeated Misure [14](#)

## P

parametro di scala

- in Equazioni di stima Generalizzate [45](#)
- in Modelli lineari generalizzati [33](#)

priori a effetto casuale

- in Varianza Componenti [19](#)

## R

R-E-G-W F

- in GLM Multivariato [5](#)
- in GLM Repeated Misure [14](#)

R-E-G-W Q

- in GLM Multivariato [5](#)
- in GLM Repeated Misure [14](#)

rapporto di odds

- in Analisi Loglineare Generale [66](#)

Rapporto odd logaritmico generalizzato

- in Analisi Loglineare Generale [66](#)

Regressione di Cox

- contrastati [95](#)
- covariate [94](#)
- Covariate categoriali [95](#)
- covariate di stringa [95](#)
- Covariate dipendenti dal tempo [98](#)

## Regressione di Cox (*Continua*)

- Definisci evento [97](#)
  - DfBeta(s) [96](#)
  - esempio [94](#)
  - funzione di rischio [96](#)
  - funzione di sopravvivenza [96](#)
  - funzioni aggiuntive del comando [97](#)
  - funzioni di base [97](#)
  - grafici [96](#)
  - ingresso e rimozione stepwise [97](#)
  - iterazioni [97](#)
  - Residui parziali [96](#)
  - salvataggio di nuove variabili [96](#)
  - Statistiche [94](#), [97](#)
  - variabile stato di sopravvivenza [97](#)
- regressione di Poisson
- in Analisi Loglineare Generale [66](#)
  - modelli misti lineari generalizzati [50](#)
- regressione logistica
- modelli misti lineari generalizzati [50](#)
- regressione logistica multinomiale
- modelli misti lineari generalizzati [50](#)
- regressione multivariata [1](#)
- residui
- in Analisi Loglineare Generale [69](#)
  - in Analisi Loglineare Logit [72](#)
  - in Equazioni di stima Generalizzate [48](#)
  - in Modelli lineari generalizzati [37](#)
  - in Selezione del modello Analisi lineare [66](#)
  - nei modelli misti lineari [28](#)
- residui cancellati
- in GLM [6](#)
  - in GLM Repeated Misure [16](#)
- Residui di devianza
- in Modelli lineari generalizzati [37](#)
- residui di Pearson
- in Equazioni di stima Generalizzate [48](#)
  - in Modelli lineari generalizzati [37](#)
- residui di probabilità
- in Modelli lineari generalizzati [37](#)
- residui non standardizzati
- in GLM [6](#)
  - in GLM Repeated Misure [16](#)
- residui standardizzati
- in GLM [6](#)
  - in GLM Repeated Misure [16](#)
- Riepilogo dell'elaborazione dei casi
- in Equazioni di stima Generalizzate [46](#)
  - in Modelli lineari generalizzati [35](#)

## S

- Selezione modello: Analisi loglineare
- definizione intervalli di fattori [65](#)
  - funzioni aggiuntive del comando [66](#)
  - modelli [65](#)
  - opzioni [66](#)
- somma dei quadrati [3](#), [12](#)
- Somme dei quadrati
- in Varianza Componenti [20](#)
  - nei modelli misti lineari [24](#)
- Sopravvivenza AFT
- Finestra di dialogo di sopravvivenza - Variabili categoria [80](#)

## statistica descrittiva

- in Equazioni di stima Generalizzate [46](#)
  - in Modelli lineari generalizzati [35](#)
  - nei modelli misti lineari [27](#)
- statistica di Wald
- in Analisi Loglineare Generale [66](#)
  - in Analisi Loglineare Logit [69](#)
- stima massima della probabilità
- in Varianza Componenti [19](#)
- stima massima della probabilità limitata
- in Varianza Componenti [19](#)
- stime del parametro
- in Analisi Loglineare Generale [66](#)
  - in Analisi Loglineare Logit [69](#)
  - in Equazioni di stima Generalizzate [46](#)
  - in Modelli lineari generalizzati [35](#)
  - in Selezione del modello Analisi lineare [66](#)
  - nei modelli misti lineari [27](#)
- Strutture covarianza
- nei modelli misti lineari [102](#)
- Student-Newman-Keuls
- in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)

## T

- T3 di Dunnett
- in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- tabelle di contingenza
- in Analisi Loglineare Generale [66](#)
- Tabelle vita
- confronto dei livelli dei fattori [74](#)
  - esempio [72](#)
  - funzione di sopravvivenza [72](#)
  - funzioni aggiuntive del comando [74](#)
  - grafici [74](#)
  - Prova di Wilcoxon (Gehan) [74](#)
  - Statistiche [72](#)
  - Tasso di rischio [72](#)
  - variabili di stato di sopravvivenza [80](#)
  - variabili fattore [74](#)
  - visualizzazione della visualizzazione della tabella [74](#)
- Tamhane (T2)
- in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- Tarone - Test Ware
- a Kaplan - Meier [93](#)
- Tasso di rischio
- in Tabelle di vita [72](#)
- tavola di contingenza
- in Selezione del modello Analisi lineare [64](#)
- termini di interazione
- nei modelli misti lineari [24](#)
- termini nidificati
- in Equazioni di stima Generalizzate [44](#)
  - in Modelli lineari generalizzati [33](#)
  - nei modelli misti lineari [24](#)
- test b di Tukey
- in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- test dei parametri di covarianza
- nei modelli misti lineari [27](#)
- Test del moltiplicatore di Lagrange

- Test del moltiplicatore di Lagrange (*Continua*)
  - in Modelli lineari generalizzati [35](#)
- Test di Breslow
  - a Kaplan - Meier [93](#)
- Test di Gehan
  - in Tabelle di vita [74](#)
- test di intervallo multiplo di Duncan
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- test di Scheffé
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- Test di Wilcoxon
  - in Tabelle di vita [74](#)
- Test per confronti pairwise di Gabriel
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- Test per confronti pairwise di Games e Howell
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- test rank rank
  - a Kaplan - Meier [93](#)
- test t di Dunnett
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- test t di Sidak
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- test t di Waller-Duncan
  - in GLM Multivariato [5](#)
  - in GLM Repeated Misure [14](#)
- tolleranza di singolarità
  - nei modelli misti lineari [26](#)

## V

- valori attesi
  - in Analisi Loglineare Generale [69](#)
  - in Analisi Loglineare Logit [72](#)
  - nei modelli misti lineari [28](#)
- valori attesi pesati
  - in GLM [6](#)
  - in GLM Repeated Misure [16](#)
- valori di leva
  - in GLM [6](#)
  - in GLM Repeated Misure [16](#)
  - in Modelli lineari generalizzati [37](#)
- Valori previsti fissi
  - nei modelli misti lineari [28](#)
- variabili dei soggetti
  - nei modelli misti lineari [22](#)
- variabili misure ripetute
  - nei modelli misti lineari [22](#)
- vista modello
  - nei modelli misti lineari generalizzati [60](#)



