

*IBM SPSS Regression 30*

**IBM**

**Nota**

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in [“Informazioni particolari” a pagina 49](#).

**Informazioni sul prodotto**

Questa edizione si applica alla versione 30, release 0, modifica 0 di IBM® SPSS Statistiche e a tutte le release e modifiche successive se non diversamente indicato in nuove edizioni.

© **Copyright International Business Machines Corporation .**

---

# Indice

<b>Capitolo 1. Regressione.....</b>	<b>1</b>
Scelta di una procedura per la regressione logistica binaria.....	1
Regressione logistica.....	2
Regola Regressione Logistica.....	3
Metodi Di Selezione Variabile di Regressione Logistica.....	3
Regressione Logistica Definire Variabili Categoriali.....	3
Regressione Logistica Salva nuove variabili.....	4
Opzioni Di Regressione Logistica.....	5
REGRESSIONE LOGISTICA Comando Funzioni aggiuntive.....	5
Regressione logistica multinomiale.....	6
Regressione logistica multinomiale.....	6
Multinomiale Logistic Regression Reference Categoria.....	7
Statistiche Di Regressione Logistica Multinomiale.....	7
Criteri Di Regressione Logistica Multinomiale.....	8
Opzioni Di Regressione Logistica multinomiale.....	8
Multinomiale Logistic Regression Save.....	9
Funzioni Aggiuntive di comando NOMREG.....	9
Regressione Probit.....	10
Analisi probit Definisci intervallo.....	11
Opzioni di analisi probit.....	11
PROBIT Command Aggiuntivo Features.....	11
Regressione quantile.....	12
Regressione quantile: criteri.....	13
Regressione quantile: modello.....	14
Regressione quantile: visualizzazione.....	15
Regressione quantile: salva.....	16
Regressione quantile: esporta.....	17
Regressione non lineare.....	17
Logica Condizionale (Regressione Non Lineare).....	18
Parametri Di Regressione Non Lineare.....	19
Modelli Comuni Di Regressione Non Lineare.....	19
Funzione Perdita di regressione non lineare.....	20
Vincoli Parametri di regressione non lineare.....	20
Regressione Non Lineare Save New Variabili.....	21
Opzioni Di Regressione Non Lineare.....	21
Interpretazione Risultati Regression Non Lineari.....	21
Funzioni Aggiuntive di comando NLR.....	22
Stima del peso.....	22
Opzioni Di Stima Del Peso.....	23
Funzioni Aggiuntive Comando WLS.....	23
Regressione minimi quadrati a due fasi.....	23
Opzioni Di Regressione di Due Stage - Quadrati.....	24
2SLS Funzioni aggiuntive di comando.....	25
Schemi Di Codifica Variabile Categoriale.....	25
Deviazione.....	25
Semplice.....	25
Helmert.....	26
Differenza.....	26
Polinomiale.....	26
Ripetuto.....	27
Speciale.....	27

Indicatore.....	28
KRR (Kernel Ridge Regression).....	28
Parametri kernel.....	30
Regressione Ridge kernel: opzioni.....	30
Modelli Temporali Accelerati Parametrici accelerati.....	31
Parametric Accelerated Failure Time Models: Criteria.....	32
Parametric Accelerated Failure Time Models: Modello.....	32
Parametric Accelerated Failure Time Models: Preventivo.....	33
Parametric Accelerated Failure Time Models: Stampa.....	34
Parametric Accelerated Failure Time Models: Prevedere.....	35
Parametric Accelerated Failure Time Models: Plot.....	35
Parametric Accelerated Failure Time Models: Export.....	36
Eventi di definizione AFT di sopravvivenza per variabili di stato.....	36
Modelli di tempo di errore accelerati parametrici: Seleziona categoria.....	36
Modelli di fragilità condivisa parametrici.....	37
Modelli di fragilità condivisa parametrica: criteri.....	38
Modelli di fragilità condivisa parametrica: Modello.....	39
Modelli di fragilità condivisa parametrica: Stima.....	40
Modelli di fragilità condivisa parametrica: Stampa.....	41
Modelli di fragilità condivisa parametrici: Prevedi.....	41
Modelli di fragilità parametrici condivisi: grafico.....	42
Modelli di fragilità condivisa parametrici: Esporta.....	43
Modelli di fragilità parametrica condivisa: Definisci eventi.....	43
Modelli di fragilità condivisa parametrici - Esempi.....	44
Modelli di fragilità parametrici condivisi - Un caso di studio per i dati più di recente.....	46
<b>Informazioni particolari.....</b>	<b>49</b>
Marchi.....	50
<b>Indice analitico.....</b>	<b>53</b>

---

# Capitolo 1. Regressione

Le seguenti funzioni di regressione sono incluse in SPSS Statistiche Standard Edition o l'opzione Regression.

---

## Scelta di una procedura per la regressione logistica binaria

---

I modelli di regressione logistica binari possono essere montati utilizzando la procedura di regressione logistica e la procedura di regressione logistica multinomiale. Ogni procedura ha opzioni non disponibili nell'altra. Un'importante distinzione teorica è che la procedura di Regressione Logistica produce tutte le previsioni, i residui, le statistiche di influenza e le prove di bontà di adattamento che utilizzano i dati a livello di singolo caso, indipendentemente dal modo in cui i dati sono inseriti e se il numero di schemi di covariata sia inferiore al numero totale dei casi, mentre la procedura Regressione Logistica Multinomiale aggrega internamente i casi per formare sottopopolazioni con identici modelli di covariata per i predittori, producendo previsioni, residui e test di bontà basati su queste sottopopolazioni. Se tutti i predittori sono categoriali o eventuali predittori continui assumono solo un numero limitato di valori - in modo che ci siano diversi casi ad ogni modello di covariato distinto - l'approccio di sottopopolazione può produrre validi test di bontà e residui informativi, mentre l'approccio di livello individuale di singolo caso non può.

### **Regressione logistica**

Fornisce le seguenti funzioni uniche:

- Hosmer-Lemeshow prova di bontà di adattamento per il modello
- Analisi stepwise
- Contrasti per definire la parametrizzazione del modello
- Punti di taglio alternativi per la classificazione
- Grafici di classificazione
- Modello montato su una serie di casi ad un insieme di casi tenuto
- Salva le previsioni, i residui e le statistiche di influenza

### **Regressione logistica multinomiale**

Fornisce le seguenti funzioni uniche:

- Pearson e deviance chi - prova piazza per bontà di adattamento del modello
- Specifica delle sottopopolazioni per il raggruppamento dei dati per le prove di bontà
- Quotazione dei conteggi, conteggi previsti e residui per sottopopolazioni
- Correzione delle stime di varianza per eccesso di dispersione
- Matrice di covarianza delle stime dei parametri
- Test di combinazioni lineari di parametri
- Specifica esplicita dei modelli nidificati
- Fit 1 - 1 corrispondenza modelli di regressione logistica condizionale utilizzando variabili differenziate

### **Note:**

- Entrambe queste procedure si adattano ad un modello per i dati binari che è un modello lineare generalizzato con funzione di distribuzione binomiale e collegamento logit. Se una funzione di collegamento differente è più appropriata per i tuoi dati, allora dovresti utilizzare la procedura Generalized Linear Models.
- Se si hanno ripetute misurazioni di dati binari o record altrimenti correlati, allora si dovrebbero considerare le procedure Generalizzate Linear Mixed Models o Generalized Estimating Equations.

## Regressione logistica

---

La regressione logistica è utile per le situazioni in cui si desidera prevedere la presenza o l'assenza di una caratteristica o di un esito in base ai valori di un insieme di variabili predittore. È simile a un modello di regressione lineare ma è adatto per i modelli in cui la variabile dipendente è dicotomica. Il coefficiente di regressione logistica può essere utilizzato per stimare i rapporti odd per ciascuna delle variabili indipendenti nel modello. La regressione logistica è applicabile a una gamma più ampia di situazioni di ricerca rispetto all'analisi discriminante.

**Esempio.** Quali caratteristiche di stile di vita sono fattori di rischio per la cardiopatia coronarica (CHD)? Dato un campione di pazienti misurati sullo stato di fumo, la dieta, l'esercizio fisico, l'uso di alcol e lo stato CHD, si potrebbe costruire un modello utilizzando le quattro variabili di stile di vita per prevedere la presenza o l'assenza di CHD in un campione di pazienti. Il modello può quindi essere utilizzato per ricavare stime dei rapporti di probabilità per ogni fattore di rischio, ad esempio, quanto i fumatori più probabili sviluppino CHD rispetto ai non fumatori.

**Statistiche.** Per ogni analisi: casi totali, casi selezionati, casi validi. Per ogni variabile categoriale: codifica dei parametri. Per ogni passo: variabile (s) entrata o rimossa, storia dell'iterazione, -2 logaritmo-verosimiglianza, bontà di adattamento, Hosmer-Lemeshow bontà - di - fit statistica, modello chi - quadrato, miglioramento chi - quadrato, tabella di classificazione, correlazioni tra variabili, gruppi osservati e probabilità probabili, residua chi - quadrato. Per ciascuna variabile nell'equazione: coefficiente ( $B$ ), errore standard di  $B$ , Wald statistica, odds ratio stimato ( $\exp(B)$ ), intervallo di confidenza per l' $\exp(B)$ , logaritmo verosimilare se termine rimosso dal modello. Per ogni variabile non nell'equazione: statistica del punteggio. Per ogni caso: gruppo osservato, probabilità prevista, gruppo previsto, residuo, residuo standardizzato.

**Metodi.** È possibile stimare i modelli utilizzando l'ingresso a blocchi di variabili o uno qualsiasi dei seguenti metodi stepwise: forward condizionale, forward LR, forward Wald, backward condizionale, backward LR o backward Wald.

### considerazioni sui dati di regressione logistica

**Dati.** La variabile dipendente deve essere dicotomica. Le variabili indipendenti possono essere al livello dell'intervallo o categoriali; se categoriali, devono essere codificate da indicatori o dummy (nella procedura è presente un'opzione per ricodificare automaticamente le variabili categoriali).

**Ipotesi.** La regressione logistica non si affida a ipotesi distributive nello stesso senso che l'analisi discriminante fa. Tuttavia, la vostra soluzione potrebbe essere più stabile se i vostri predittori hanno una distribuzione normale multivariata. Inoltre, come per altre forme di regressione, la multicollinearità tra i predittori può portare a stime biased e ad errori standard gonfiati. La procedura è più efficace quando l'appartenenza di gruppo è una variabile veramente categoriale; se l'appartenenza di gruppo si basa su valori di una variabile continua (ad esempio "IQ alto" versus "basso IQ"), si deve considerare l'utilizzo di regressione lineare per sfruttare le informazioni più ricche offerte dalla variabile continua stessa.

**Procedure correlate.** Usa la procedura Scatterplot per visualizzare i tuoi dati per la multicollinearità. Se le ipotesi di normalità multivariata e le matrici di covarianza uguali sono soddisfatte, è possibile ottenere una soluzione più veloce utilizzando la procedura di Analisi Discriminante. Se tutte le variabili del tuo predittore sono categoriali, puoi anche utilizzare la procedura Loglineare. Se la tua variabile dipendente è continua, utilizzare la procedura di Regressione Lineare. È possibile utilizzare la procedura ROC Curve per tracciare le probabilità salvate con la procedura di Regressione Logistica.

### Ottenimento di un'analisi di regressione logistica

1. Dai menu, scegliere:

**Analizzare > Regressione > Binario Logistico ...**

2. Selezionare una variabile dipendente dicotomica. Questa variabile può essere numerica o stringa.

3. Selezionare una o più covariate. Per includere i termini di interazione, selezionare tutte le variabili coinvolte nell'interazione e quindi selezionare **> a\*b>**.

Per immettere le variabili in gruppi (**blocchi**), selezionare le covariate per un blocco e fare clic su **Avanti** per specificare un nuovo blocco. Ripetere fino a quando non sono stati specificati tutti i blocchi.

Facoltativamente, è possibile selezionare i casi per l'analisi. Scegliere una variabile di selezione ed inserire i criteri di regola.

## Regola Regressione Logistica

Nella stima del modello verranno inseriti i casi definiti dalla regola di selezione impostata. Ad esempio, se hai selezionato una variabile e **uguale** e specificato un valore di 5, allora solo i casi per i quali la variabile selezionata ha un valore pari a 5 sono inclusi nella stima del modello.

Le statistiche e i risultati della classificazione vengono generati sia per i casi selezionati che per i casi non selezionati. Questo fornisce un meccanismo per classificare nuovi casi in base ai dati precedentemente esistenti, o per partizionare i propri dati in sottoinsiemi di formazione e verifica, per eseguire la validazione sul modello generato.

## Metodi Di Selezione Variabile di Regressione Logistica

La selezione del metodo consente di specificare come vengono inserite nell'analisi le variabili indipendenti. Utilizzando diversi metodi, è possibile creare molteplici modelli di regressione dallo stesso insieme di variabili.

- *Invio*. Una procedura per la selezione delle variabili nella quale tutte le variabili di un blocco sono inserite in un unico passo.
- *Selezione in avanti (Condizionale)*. Metodo di selezione stepwise con test di inserimento basato sulla significatività della statistica di punteggio e con test di rimozione basato sulla probabilità di un rapporto di verosimiglianza fissato in base a stime condizionali dei parametri.
- *Selezione in avanti (Rapporto di verosimiglianza)*. Metodo di selezione stepwise con test di inserimento basato sulla significatività della statistica di punteggio e con test di rimozione basato sulla probabilità di un rapporto di verosimiglianza fissato in base a stime di massima verosimiglianza parziali.
- *Selezione in avanti (Wald)*. Metodo di selezione stepwise con test di inserimento basato sulla significatività della statistica di punteggio e con test di rimozione basato sulla probabilità della statistica di Wald.
- *Eliminazione all'indietro (Condizionale)*. Selezione a fasi all'indietro. Il test di rimozione è basato sulla probabilità della statistica dei rapporti di verosimiglianza sulla base delle stime condizionali dei parametri.
- *Eliminazione all'indietro (Rapporto di verosimiglianza)*. Selezione a fasi all'indietro. Il test di rimozione è basato sulla probabilità della statistica dei rapporti di verosimiglianza sulla base delle stime della massima verosimiglianza parziale.
- *Eliminazione all'indietro (Wald)*. Selezione a fasi all'indietro. Il test di rimozione è basato sulla probabilità della statistica di Wald.

I valori di significatività dell'output si basano sull'adattamento di un singolo modello. Pertanto, i valori di significatività sono generalmente non validi quando viene utilizzato un metodo stepwise.

Tutte le variabili indipendenti selezionate vengono aggiunte a un solo modello di regressione. È tuttavia possibile specificare diversi metodi di inserimento per diversi sottoinsiemi di variabili. Ad esempio, è possibile inserire un blocco di variabili nel modello di regressione utilizzando la selezione stepwise e un secondo blocco utilizzando la selezione in avanti. Per aggiungere un secondo blocco di variabili a un modello di regressione, fare clic su **Avanti**.

## Regressione Logistica Definire Variabili Categoricali

È possibile specificare i dettagli di come la procedura di Regressione Logistica gestirà le variabili categoriali:

**Covariati.** Contiene un elenco di 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**. Selezionare qualsiasi altra covariata categoriale dalla lista dei **Covariati** e spostarle nella lista **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) viene 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 predittore (tranne l'ultima categoria) viene paragonata alla categoria successiva.
- **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** o **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**, è necessario rimuovere tutti i termini contenenti la variabile dall'elenco **Covariati** nella finestra di dialogo principale.

## Regressione Logistica Salva nuove variabili

È possibile salvare i risultati della regressione logistica come nuove variabili nel dataset attivo:

**Valori previsti.** Salva i valori previsti dal modello. Le opzioni disponibili sono **Probabilità** e **appartenenza del Gruppo**.

- **Probabilità.** Per ogni caso, salva la probabilità prevista di ricorrenza dell'evento. Una tabella nell'output visualizza il nome e il contenuto di eventuali nuove variabili. L'"evento" è la categoria della variabile dipendente con il valore superiore; ad esempio, se la variabile dipendente assume valori 0 e 1, viene salvata la probabilità prevista di categoria 1.
- **Appartenenza al gruppo previsto.** Il gruppo con la probabilità a posteriori più elevata, basato sui punteggi discriminanti. Il gruppo di appartenenza del caso previsto dal modello.

**Influenza.** Salva i valori dalle statistiche che misurano l'influenza dei casi sui valori previsti. Le opzioni disponibili sono **Cook's**, **Leverage values** e **DfBeta(s)**.

- **Di Cook.** La regressione logistica analogica della statistica di influenza 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.
- **Valore di influenza.** L'influenza relativa di ogni osservazione sull'adattamento del modello.
- **DfBetas.** La differenza nel valore beta è la variazione del coefficiente di regressione risultante dall'esclusione di uno specifico caso. Un valore viene calcolato per ogni termine nel modello, compreso quello costante.



**Residui.** Salva i residui. Le opzioni disponibili sono Unstandardized, Logit, Studentized, Standardized e Deviance.

- *Residui non standardizzati.* La differenza tra un valore osservato e il valore stimato dal modello.
- *Residuo logit.* Il residuo per il caso se è previsto nella scala logit. Il residuo logit è il residuo diviso per la probabilità prevista e moltiplicato per 1 meno la probabilità prevista.
- *Residuo studentizzato.* La variazione della devianza nel modello se un caso viene escluso.
- *Residui 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.
- *Devianza.* Residui basati sulla devianza del modello.

**Esporta informazioni modello in file XML.** Le stime del parametro e, se si desidera, le relative covarianze vengono esportati nel file specificato 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.

## Opzioni Di Regressione Logistica

È possibile specificare le opzioni per l'analisi di regressione logistica:

**Statistiche e grafici.** Consente di richiedere statistiche e complotti. Le opzioni disponibili sono i tracciati di classificazione, Hosmer-Lemeshow bontà - di - fit, Casewise elenco dei residui, Correlazioni di stime, cronologia di Iterazione e CI per exp (*B*). Seleziona una delle alternative nel gruppo Display per visualizzare statistiche e complotti Ad ogni passo o, solo per il modello finale, all'ultimo passo.

- *Statistica di bontà di adattamento di Hosmer-Lemeshow.* Questa statistica della bontà di adattamento è più robusta della tradizionale statistica della bontà di adattamento utilizzata nella regressione logistica, in particolare per i modelli con covariate continue e gli studi con dimensioni del campione ridotte. È basata sul raggruppamento dei casi in decili di rischio e sul confronto della probabilità osservata con la probabilità prevista in ciascun decile.

**Probabilità per Stepwise.** Consente di controllare i criteri con cui le variabili vengono inserite e eliminate dall'equazione. È possibile specificare criteri per la Voce o la rimozione delle variabili.

- *Probabilità per passi.* Una variabile viene immessa nel modello se la probabilità della sua statistica di punteggio è inferiore al valore di inserimento e viene rimossa se la probabilità è maggiore del valore di rimozione. Specificare valori di inserimento e rimozione positivi per annullare le impostazioni predefinite. Il valore di inserimento deve essere inferiore al valore di rimozione.

**Valore di interruzione della classificazione.** Consente di determinare il punto di taglio per classificare i casi. I casi con valori attesi che superano il valore di riferimento sono classificati come positivi, mentre i casi con valori previsti minori del valore di riferimento sono classificati come negativi. Per modificare il valore predefinito, inserire un valore compreso tra 0,01 e 0,99.

**Numero massimo di iterazioni.** Consente di modificare il numero massimo di volte che il modello itera prima della terminazione.

**Includi costante nel modello.** Consente di indicare se il modello deve includere un termine costante. Se disabilitato, il termine costante sarà uguale a 0.

## REGRESSIONE LOGISTICA Comando Funzioni aggiuntive

Il linguaggio della sintassi dei comandi consente inoltre di:

- Identificare l'output casewise per i valori o le etichette variabili di una variabile.
- Controllare la spaziatura dei rapporti di iterazione. Piuttosto che stampare le stime dei parametri dopo ogni iterazione, è possibile richiedere le stime dei parametri dopo ogni *nesimo* iterazione.
- Modificare i criteri per la terminazione e il controllo della ridondanza.
- Specificare un elenco di variabili per i listini caseari.

- Conserve la memoria tenendo i dati per ogni gruppo di file suddiviso in un file di scratch esterno durante l'elaborazione.

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

## Regressione logistica multinomiale

la regressione logistica multinomiale è utile per le situazioni in cui si desidera poter classificare i soggetti in base ai valori di una serie di variabili predittive. Questo tipo di regressione è simile alla regressione logistica, ma è più generale perché la variabile dipendente non è limitata a due categorie.

**Esempio.** Per commercializzare i film in modo più efficace, gli studios cinematografici vogliono prevedere quale tipo di film un moviegoer possa vedere. Eseguendo una Regressione Logistica Multinomiale, lo studio può determinare la forza di influenzare l'età di una persona, il sesso e lo status di frequentazione ha sul tipo di film che preferiscono. Lo studio può poi sbattere la campagna pubblicitaria di un film particolare verso un gruppo di persone che probabilmente la vedranno.

**Statistiche.** Storia dell'iterazione, coefficienti di parametro, covarianza asintotica e matrici di correlazione, test di verosimiglianza per modello e effetti parziali,  $-2$  logarit-verosimiglianza. Pearson e devianza chi - quadrato bontà di adattamento. Cox e Snell, Nagelkerke e McFadden  $R^2$ . Classificazione: osservata versus frequenze previste per categoria di risposta. Crosstabulazione: frequenze osservate e previste (con residui) e proporzioni per modello covariato e categoria di risposta.

**Metodi.** Un modello di logit multinomiale è adatto al modello factoriale completo o a un modello specificato dall'utente. La stima dei parametri viene eseguita attraverso un algoritmo iterativo di massima verosimiglianza.

### considerazioni sui dati di regressione logistica multinomiale

**Dati.** La variabile dipendente dovrebbe essere categoriale. Le variabili indipendenti possono essere fattori o covariate. In generale, i fattori dovrebbero essere variabili categoriali e le covariate dovrebbero essere variabili continue.

**Ipotesi.** Si ipotizza che il rapporto di odds ratio di ogni due categorie sia indipendente da tutte le altre categorie di risposta. Ad esempio, se un nuovo prodotto viene introdotto in un mercato, questa supposizione afferma che le quote di mercato di tutti gli altri prodotti sono influenzate proporzionalmente. Inoltre, dato uno schema covariato, le risposte sono assunte per essere variabili multinazionali indipendenti.

### Ottenimento di una regressione logistica multinomiale

1. Dai menu, scegliere:

**Analizzare > Regressione > Regressione Logistica Multinomiale ...**

2. Selezionare una variabile dipendente.

3. I fattori sono facoltativi e possono essere numerici o categoriali.

4. Le covariate sono facoltative ma devono essere numeriche se specificate.

## Regressione logistica multinomiale

Per impostazione predefinita, la procedura di Regressione Logistica Multinomiale produce un modello con il fattore e covariando gli effetti principali, ma è possibile specificare un modello personalizzato o richiedere una selezione del modello stepwise con questa finestra di dialogo.

**Specifica modello.** Un modello di effetti principali include gli effetti principali di covariate e fattori, ma non gli effetti di interazione. Un modello fattoriale completo contiene tutti gli effetti principali e tutte le interazioni fattore - by - factor. Non contiene interazioni di covariate. È possibile creare un modello personalizzato per specificare sottoinsiemi di interazioni del fattore o interazioni covariate o richiedere la selezione stepwise dei termini del modello.

**Fattori & Covariati.** Vengono elencati i fattori e le covariate.

**Termini di ingresso forzato.** Termini aggiunti all'elenco di ingresso forzato sono sempre inclusi nel modello.

**Termini Stepwise.** Termini aggiunti alla lista stepwise sono inclusi nel modello secondo uno dei seguenti Metodi Stepwise selezionati dall'utente:

- **Ingresso in avanti.** Questo metodo inizia senza termini di stepwise nel modello. Ad ogni passo, il termine più significativo viene aggiunto al modello fino a quando nessuno dei termini stepwise lasciati fuori dal modello avrebbe un contributo statisticamente significativo se aggiunto al modello.
- **eliminazione indietro.** Questo metodo inizia inserendo tutti i termini specificati nella lista stepwise nel modello. Ad ogni passo, il termine stepwise meno significativo viene rimosso dal modello fino a quando tutti i restanti termini stepwise hanno un contributo statisticamente significativo al modello.
- **Stepwise in avanti.** Questo metodo inizia con il modello che sarebbe stato selezionato con il metodo di ingresso in avanti. Da lì, l'algoritmo si alterna all'eliminazione all'indietro sui termini stepwise nel modello e all'ingresso in avanti sui termini lasciati fuori dal modello. Questo continua fino a quando non vengono soddisfatti i criteri di ingresso o di rimozione.
- **Backward stepwise.** Questo metodo inizia con il modello che sarebbe stato selezionato con il metodo di eliminazione all'indietro. Da lì, l'algoritmo si alterna tra l'ingresso in avanti sui termini lasciati fuori dal modello e l'eliminazione all'indietro sui termini stepwise nel modello. Questo continua fino a quando non vengono soddisfatti i criteri di ingresso o di rimozione.

**Includi intercetta nel modello.** Consente di includere o escludere un termine di intercettazione per il modello.

## Crea termini

Per i fattori e le covariate selezionati:

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

**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.

## Multinomiale Logistic Regression Reference Categoria

Per impostazione predefinita, la procedura Regressione Logistica Multinomiale rende l'ultima categoria la categoria di riferimento. Questa finestra di dialogo ti dà il controllo della categoria di riferimento e il modo in cui vengono ordinate le categorie.

**Categoria di riferimento.** Specificare la prima, ultima o una categoria personalizzata.

**Ordine di categoria.** In ordine crescente, il valore più basso definisce la prima categoria e il valore più alto definisce l'ultimo. In ordine decrescente, il valore più alto definisce la prima categoria e il valore più basso definisce l'ultimo.

## Statistiche Di Regressione Logistica Multinomiale

È possibile specificare le seguenti statistiche per la Regressione Logistica Multinomiale:

**Riepilogo elaborazione casi.** Questa tabella contiene informazioni sulle variabili categoriali specificate.

**Modello.** Statistiche per il modello complessivo.

- **Pseudo R-square.** Stampa le statistiche Cox e Snell, Nagelkerke e McFadden  $R^2$ .

- **Riepilogo passo.** Questa tabella riepiloga gli effetti immessi o rimossi ad ogni passo in un metodo stepwise. Non viene prodotto a meno che non venga specificato un modello stepwise nella finestra di dialogo Modello.
- **Inform. su adattam. modello.** Questa tabella confronta i modelli montati e intercept - only o null.
- **Criteri di informazione.** Questa tabella stampa il criterio di informazione di Akaike (AIC) e il criterio di informazione Bayesiana di Schwarz (BIC).
- **Probabilità di cella.** Stampa una tabella delle frequenze osservate e attese (con residui) e proporzioni per modello covariato e categoria di risposta.
- **Tabella di classificazione.** Stampa una tabella delle risposte osservate versus previste.
- **Bontà delle statistiche di chi - quadrato.** Stampa Pearson e probabilità - rapporto chi - quadrato. Le statistiche sono calcolate per i modelli covariati determinati da tutti i fattori e le covariate o da un sottoinsieme definito dall'utente dei fattori e delle covariate.
- **Misure di Monotonicità.** Visualizza una tabella con informazioni sul numero di coppie concordanti, coppie discordanti e coppie legate. In questa tabella sono inoltre visualizzati D di Somers, Gamma di Goodman e Kruskal, tau-a di Kendall e l'indice di concordanza C.

**Parametri.** Statistiche relative ai parametri del modello.

- **stime.** Stampa le stime dei parametri del modello, con un livello di confidenza specificato dall'utente.
- **Test rapporto di probabilità.** Stampa prove di rapporto di verosimiglianza per gli effetti parziali del modello. Il test per il modello complessivo viene stampato automaticamente.
- **Correlazioni asintotiche.** Stampa matrice di stima dei parametri correlazioni.
- **Covarianze asintotiche.** Stampa matrice di covarianze di stima dei parametri.

**Definisci sottopopolazioni.** Consente di selezionare un sottoinsieme dei fattori e delle covariate al fine di definire i modelli di covariata utilizzati dalle probabilità di cella e i test di bontà.

## Criteri Di Regressione Logistica Multinomiale

È possibile specificare i seguenti criteri per la Regressione Logistica Multinomiale:

**Iterazioni.** Consente di specificare il numero massimo di volte in cui si desidera effettuare il ciclo attraverso l'algoritmo, il numero massimo di passi nel passo - dimezzamento, le tolleranze di convergenza per le modifiche nel log - verosimiglianza e i parametri, quanto spesso viene stampato l'avanzamento dell'algoritmo iterativo e a quale iterazione la procedura deve iniziare la verifica della separazione completa o quasi completa dei dati.

- **Convergenza logaritmo della verosimiglianza.** Si ipotizza la convergenza se la modifica assoluta nella funzione log - verosimigliante è inferiore al valore specificato. Il criterio non viene utilizzato se il valore specificato è 0. Specificare un valore non negativo.
- **Convergenza parametri.** Convergenza si assume se il cambiamento assoluto nelle stime dei parametri è inferiore a questo valore. Il criterio non viene utilizzato se il valore specificato è 0.

**Delta.** Consente di specificare un valore non negativo inferiore a 1. Questo valore viene aggiunto a ciascuna cella vuota della crosstabulazione della categoria di risposta mediante modello covariato. Questo aiuta a stabilizzare l'algoritmo e prevenire i pregiudizi nelle stime.

**Tolleranza della singolarità.** Consente di specificare la tolleranza utilizzata nel controllo delle singolarità.

## Opzioni Di Regressione Logistica multinomiale

È possibile specificare le seguenti opzioni per la Regressione Logistica Multinomiale:

**Scala di dispersione.** Consente di specificare il valore di scaling di dispersione che verrà utilizzato per correggere la stima della matrice di covarianza dei parametri. L'opzione **Devianza** effettua una stima del valore di scala usando la funzione di devianza (statistica Chi-quadrato del rapporto di verosimiglianza). L'opzione **Pearson** effettua una stima del valore di scala usando la statistica Chi-quadrato di Pearson. È anche possibile specificare il proprio valore di scaling. Deve essere un valore numerico positivo.

**Opzioni Stepwise.** Queste opzioni ti danno il controllo dei criteri statistici quando vengono utilizzati metodi stepwise per costruire un modello. Vengono ignorati a meno che non venga specificato un modello stepwise nella finestra di dialogo Modello.

- **Probabilità di ingresso.** Questa è la probabilità della statistica di probabilità - rapporto per l'ingresso variabile. Più grande è la probabilità specificata, più facile è per una variabile entrare nel modello. Questo criterio viene ignorato a meno che non venga selezionato l'ingresso in avanti, il metodo stepwise in avanti o indietro.
- **Test di ingresso.** Questo è il metodo per inserire termini in metodi stepwise. Scegli tra il test di probabilità - rapporto e test di punteggio. Questo criterio viene ignorato a meno che non venga selezionato l'ingresso in avanti, il metodo stepwise in avanti o indietro.
- **Probabilità di rimozione.** Questa è la probabilità della statistica di probabilità - rapporto per la rimozione variabile. Più grande è la probabilità specificata, più facile è per una variabile rimanere nel modello. Questo criterio viene ignorato a meno che non venga selezionata l'eliminazione all'indietro, il metodo stepwise in avanti o indietro.
- **Test di rimozione.** Questo è il metodo per la rimozione dei termini in metodi stepwise. Scegli tra la prova di probabilità - rapporto e test Wald. Questo criterio viene ignorato a meno che non venga selezionata l'eliminazione all'indietro, il metodo stepwise in avanti o indietro.
- **Minimo degli effetti in modello.** Quando si utilizzano i metodi di eliminazione all'indietro o all'indietro, questo specifica il numero minimo di termini da includere nel modello. L'intercettazione non è conteggiata come termine modello.
- **Massimo degli effetti in modello.** Quando si utilizzano i metodi di inoltro o stepwise in avanti, questo specifica il numero massimo di termini da includere nel modello. L'intercettazione non è conteggiata come termine modello.
- **Hierarchicamente vincola l'ingresso e la rimozione dei termini.** Questa opzione consente di scegliere se porre limitazioni all'inclusione dei termini del modello. La gerarchia richiede che per qualsiasi termine da inserire, tutti i termini di ordine inferiore che fanno parte del termine da includere devono essere nel modello prima. Ad esempio, se il requisito della gerarchia è in vigore, i fattori *Stato Marital* e *Gender* devono essere entrambi nel modello prima che si possa aggiungere l'interazione *Stato Marital \* Gender*. Le tre opzioni del pulsante radio determinano il ruolo delle covariate nella determinazione della gerarchia.

## Multinomiale Logistic Regression Save

La finestra di dialogo Salva consente di salvare le variabili sul file di lavoro ed esportare le informazioni del modello in un file esterno.

**Variabili salvate.** Le seguenti variabili possono essere salvate:

- **Probabilità di risposta stimate.** Sono le probabilità stimate di classificare un modello di fattore / covariato nelle categorie di risposta. Ci sono altrettante probabilità stimate poiché ci sono categorie della variabile di risposta; fino a 25 verranno salvate.
- **Categoria prevista.** Questa è la categoria di risposta con la più grande probabilità prevista per uno schema di fattore / covariato.
- **Probabilità di categoria previste.** Questo è il massimo delle probabilità di risposta stimate.
- **Probabilità di categoria reale.** Questa è la probabilità stimata di classificare un modello di fattore / covariato nella categoria osservata.

**Esporta informazioni modello in file XML.** Le stime del parametro e, se si desidera, le relative covarianze vengono esportati nel file specificato 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 NOMREG

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare la categoria di riferimento della variabile dipendente.

- Includere i casi con valori mancanti dell'utente.
- Personalizzare i test di ipotesi specificando ipotesi nulle come combinazioni lineari di parametri.

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

## Regressione Probit

Questa procedura misura la relazione tra l'intensità di uno stimolo e la proporzione dei casi che mostrano una determinata risposta allo stimolo. È utile per le situazioni in cui è presente un output dicotomico che si presume sia influenzato o causato da livelli di alcune variabili indipendenti ed è particolarmente idoneo per i dati sperimentali. Questa procedura consentirà di stimare l'intensità di uno stimolo richiesta per indurre una determinata proporzione di risposte, ad esempio la mediana della dose effettiva.

**Esempio.** Quanto è efficace un nuovo pesticida per l'uccisione di formiche, e qual è una concentrazione adeguata da utilizzare? Potreste eseguire un esperimento in cui esporre campioni di formiche a diverse concentrazioni del pesticida e poi registrare il numero di formiche uccise e il numero di formiche esposte. Applicando la regressione probit a questi dati, è possibile determinare la forza del rapporto tra concentrazione e uccisione, e si può determinare quale sarebbe l'adeguata concentrazione di pesticidi se si volesse essere sicuri di uccidere, diciamo, il 95% delle formiche esposte.

**Statistiche.** Coefficienti di regressione e errori standard, intercettazione e errore standard, Pearson bontà - di - fit chi - quadrato, frequenze osservate e attese e intervalli di confidenza per livelli efficaci di variabili indipendenti. Complotti: complotti di risposta trasformati.

Questa procedura utilizza gli algoritmi proposti e implementati in NPSOL<sup>®</sup> da Gill, Murray, Saunders & Wright per stimare i parametri del modello.

### considerazioni sui dati di Probit Regression

**Dati.** Per ciascun valore della variabile indipendente (o ciascuna combinazione di valori per più variabili indipendenti), la variabile della risposta deve essere un conteggio del numero di casi con i valori che mostrano la risposta di interesse e la variabile osservata totale deve essere un conteggio del numero totale di casi con tali valori per la variabile indipendente. La variabile fattore deve essere categoriale, codificata come numeri interi.

**Ipotesi.** Le osservazioni dovrebbero essere indipendenti. Se si dispone di un gran numero di valori per le variabili indipendenti rispetto al numero di osservazioni, come si potrebbe in uno studio osservazionale, le statistiche di chi - quadrato e bontà possono non essere valide.

**Procedure correlate.** L'analisi probit è strettamente correlata alla regressione logistica; infatti, se si sceglie la trasformazione logit, questa procedura calcola essenzialmente una regressione logistica. In generale, l'analisi probit è appropriata per esperimenti progettati, mentre la regressione logistica è più appropriata per gli studi osservazionali. Le differenze in output riflettono questi diversi punti di forza. La procedura di analisi probit riporta stime di valori efficaci per vari tassi di risposta (compresa la dose efficace mediana), mentre la procedura di regressione logistica riporta stime di odds ratio per variabili indipendenti.

### Ottenimento di un'analisi Probit Regression

1. Dai menu, scegliere:

**Analizzare > Regressione > Probit ...**

2. Selezionare una variabile di frequenza di risposta. Questa variabile indica il numero di casi che esibisce una risposta allo stimolo di prova. I valori di questa variabile non possono essere negativi.
3. Selezionare una variabile totale osservata. Questa variabile indica il numero di casi a cui è stato applicato lo stimolo. I valori di questa variabile non possono essere negativi e non possono essere inferiori ai valori della variabile di frequenza di risposta per ciascun caso.

Facoltativamente è possibile selezionare una variabile di Fattore. Se si fa, utilizzare **Definisci intervallo** per definire la gamma per i gruppi.

4. Selezionare una o più covariate. Questa variabile contiene il livello dello stimolo applicato ad ogni osservazione. Se si desidera trasformare la covariata, selezionare una trasformazione dall'elenco a discesa **Trasforma** . Se non viene applicata alcuna trasformazione e c'è un gruppo di controllo, allora il gruppo di controllo è incluso nell'analisi.
5. Selezionare il modello **Probit** o **Logit** .

#### **Modello Probit**

Applica la trasformazione Probit (l'inversa della funzione di distribuzione cumulata normale standard) alle proporzioni di risposta.

#### **Modello logit**

Applica la trasformazione logit (odds logaritmici) alle proporzioni di risposta.

## **Analisi probit Definisci intervallo**

Ciò consente di specificare i livelli della variabile fattore che verranno analizzati. I livelli dei fattori devono essere codificati come numeri interi consecutivi e tutti i livelli nella gamma che si specifica verranno analizzati.

## **Opzioni di analisi probit**

È possibile specificare le opzioni per l'analisi probit:

**Statistiche.** Consente di richiedere le seguenti statistiche facoltative: Frequenze, Potenza mediana relativa, test di Parallelismo e intervalli di confidenza Fiduale.

- *Potenza mediana relativa.* Visualizza il rapporto delle potenze mediane per ogni coppia di livelli dei fattori. Mostra anche i limiti di confidenza al 95% per ogni potenza mediana relativa. Le potenze mediane relative non sono disponibili se non si ha una variabile fattore o se si ha più di una covariata.
- *Test di parallelismo.* Un test dell'ipotesi che tutti i livelli di un fattore abbiano una pendenza comune.
- *Intervalli di confidenza fiduciari.* Intervalli di confidenza per il dosaggio richiesto per ottenere una certa probabilità di risposta.

Gli intervalli di confidenza di riferimento e la potenza mediana Relativa non sono disponibili se hai selezionato più di una covariata. La potenza mediana relativa e il test di Parallelismo sono disponibili solo se hai selezionato una variabile fattore.

**Tasso di risposta naturale.** Consente di indicare un tasso di risposta naturale anche in assenza dello stimolo. Le alternative disponibili sono None, Calcola da dati o Value.

- *Calcola dai dati.* Stima il tasso di risposta naturale dai dati del campione. I tuoi dati dovrebbero contenere un caso che rappresenta il livello di controllo, per il quale il valore delle covariate è di 0. Probit stima il tasso di risposta naturale utilizzando la proporzione di risposte per il livello di controllo come valore iniziale.
- *Valore.* Imposta il tasso di risposta naturale nel modello (selezionare questo elemento quando si conosce il tasso di risposta naturale in anticipo). Il valore da inserire è la proporzione di risposte, ovvero un valore fra 0 e 1. Ad esempio, se la risposta si verifica il 10% delle volte che lo stimolo è 0, immettere 0.10.

**Criteri.** Consente di controllare i parametri dell'algoritmo di stima dei parametri iterativi. È possibile sovrascrivere i valori predefiniti per le iterazioni massime, il limite di passo e la tolleranza di Optimality.

## **PROBIT Command Aggiuntivo Features**

Il linguaggio della sintassi dei comandi consente inoltre di:

- Richiedere un'analisi su entrambi i modelli probit e logit.
- Controllare il trattamento dei valori mancanti.
- Trasformare le covariate per basi diverse dalla base 10 o log naturale.

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

## Regressione quantile

Regressione è un metodo statistico ampiamente utilizzato nella modellazione quantitativa. La regressione lineare multipla è un approccio di base e standard in cui i ricercatori utilizzano i valori di diverse variabili per spiegare o prevedere i valori medi di un risultato su scala. Tuttavia, in molte circostanze, siamo più interessati alla mediana, o a un quantile arbitrario del risultato su scala.

I modelli di regressione quantile la relazione tra una serie di variabili predittiva (indipendenti) e percentili specifici (o "quantili") di una variabile target (dipendente), più spesso la mediana. Ha due principali vantaggi rispetto alla regressione Ordinaria Minimi Quadrati:

- La regressione quantile non fa supposizioni sulla distribuzione della variabile target.
- La regressione quantile tende a resistere all'influenza delle osservazioni stravaganti

La regressione quantile è ampiamente utilizzata per la ricerca in industrie come l'ecologia, l'assistenza sanitaria e l'economia finanziaria.

### Esempio

Qual è il rapporto tra il reddito familiare totale e la percentuale di reddito che viene speso per il cibo? La legge di Engel è un'osservazione in economia affermando che come aumenta il reddito, la percentuale di reddito speso per il cibo cade, anche se la spesa assoluta per gli alimenti sale. Applicando la regressione quantile a questi dati, è possibile determinare quali spese alimentari possono coprire il 90% delle famiglie (per 100 famiglie con un determinato reddito) quando non interessate alla spesa media alimentare.

### Statistica

Quantitative Regression, Simplex approach, Frisch - Newton interior - point non - linear optimization algoritmo, Barrodale e Roberts, Bofito, Hall Sheather, larghezza di banda, livello di significatività, manipolazioni di matrice, criterio di convergenza, pesi di regressione, termine di intercettazione, residui di previsione, tabulazione, tabelle di previsione, stime dei parametri, matrice di covarianza, matrice di correlazione, valori osservati, intervallo di confidenza, intervallo di confidenza.

Questa procedura utilizza gli algoritmi proposti da Koenker, R. W. e Bassett, G. W. (1978). Regressione quantili, *Econometrica*, 46, 33-50.

## considerazioni sui dati di regressione quantile

### Dati

È necessaria una singola variabile dipendente numerica. La variabile di destinazione deve essere una variabile continua. I predittori possono essere variabili continue o variabili dummy per i predittori categoriali. O il termine di intercettazione o almeno un predittore è richiesto per eseguire un'analisi.

### Ipotesi

La regressione quantile non fa supposizioni sulla distribuzione della variabile target e resiste all'influenza delle osservazioni stravaganti.

### Procedure correlate

L'analisi quantile è relativa alla regressione Ordinaria Minimi Quadrati.

## Ottenimento di un'analisi Quantile Regression

1. Dai menu, scegliere:

**Analizzare > Regressione > Quantile ...**

La finestra di dialogo consente di specificare le variabili target, fattore, covariato e di peso da utilizzare per l'analisi di regressione quantile. La finestra di dialogo fornisce anche l'opzione di conservare la memoria per analisi complesse o grandi datasets.

2. Selezionare una variabile di destinazione numerica. È necessaria una sola variabile di destinazione per eseguire un'analisi. Sono consentite solo variabili numeriche.
3. Facoltativamente, selezionare una o più variabili del fattore. Le variabili di scala non sono consentite.
4. Facoltativamente, selezionare una o più variabili covariate. Le variabili di stringa non sono consentite.



**Nota:** Quando sia gli elenchi **Fattore (s)** che **Covariate (s)** sono vuoti e **Includi l'intercettazione nel modello** viene selezionato sulla finestra di dialogo Modello, i seguenti pannelli di messaggio:

No effects have been specified. Therefore, an intercept only model will be fit.  
Do you want to fit an intercept-only model?

5. Opzionalmente, selezionare una variabile peso di regressione. Le variabili di stringa non sono consentite.
6. Opzionalmente, selezionare **Conserva memoria per analisi complesse o grandi datasets**. Questa impostazione controlla se i dati si tengono o meno in un file esterno durante l'elaborazione. L'abilitazione dell'impostazione può aiutare a conservare le risorse di memoria durante l'esecuzione di analisi complesse, o le analisi con i dataset di grandi dimensioni.

## Regressione quantile: criteri

Dialogo Criteri fornisce le opzioni per

### Quantile

Fornisce le opzioni per specificare i quantili.

#### Specifica quantili singoli

Quando viene selezionato, è necessario almeno un valore per eseguire l'analisi. Sono consentiti più valori e ogni valore deve appartenere a  $[0, 1]$ . È possibile specificare più valori con ogni valore separato da uno spazio vuoto (o spazi vuoti). Utilizzare i pulsanti **Aggiungi**, **Modifica** e **Rimuovi** per lavorare con i valori nella lista dei valori quantili.

Tutti i valori devono essere univoci (non sono consentiti valori duplicati). Il valore predefinito è 0.5.

#### Specifica quantili griglia

Quando selezionato, una griglia di quantili può essere specificata da un valore **Start** (value1) ad un valore **Fine** (value2) con l'incremento di **By** (value3). Se specificato, è consentito solo una serie valida di  $[\text{value1 TO value2 BY value3}]$ . Deve soddisfare tale  $0 \leq \text{value1} \leq \text{value2} \leq 1$ . Nei casi in cui  $\text{value1} = \text{value2}$ , equivale a specificare un singolo value1, indipendentemente da value3.

### Metodo di stima

Fornisce le opzioni per specificare il metodo di stima del modello.

#### Scelto automaticamente dal programma

Consente alla procedura di selezionare automaticamente il metodo di stima appropriato. Questa è l'impostazione predefinita.

#### Algoritmo semplice

Richiama l'algoritmo simplex sviluppato da Barrodale e Roberts.

#### Ottimizzazione non lineare punto interno Frisch-Newton

Chiamate per l'algoritmo di ottimizzazione non lineare di Frisch - Newton.

### Post-stima

Fornisce opzioni per la post - stima della varianza - covarianza delle stime dei parametri e degli intervalli di confidenza per i valori di destinazione previsti.

#### Presumi che i casi siano IID

Quando selezionata, questa impostazione presuppone che i termini di errore siano indipendenti e identicamente distribuiti. Quando l'impostazione non viene selezionata, il tempo di calcolo può aumentare in modo significativo per i modelli di grandi dimensioni. L'impostazione viene selezionata per impostazione predefinita.

#### Tipo di banda

Determina quale metodo di banda viene utilizzato per stimare la matrice di varianza - covarianza delle stime dei parametri (**Bofinger** o **Hall - Sheather**). **Bofinger** è l'impostazione predefinita.

### Metodo numerico

Fornisce le seguenti opzioni:

**Tolleranza della singolarità**

Specifica il valore di tolleranza per le manipolazioni a matrice nel metodo dell'interno. Il valore specificato deve essere un singolo, doppio valore in (0,  $10^{-3}$ ), con  $10^{-12}$  come impostazione predefinita.

**Convergenza**

Specifica il criterio di convergenza per il metodo numerico. Il valore specificato deve essere un singolo, doppio valore in (0,  $10^{-3}$ ), con  $10^{-6}$  come impostazione predefinita.

**Numero massimo di iterazioni**

Specifica il numero massimo di iterazioni. Il valore specificato deve essere un intero singolo positivo. Il valore predefinito è 2000.

**Valori mancanti**

Fornisce le opzioni per determinare come vengono gestiti i valori mancanti.

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

Quando vengono selezionati, sono esclusi entrambi i valori mancanti di utente e di sistema.

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

Quando vengono selezionati, i valori mancanti dell'utente sono trattati come validi.

**Intervallo di confidenza (%)**

Specifica il livello di significatività. Quando specificato, il valore deve essere un singolo doppio valore compreso tra 0 e 100. Il valore predefinito è 95.

## Regressione quantile: modello

La finestra di dialogo Modello fornisce opzioni per specificare gli effetti e i pesi utilizzati nel modello. Se omissa, o specificato da solo, il modello conterrà il termine di intercettazione e tutti gli effetti principali con le covariate nell'elenco delle covariate e i fattori in elenco fattori.

**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. Quando si seleziona **Crea termini**, sono disponibili le seguenti opzioni di effetto e interazione per i termini non nidificati.

**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.

Quando **Crea termini nidificati** è selezionato, è possibile costruire termini nidificati. 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.

**Note di nidificazione:**

- Per includere un effetto per un'interazione tra due fattori, utilizzare la parola chiave BY o l'asterisco (\*) per unire i fattori implicati nell'interazione.
- I fattori all'interno di un effetto di interazione devono essere distinti.
- Utilizzare coppie di parentesi per includere un effetto per la nidificazione di un termine all'interno di un altro.
- Quando sono presenti più coppie di parentesi, ogni coppia di parentesi deve essere racchiusa o nidificata all'interno di un'altra coppia di parentesi.
- È consentita la nidificazione multipla.
- Le interazioni tra gli effetti nidificati non sono supportate.

**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  $A(X)$ .

#### **Includi intercetta nel modello**

Quando selezionato, il termine di intercettazione è incluso nel modello. Quando non viene selezionato, almeno un predittore è necessario per eseguire l'analisi. L'impostazione è abilitata per impostazione predefinita.

## **Regressione quantile: visualizzazione**

La finestra di dialogo Visualizza fornisce impostazioni di controllo di output e trama.

#### **Stampa**

Sono disponibili le seguenti opzioni di output.

#### **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.

#### **Crea grafico e tabella**

Sono disponibili le seguenti opzioni di plotting:

#### **Traccia le stime parametro per**

È possibile selezionare le stime dei parametri per un numero specifico di effetti top o per tutti gli effetti nel modello. **Top xx effects** impostazione controlla il numero di categorie, o combinazioni di categorie in un effetto misto, che vengono tracciate all'interno dell'interazione di una covariata e di uno o due fattori. Il valore deve essere un singolo numero intero positivo (50 è l'impostazione predefinita).

#### **Note:**

- I tracciati di previsione vengono creati per tutti gli effetti quando il valore intero specificato è più grande del numero di categorie o combinazioni.
- L'impostazione è valida solo quando vengono specificati più valori per l'impostazione **Quantile valori** nella finestra di dialogo Criteri. Non vengono creati complotti quando viene specificato un solo valore quantile.

### **Visualizza il grafico Previsioni attraverso osservazioni**

Controlla la creazione della trama dei valori previsti versus osservati. Quando è abilitata, viene creata una singola trama che contiene i punti (con diversi doti colorati che rappresentano diversi quantili). L'impostazione è disabilitata per impostazione predefinita.

### **Prevedi gli effetti nel modello**

Quando è abilitata, sono disponibili le seguenti opzioni:

#### **Trama o tabula gli effetti x top**

Specificare il numero di effetti di primo piano la cui tabella di previsione o la tabella di previsione verranno create. 3 è il valore predefinito.

**Nota:** Le tabelle di previsione o le tabelle di previsione sono create per tutti gli effetti quando il valore specificato è più grande del numero di effetti validi nel modello.

#### **Grafico o tabella degli effetti specificati dall'utente**

Le linee guida ad effetto valide sono:

- Effetto con una covariata (incluso un alto potere della covariata stessa): Creare una singola trama contenente le linee previste da diversi quantili.
- Effetto con un solo fattore: Tabulare le previsioni per le categorie del fattore da diversi quantili.
- Effetto con l'interazione di due fattori: Per ogni quantile, tabellare le previsioni per le categorie di due fattori.
- Effetto con l'interazione di una covariata e uno o due fattori: Per ogni quantile, creare una trama contenente le righe per ogni categoria o combinazione delle categorie all'interno dell'effetto interazione.
- Il numero massimo delle combinazioni da tracciare è controllato dal valore specificato per **Categorie massime xx di combinazioni di categorie in un effetto misto**.

Gli effetti che vengono spostati da **Modello Effetti** all'elenco **Linee di previsione** vengono utilizzati per il tracciamento. I tracciati non vengono creati nei casi in cui gli effetti specificati sono costanti (rimossi dall'edificio del modello).

Gli effetti che vengono spostati da **Modello Effetti** all'elenco **Tabelle di previsione** vengono utilizzati per la tabulazione. Le tabelle non vengono create nei casi in cui gli effetti specificati sono costanti (rimossi dall'edificio del modello).

#### **Trama massima xx categorie o combinazioni di categorie in un effetto misto**

Controlla il numero massimo delle combinazioni di categoria alla trama. Il valore predefinito è 10.

## **Regressione quantile: salva**

La finestra di dialogo Salva fornisce le opzioni per la valutazione del modello.

### **Valore previsto della risposta**

Quando viene selezionato, viene segnato il valore di destinazione previsto.

### **Residuo**

Quando vengono selezionati, i residui di previsione vengono segnati.

### **Limite inferiore dell'intervallo di previsione**

Quando vengono selezionati, vengono segnati i limiti inferiori degli intervalli di previsione.

### **Limite superiore dell'intervallo di previsione**

Quando vengono selezionati, vengono segnati i limiti superiori degli intervalli di previsione.

**Nota:** Un nome variabile può essere specificato per ogni opzione di salvataggio. Se viene specificato un nome root, deve essere un nome di variabile valido. Un nome radice, seguito da un carattere di sottolineatura "\_" e da un suffisso quantile significativo, viene utilizzato quando vengono specificati più valori per l'impostazione **Quantile valori** nella finestra di dialogo Criteri.

## Regressione quantile: esporta

La finestra di dialogo Esporta fornisce opzioni per specificare quali statistiche vengono esportate, come le statistiche vengono esportate (file di dati esterni o dataset) e controllando come i dati vengono gestiti durante l'elaborazione (processo normalmente, oppure tenuti in un file di scratch esterno mentre si elabora).

### Matrice di covarianza delle stime dei parametri

Quando selezionate, vengono abilitate le opzioni per la scrittura della matrice di covarianza delle stime dei parametri in un file di dati esterno o un dato precedentemente dichiarato.

### Matrice di correlazione delle stime dei parametri

Quando vengono selezionate, sono abilitate le opzioni per la scrittura della matrice di correlazione delle stime dei parametri in un file di dati esterno o un dataset precedentemente dichiarato.

### La matrice di covarianza / correlazione sarà salvata in un unico dataset o file esterno in presenza di più quantili di regressione

Quando sono presenti più quantili, questa opzione attiva il salvataggio delle matrici di covarianza / correlazione a dataset singoli o multipli o file di dati esterni. Quando non è abilitata, le matrici vengono salvate in un singolo file di dati esterno o in un dataset. Quando abilitati, le matrici vengono salvate in più file di dati esterni o set di dati. L'impostazione prende effetto solo quando vengono specificati più valori per l'impostazione **Quantile valori** nella finestra di dialogo Criteri.

**Nota:** Questa opzione è disponibile solo quando viene selezionata **Matricianza matrice delle stime dei parametri** o **Correlazione matrice delle stime dei parametri**.

### Esporta informazioni modello in file XML

Quando selezionato, fornisce le opzioni per l'esportazione delle informazioni del modello in un nome e posizione file XML specifico.

### Esporta come XML

Quando si seleziona **Esporta informazioni modello su file XML**, è possibile selezionare di esportare le stime dei parametri e le matrici di covarianza o solo le stime dei parametri. **Le stime dei parametri e la matrice di covarianza** è l'impostazione predefinita.

## Convenzioni nome file

- Quando viene specificato un solo valore per l'impostazione **Quantile valori** nella finestra di dialogo Criteri, savefile e dataset vengono utilizzati per denominare il file di dati esterno o il dataset.
- Quando vengono specificati più valori per l'impostazione **Quantile valori** nella finestra di dialogo Criteri, ogni quantile viene salvato su un file di dati o dataset esterni.
- Un carattere di sottolineatura "\_", seguito da un suffisso quantile significativo, viene accostato automaticamente al file di dati o al nome data set. Ad esempio, quando 0.25, 0.50e 0.75 vengono specificati come **Valori quantità**, il suffisso \_25, \_50e \_75 vengono accodati ai nomi file di dati (prima dell'estensione .sav).
- Ulteriori cifre possono essere specificate per ogni suffisso quantile (se necessario).
- I **Valori Quantili** leader zero e virgola decimale non vengono utilizzati nel suffisso.
- Quando la notazione scientifica viene specificata per **Valori Quantile**, viene convertita in un valore decimale quando viene visualizzata nel suffisso.

## Regressione non lineare

La regressione non lineare è un metodo per trovare un modello non lineare del rapporto tra la variabile dipendente e un insieme di variabili indipendenti. A differenza della regressione lineare tradizionale, che si limita a stimare modelli lineari, la regressione non lineare può stimare modelli con relazioni arbitrarie tra variabili indipendenti e dipendenti. Questo è compiuto utilizzando algoritmi di stima iterativi. Da notare che questa procedura non è necessaria per semplici modelli polinomiali del modulo  $Y = A + BX^{**2}$ . Definendo  $W = X^{**2}$ , otteniamo un modello lineare semplice,  $Y = A + BW$ , che può essere stimato utilizzando metodi tradizionali come la procedura di regressione lineare.

**Esempio.** La popolazione può essere prevista in base al tempo? Una trama scenica dimostra che sembra esserci un forte rapporto tra popolazione e tempo, ma la relazione è nonlineare, quindi richiede i metodi di stima speciali della procedura di Regressione Nonlineare. Impostando un'equazione appropriata, come un modello di crescita della popolazione logistica, possiamo ottenere una buona stima del modello, permettendoci di fare previsioni sulla popolazione per i tempi che non sono stati effettivamente misurati.

**Statistiche.** Per ogni iterazione: stime dei parametri e somma residua di quadrati. Per ogni modello: somma di quadrati per regressione, residuo, totale non corretto e totale corretto, stime dei parametri, errori standard asintotici e matrice di correlazione asintotica delle stime dei parametri.

**Nota:** La regressione non lineare vincolata utilizza gli algoritmi proposti e implementati in NPSOL<sup>®</sup> da Gill, Murray, Saunders e Wright per stimare i parametri del modello.

## considerazioni sui dati di regressione non lineare

**Dati.** Le variabili dipendenti ed indipendenti devono essere quantitative. Le variabili categoriali, quali la religione, la specializzazione o l'area di residenza, devono essere ricodificate in variabili binarie (dummy) o altri tipi di variabili di contrasto.

**Ipotesi.** I risultati sono validi solo se hai specificato una funzione che descrive con precisione il rapporto tra variabili dipendenti e indipendenti. Inoltre, la scelta dei buoni valori di partenza è molto importante. Anche se hai specificato la forma funzionale corretta del modello, se usi i valori di partenza poveri, il tuo modello potrebbe non riuscire a convergere o puoi ottenere una soluzione localmente ottimale piuttosto che una che sia globalmente ottimale.

**Procedure correlate.** Molti modelli che appaiono non lineari all'inizio possono essere trasformati in un modello lineare, che può essere analizzato utilizzando la procedura di regressione lineare. Se si è incerti quale deve essere il modello corretto, la procedura di Curva Stima può aiutare a identificare utili relazioni funzionali nei propri dati.

## Ottenimento di un'analisi di Regressione Nonlineare

1. Dai menu, scegliere:

**Analizzare > Regressione > Nonlinear ...**

2. Selezionare una variabile dipendente numerica dall'elenco delle variabili nel dataset attivo.

3. Per costruire un'espressione del modello, inserire l'espressione nel campo **Modello Expression** o incollare componenti (variabili, parametri, funzioni) in campo.

4. Identificare i parametri nel proprio modello facendo clic su **Parametri**.

Un modello segmentato (uno che prende forme diverse in diverse parti del suo dominio) deve essere specificato utilizzando la logica condizionale all'interno del singolo modello statement.

## Logica Condizionale (Regressione Non Lineare)

È possibile specificare un modello segmentato utilizzando la logica condizionale. Per utilizzare la logica condizionale all'interno di un'espressione di modello o di una funzione di perdita, si forma la somma di una serie di termini, uno per ogni condizione. Ogni termine consiste in un'espressione logica (tra parentesi) moltiplicata per l'espressione che dovrebbe risultare quando quell'espressione logica è vera.

Per esempio, considerare un modello segmentato che equivale a 0 per  $X \leq 0$ ,  $X$  per  $0 < X < 1$  e 1 per  $X \geq 1$ . L'espressione per questo è:

$$(X \leq 0) * 0 + (X > 0 \ \& \ X < 1) * X + (X \geq 1) * 1.$$

Le espressioni logiche tra parentesi tutte valutano a 1 (true) o 0 (false). Pertanto:

Se  $X \leq 0$ , il precedente si riduce a  $1 * 0 + 0 * X + 0 * 1 = 0$ .

Se  $0 < X < 1$  si riduce a  $0 * 0 + 1 * X + 0 * 1 = X$ .

Se  $X \geq 1$ , si riduce a  $0 * 0 + 0 * X + 1 * 1 = 1$ .

Esempi più complicati possono essere facilmente costruiti sostituendo espressioni logiche diverse e espressioni di risultato. Ricordate che le doppie disuguaglianze, come  $0 < X < 1$ , devono essere scritte come espressioni composti, come  $(X > 0 \ \& \ X < 1)$ .

Le variabili di stringa possono essere utilizzate all'interno delle espressioni logiche:

`(city= 'New York') * costliv + (city= 'Des Moines')*0.59*costliv`

Questo rese una espressione (il valore della variabile *costliv*) per i newyorkesi e un altro (59% di quel valore) per Des Moines residenti. Le costanti di stringa devono essere racchiuse tra virgolette o apostrofi, come mostrato qui.

## Parametri Di Regressione Non Lineare

I parametri sono le parti del tuo modello che le stime della procedura Nonlinear Regression. I parametri possono essere costanti additivi, coefficienti moltiplicativi, esponenti o valori utilizzati nella valutazione delle funzioni. Tutti i parametri definiti verranno visualizzati (con i relativi valori iniziali) nella lista dei Parametri nella finestra di dialogo principale.

**Nome.** È necessario specificare un nome per ogni parametro. Questo nome deve essere un nome variabile valido e deve essere il nome utilizzato nell'espressione del modello nella finestra di dialogo principale.

**Valore iniziale.** Consente di specificare un valore di partenza per il parametro, preferibilmente il più vicino possibile alla soluzione finale prevista. I valori di partenza poveri possono determinare un mancato convergere o in convergenza su una soluzione che sia locale (piuttosto che globale) o sia fisicamente impossibile.

**Utilizza i valori di partenza dall'analisi precedente.** Se già si esegue una regressione non lineare da questa finestra di dialogo, è possibile selezionare questa opzione per ottenere i valori iniziali dei parametri dai loro valori nella precedente esecuzione. Questo permette di continuare alla ricerca quando l'algoritmo sta convergendo lentamente. (I valori iniziali iniziali appariranno ancora nella lista dei parametri nella finestra di dialogo principale).

*Nota:* questa selezione persiste in questa finestra di dialogo per il resto della sessione. Se si modifica il modello, assicurarsi di deselectionarlo.

## Modelli Comuni Di Regressione Non Lineare

La tabella sottostante fornisce la sintassi del modello di esempio per molti modelli di regressione non lineari pubblicati. Un modello selezionato a caso non è probabilmente adatto ai tuoi dati. I valori di partenza appropriati per i parametri sono necessari e alcuni modelli richiedono vincoli per convergere.

*Tabella 1. Sintassi del modello di esempio*

Nome	Espressione del modello
Regressione Asintotica	$b1 + b2 * \exp(b3 * x)$
Regressione Asintotica	$b1 - (b2 * (b3 ** x))$
Densità	$(b1 + b2 * x) ** (-1 / b3)$
Gauss	$b1 * (1 - b3 * \exp(-b2 * x ** 2))$
Gompertz	$b1 * \exp(-b2 * \exp(-b3 * x))$
Johnson - Schumacher	$b1 * \exp(-b2 / (x + b3))$
Log - Modificato	$(b1 + b3 * x) ** b2$
Logaritmica logistica	$b1 - \ln(1 + b2 * \exp(-b3 * x))$
Legge Metcherlich di Diminuzione Ritorni	$b1 + b2 * \exp(-b3 * x)$
Michaelis Menten	$b1 * x / (x + b2)$

Tabella 1. Sintassi del modello di esempio (Continua)

Nome	Espressione del modello
Morgan - Mercer - Florin	$(b1 * b2 + b3 * x^{**} b4) / (b2 + x^{**} b4)$
Peal - Reed	$b1 / (1 + b2 * \exp(- (b3 * x + b4 * x^{**} 2 + b5 * x^{**} 3)))$
Rapporto di Cubics	$(b1 + b2 * x + b3 * x^{**} 2 + b4 * x^{**} 3) / (b5 * x^{**} 3)$
Rapporto di Quadratalità	$(b1 + b2 * x + b3 * x^{**} 2) / (b4 * x^{**} 2)$
Richards	$b1 / ((1 + b3 * \exp(-b2 * x))^{**} (1 / b4))$
Verhulst	$b1 / (1 + b3 * \exp(-b2 * x))$
Von Bertalanffy	$(b1^{**} (1 - b4) - b2 * \exp(-b3 * x))^{**} (1 / (1 - b4))$
Weibull	$b1 - b2 * \exp(-b3 * x^{**} b4)$
Densità Resa	$(b1 + b2 * x + b3 * x^{**} 2)^{**} (-1)$

## Funzione Perdita di regressione non lineare

La **funzione di perdita** nella regressione non lineare è la funzione che viene minimizzata dall'algorithm. Selezionare **Sum dei residui quadrati** per ridurre al minimo la somma dei residui quadrati o **funzione di perdita definita dall'utente** per ridurre al minimo una diversa funzione.

Se si seleziona **Funzione di perdita definita dall'utente**, è necessario definire la funzione di perdita la cui somma (in tutti i casi) deve essere minimizzata dalla scelta dei valori dei parametri.

- La maggior parte delle funzioni di perdita comporta la variabile speciale *RESID\_*, che rappresenta il residuo. (La funzione di perdita della somma dei residui quadrati può essere immessa esplicitamente come *RESID\_\*\*2*.) Se è necessario utilizzare il valore previsto nella funzione di perdita, è uguale alla variabile dipendente meno il residuo.
- È possibile specificare una funzione di perdita condizionale utilizzando la logica condizionale.

È possibile digitare un'espressione nel campo funzione di perdita definita dall'Utente o incollare componenti dell'espressione in campo. Le costanti di stringa devono essere racchiuse tra virgolette o apostrofi e le costanti numeriche devono essere digitate in formato americano, con il dot come delimitatore decimale.

## Vincoli Parametri di regressione non lineare

Un **vincolo** è una limitazione sui valori consentiti per un parametro durante la ricerca iterativa per una soluzione. Le espressioni lineari vengono valutate prima di un passo, quindi è possibile utilizzare dei vincoli lineari per impedire passaggi che potrebbero risultare in overflow. Le espressioni non lineari vengono valutate dopo che si fa un passo.

Ogni equazione o disuguaglianza richiede i seguenti elementi:

- Un'espressione che coinvolge almeno un parametro nel modello. Digitare l'espressione o utilizzare il tastierino, che consente di incollare numeri, operatori o parentesi nell'espressione. È possibile digitare nei parametri richiesti insieme al resto dell'espressione o incollare dalla lista dei parametri a sinistra. Non è possibile utilizzare variabili ordinarie in un vincolo.
- Uno dei tre operatori logici  $< =, =, > =$ .
- Una costante numerica, a cui viene confrontata l'espressione utilizzando l'operatore logico. Digitare la costante. Le costanti numeriche devono essere digitate in formato americano, con il dot come delimitatore decimale.



## Regressione Non Lineare Save New Variabili

È possibile salvare un numero di nuove variabili nel file dei dati attivi. Le opzioni disponibili sono Residui, Valori previsti, Derivati e valori di funzione Loss. Queste variabili possono essere utilizzate nelle successive analisi per testare l'adattamento del modello o per identificare i casi di problema.

- *Residui*. Salva i residui con il nome di variabile resid.
- *Valori previsti*. Salva i valori previsti con il nome variabile pred\_.
- *Derivate*. Una derivata viene salvata per ciascun parametro del modello. I nomi derivata sono creati aggiungendo il prefisso 'd.' per i primi sei caratteri del nome del parametro.
- *Valori della funzione di perdita*. Questa opzione è disponibile se si specifica la propria funzione di perdita. La variabile name loss\_ è assegnata ai valori della funzione di perdita.

## Opzioni Di Regressione Non Lineare

Le opzioni consentono di controllare vari aspetti della tua analisi di regressione non lineare:

*Stime di bootstrap*. Un metodo di stima dell'errore standard di una statistica utilizzando campioni ripetuti dell'insieme di dati originale. Questo risultato viene ottenuto campionando (con sostituzione) per ottenere molti campioni della stessa dimensione del dataset originale. L'equazione non lineare è stimata per ciascuno di questi campioni. L'errore standard di ogni stima del parametro viene quindi calcolato come la deviazione standard delle stime di bootstrap. I valori di parametro dei dati originali vengono usati come valori di partenza per ogni campione di bootstrap. Ciò richiede l'algoritmo di programmazione quadratica sequenziale.

**Metodo di stima**. Consente di selezionare un metodo di stima, se possibile. (Alcune scelte in questa o altre finestre di dialogo richiedono l'algoritmo di programmazione quadratica sequenziale.) Le alternative disponibili includono la programmazione quadratica Sequenziale e Levenberg-Marquardt.

- *Programmazione quadratica sequenziale*. Questo metodo è disponibile per modelli vincolati e non vincolati. Viene proposto come metodo predefinito per i modelli vincolati, in presenza di una funzione di perdita personalizzata o se si scelgono le stime bootstrap. È possibile immettere nuovi valori per Massimo numero di iterazioni e Limite di fase ed è possibile modificare la selezione negli elenchi a discesa Tolleranza di ottimalità, Precisione della funzione e Dimensione fase infinita.
- *Levenberg-Marquardt*. Questo è l'algoritmo predefinito per i modelli non vincolati. Questo metodo non è disponibile per i modelli se è stato specificato un modello vincolato, una funzione di perdita personalizzata o una stima bootstrap. È possibile immettere nuovi valori per Massimo numero di iterazioni ed è possibile modificare la selezione negli elenchi a discesa Convergenza per la somma dei quadrati e Convergenza parametri.

## Interpretazione Risultati Regression Non Lineari

I problemi di regressione non lineari spesso presentano difficoltà computazionali:

- La scelta dei valori iniziali per i parametri influenza la convergenza. Cercate di scegliere valori iniziali ragionevoli e, se possibile, vicini alla soluzione finale prevista.
- A volte un algoritmo esegue meglio dell'altro su un problema particolare. Nella finestra di dialogo Opzioni, selezionare l'altro algoritmo se disponibile. (Se si specifica una funzione di perdita o determinati tipi di vincoli non è possibile utilizzare l'algoritmo di Levenberg-Marquardt).
- Quando l'iterazione si interrompe solo perché si è verificato il numero massimo di iterazioni, il modello "finale" non è probabilmente una buona soluzione. Selezionare **Utilizza i valori di partenza dall'analisi precedente** nella finestra di dialogo Parametri per continuare l'iterazione o, meglio ancora, scegliere diversi valori iniziali.
- I modelli che richiedono l'esponenziamento di o da grandi valori dati possono causare overflow o underflow (numeri troppo grandi o troppo piccoli per il computer da rappresentare). A volte si possono evitare queste da una scelta adeguata dei valori iniziali o da vincoli imponenti sui parametri.

## Funzioni Aggiuntive di comando NLR

Il linguaggio della sintassi dei comandi consente inoltre di:

- Denominare un file da cui leggere i valori iniziali per le stime dei parametri.
- Specificare più di una istruzione modello e della funzione di perdita. Questo rende più facile specificare un modello segmentato.
- Fornire i propri derivati piuttosto che utilizzare quelli calcolati dal programma.
- Specificare il numero di campioni di bootstrap da generare.
- Specificare ulteriori criteri di iterazione, inclusa l'impostazione di un valore critico per la verifica dei derivati e la definizione di un criterio di convergenza per la correlazione tra i residui e i derivati.

Ulteriori criteri per il comando CNLR (regressione non lineare) consentono di:

- Specificare il numero massimo di iterazioni minori consentite all'interno di ogni iterazione maggiore.
- Impostare un valore critico per il controllo derivato.
- Impostare un limite di passo.
- Specificare una tolleranza di arresto per determinare se i valori iniziali sono all'interno dei loro limiti specificati.

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

## Stima del peso

---

I modelli di regressione lineare standard presumono che la varianza sia costante nell'ambito della popolazione oggetto di studio. Quando questo non è il caso (ad esempio, quando i casi che sono alti su qualche attributo mostrano più variabilità rispetto ai casi che sono bassi su quell'attributo) la regressione lineare utilizzando i minimi quadrati ordinari (OLS) non fornisce più stime di modello ottimali. Se le differenze in variabilità possono essere previste da un'altra variabile, la procedura Stime del peso può calcolare i coefficienti di un modello di regressione lineare utilizzando i minimi quadrati pesati (WLS, Weighted Least Squares) per far sì che alle osservazioni più precise (ovvero le osservazioni con meno variabilità) sia assegnato un peso maggiore per determinare i coefficienti di regressione. La procedura Stima del peso verifica una serie di trasformazioni del peso e indica quale sia maggiormente adatta per i dati.

**Esempio.** Quali sono gli effetti dell'inflazione e della disoccupazione sulle variazioni dei prezzi azionari? Poiché le scorte con valori di condivisione più elevati spesso mostrano una maggiore variabilità rispetto a quelle con valori di condivisione bassi, i minimi quadrati ordinari non produrranno stime ottimali. La stima del peso consente di rendere conto dell'effetto del prezzo di condivisione sulla variabilità delle variazioni di prezzo nel calcolo del modello lineare.

**Statistiche.** Valori di verosimiglianza di verosimiglianza per ogni potenza della variabile di origine peso collaudata, più  $R$ ,  $R$ -quadrato, regolato  $R$ -quadrato, tabella ANOVA per modello WLS, stime dei parametri non standardizzate e standardizzate, e logaritmo-verosimiglianza per il modello WLS.

### considerazioni sui dati di stima del peso

**Dati.** Le variabili dipendenti ed indipendenti devono essere quantitative. Le variabili categoriali, quali la religione, la specializzazione o l'area di residenza, devono essere ricodificate in variabili binarie (dummy) o altri tipi di variabili di contrasto. La variabile del peso deve essere quantitativa e correlata alla variabilità nella variabile dipendente.

**Ipotesi.** Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La relazione tra la variabile dipendente e ogni variabile indipendente deve essere lineare e tutte le osservazioni devono essere indipendenti. La varianza della variabile dipendente può variare tra i livelli delle variabili indipendenti, ma le differenze devono essere prevedibili in base alla variabile peso.

**Procedure correlate.** La Procedura D'Esplorazione può essere utilizzata per schermare i tuoi dati. Esplora fornisce test per la normalità e l'omogeneità della varianza, nonché i display grafici. Se la tua variabile

dipendente sembra avere uguale varianza su livelli di variabili indipendenti, è possibile utilizzare la procedura di Regressione Lineare. Se i tuoi dati sembrano violare un'assunzione (come la normalità), prova a trasformarle. Se i tuoi dati non sono correlati linearmente e una trasformazione non aiuta, utilizzare un modello alternativo nella procedura di stima della curva. Se la tua variabile dipendente è dicotomica (ad esempio, se una particolare vendita viene completata o se un articolo è difettoso) utilizzare la procedura di Regressione Logistica. Se la tua variabile dipendente è censurata (ad esempio, tempo di sopravvivenza dopo l'intervento) usare Life Tables, Kaplan-Meier o Cox Regression, disponibile nelle Tabelle personalizzate e statistiche avanzate. Se i tuoi dati non sono indipendenti (ad esempio, se osservi la stessa persona in diverse condizioni) utilizzare la procedura Repeated Misure, disponibile nelle Tabelle personalizzate e Statistiche avanzate.

## Ottenimento di un'analisi di stima del peso

1. Dai menu, scegliere:

**Analizzare > Regressione > Stima del peso ...**

2. Selezionare una variabile dipendente.

3. Selezionare una o più variabili indipendenti.

4. Selezionare la variabile che è la fonte di eteroscedasticità come variabile peso.

### **Variabile di ponderazione**

I dati vengono pesati dal reciproco di questa variabile elevato a una potenza. L'equazione di regressione è calcolata per ognuno dei valori di potenza di uno specifico intervallo e indica la potenza che massimizza la funzione del logaritmo della verosimiglianza.

### **Intervallo di potenza**

Viene utilizzato insieme alla variabile di peso per calcolare i pesi. Verrà calcolata un'equazione di regressione per ciascuna potenza inclusa nell'intervallo specificato. L'intervallo può variare al massimo fra -6,5 e +7,5, compresi. L'intervallo di valori di potenza dal valore basso al valore alto, in incrementi determinati dal valore specificato. Il numero totale di valori nell'intervallo di potenza è limitato a 150.

## Opzioni Di Stima Del Peso

È possibile specificare le opzioni per l'analisi della stima del peso:

**Salva miglior peso come nuova variabile.** Aggiunge la variabile di peso al file attivo. Questa variabile si chiama *WGT\_n*, dove *n* è un numero scelto per dare alla variabile un nome univoco.

**Visualizza ANOVA e stime.** Consente di controllare come vengono visualizzate le statistiche nell'output. Le alternative disponibili sono Per la potenza migliore e per ogni valore di potenza.

## Funzioni Aggiuntive Comando WLS

Il linguaggio della sintassi dei comandi consente inoltre di:

- Fornire un valore unico per la potenza.
- Specificare un elenco di valori di potenza, oppure mescolare una gamma di valori con un elenco di valori per la potenza.

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

## Regressione minimi quadrati a due fasi

I modelli di regressione lineari standard presumono che gli errori nella variabile dipendente non siano correlati alle variabili indipendenti. Quando questo non è il caso (ad esempio, quando le relazioni tra le variabili sono bidirezionali), la regressione lineare che utilizza OLS (Ordinary Least Squares) non fornisce più stime ottimali del modello. La regressione minimi quadrati a due fasi utilizza le variabili strumentali che non sono correlate ai termini di errore per calcolare i valori stimati dei predittori problematici (il primo stadio) e, quindi, utilizza tali valori calcolati per stimare un modello di regressione lineare della variabile

dipendente (il secondo stadio). Poiché i valori calcolati sono basati sulle variabili non correlati agli errori, i risultati del modello a due stadi sono ottimali.

**Esempio.** La domanda di una merce è legata al suo prezzo e ai redditi dei consumatori? La difficoltà di questo modello è che il prezzo e la domanda hanno un effetto reciproco l'uno sull'altro. Ovvero, il prezzo può influenzare la domanda e la domanda può anche influenzare il prezzo. Un modello di regressione meno squadrato a due stadi potrebbe utilizzare i redditi dei consumatori e il prezzo marcato per calcolare un proxy per il prezzo non correlato con gli errori di misurazione della domanda. Questo proxy viene sostituito per il prezzo stesso nel modello originariamente specificato, che viene poi stimato.

**Statistiche.** Per ogni modello: coefficienti di regressione standardizzati e non standardizzati, più  $R$ ,  $R^2$ , rettificato  $R^2$ , errore standard della tabella di stima, analisi della varianza, valori previsti, e residui. Inoltre, intervalli di confidenza di 95% per ogni coefficiente di regressione, e matrici di correlazione e covarianza delle stime dei parametri.

## Due - Stage Meno - Squares Regression dati considerazioni

**Dati.** Le variabili dipendenti ed indipendenti devono essere quantitative. Le variabili categoriali, quali la religione, la specializzazione o l'area di residenza, devono essere ricodificate in variabili binarie (dummy) o altri tipi di variabili di contrasto. *Endogene* variabili esplicative dovrebbero essere quantitative (non categoriali).

**Ipotesi.** Per ciascun valore della variabile indipendente, la distribuzione della variabile dipendente deve essere normale. La varianza della distribuzione della variabile dipendente deve essere costante per tutti i valori della variabile indipendente. Il rapporto tra la variabile dipendente e ogni variabile indipendente dovrebbe essere lineare.

**Procedure correlate.** Se si ritiene che nessuna delle variabili predittore sia correlata con gli errori nella variabile dipendente, è possibile utilizzare la procedura di regressione lineare. Se i tuoi dati sembrano violare una delle ipotesi (come la normalità o la varianza costante), provare a trasformarle. Se i tuoi dati non sono correlati linearmente e una trasformazione non aiuta, utilizzare un modello alternativo nella procedura di stima della curva. Se la tua variabile dipendente è dicotomica, ad esempio se una particolare vendita è completata o meno, utilizzare la procedura di Regressione Logistica. Se i tuoi dati non sono indipendenti - ad esempio, se osservi la stessa persona in diverse condizioni - usa la procedura Repeated Measures.

## Ottenimento di un'analisi di regressione di Two - Stage Least - Squares Regression

1. Dai menu, scegliere:

**Analizza > Regressione > Minimi quadrati a 2 stadi ...**

2. Selezionare una variabile dipendente.

3. Selezionare una o più variabili esplicative (predittore).

4. Selezionare una o più variabili strumentali.

- *Strumentali.* Sono le variabili utilizzate per calcolare i valori previsti per le variabili endogene nel primo stadio dell'analisi dei minimi quadrati a due stadi. Le stesse variabili possono figurare sia nella casella di elenco Esplicative che in quella Strumentali. Il numero di variabili strumentali deve essere almeno uguale a quello di variabili esplicative. Se tutte le variabili strumentali ed esplicative elencate sono le stesse, i risultati dell'analisi saranno gli stessi della procedura di regressione lineare.

Le variabili esplicative non specificate come strumentali sono considerate endogene. Normalmente tutte le variabili esogene dell'elenco esplicativo sono specificate anche come variabili strumentali.

## Opzioni Di Regressione di Due Stage - Quadrati

È possibile selezionare le seguenti opzioni per l'analisi:

**Salvare nuove variabili.** Consente di aggiungere nuove variabili al proprio file attivo. Le opzioni disponibili sono Previste e Residuali.

**Visualizza la covarianza dei parametri.** Consente di stampare la matrice di covarianza delle stime dei parametri.

## 2SLS Funzioni aggiuntive di comando

Il linguaggio di sintassi dei comandi consente anche di stimare simultaneamente più equazioni. Vedi il *Riferimento Syntax Reference* per informazioni complete sulla sintassi.

## 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

**contrasti 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 REGRESSION e COXREG:

```
/CONTRAST(TREATMNT)=SPECIAL( 1 1 1 1
                             3 -1 -1 -1
                             0 2 -1 -1
                             0 0 1 -1 )
```

Per LOGLINEAR è necessario specificare:

```
/CONTRAST(TREATMNT)=BASIS SPECIAL( 1 1 1 1
                                    3 -1 -1 -1
                                    0 2 -1 -1
                                    0 0 1 -1 )
```

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.

## 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 predittive 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

#### Statistica

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 iperparametri 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 30. 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 30.

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** .

## Modelli Temporali 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)**

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** .

## Parametric Accelerated Failure Time Models: Criteria

### Criteria

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 intero appartenente a [ 1, 100]. L'impostazione predefinita è 20.

### **Numero massimo di dimezzamenti**

Specifica il numero massimo di dimezzamento. Deve essere un intero 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 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.

## Parametric Accelerated Failure Time Models: Plot

### Grafico

#### Grafico dei residui di Cox-Snell

Selezionare **Visualizza la trama** per creare una trama residua 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.

##### Immettere

##### 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.

## Parametric Accelerated Failure Time Models: Export

**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'.

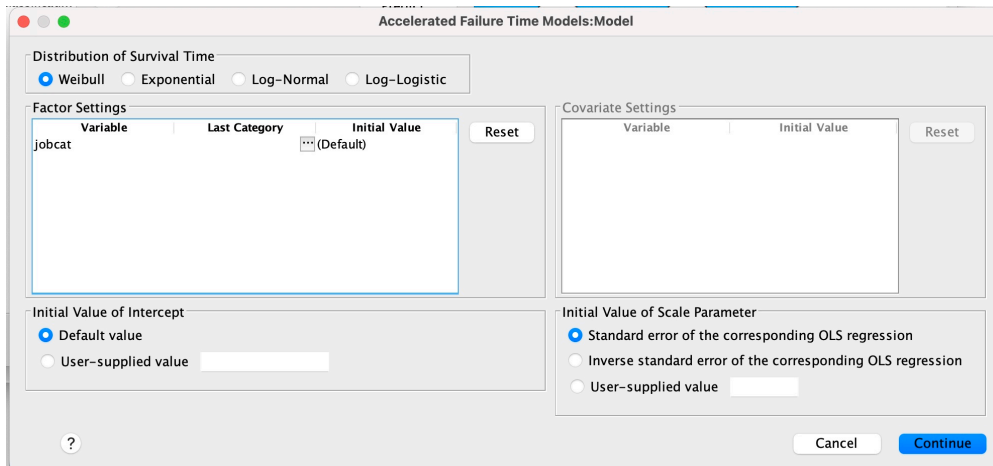


Figura 1. Modelli di durata accelerata - Finestra di dialogo - Categoria

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

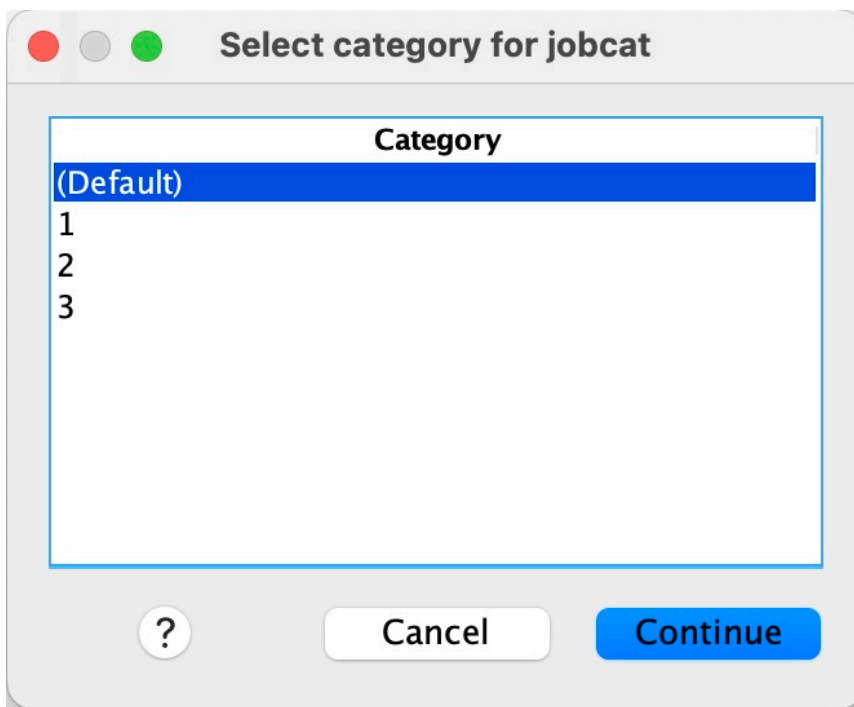


Figura 2. Modelli di durata accelerata - Finestra di dialogo - Seleziona categoria

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**

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.

## **Immettere**

### **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 ...**

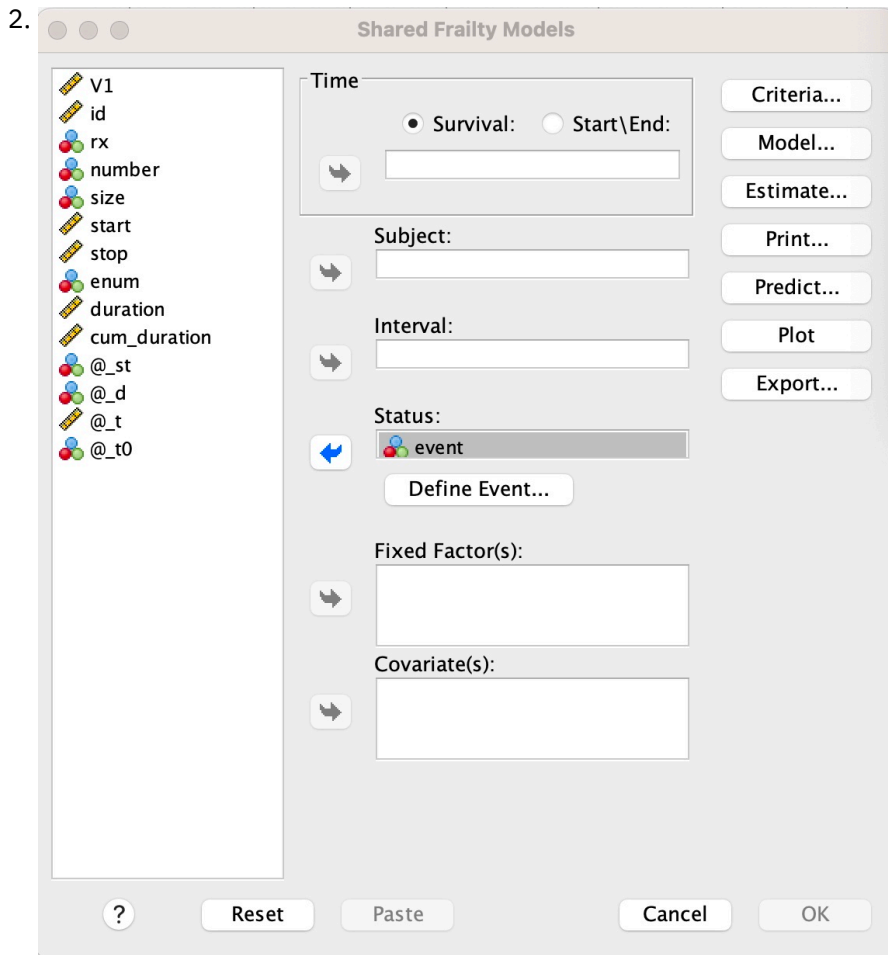


Figura 3. Modelli di fragilità condivisi - finestra di dialogo - Stato

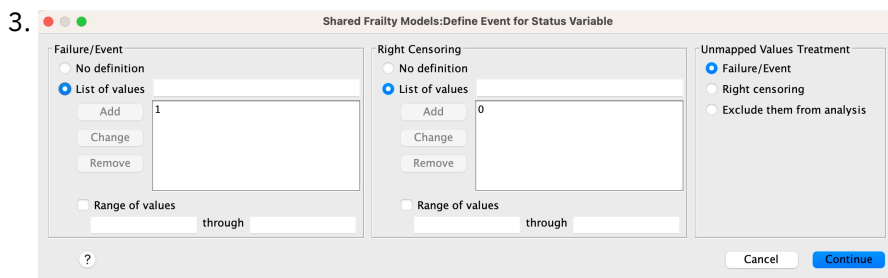


Figura 4. Modelli di fragilità condivisi - finestra di dialogo - Stato - Definisci evento

## 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 FRAILTY=GAMMA DISTRIBUTION=WEIBULL
```

```
/STATUS VARIABLE=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 VARIABLE=infect 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 univoco.

- endTime: tempo di sopravvivenza (in giorni) dell'effetto indesiderato, dopo un trattamento, misurato dall'inizio di un trattamento a un effetto indesiderato segnalato o censurato entro 60 giorni.
- sideEffect: stato dell'effetto collaterale, stato = 0 se censurato e stato = 1 se viene riportato l'effetto a lato 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 FRAILITY=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.

## Informazioni particolari

---

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

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

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

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

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

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

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

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

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

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

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

*IBM Director of Licensing  
IBM Corporation  
North Castle Drive, MD-NC119*

Armonk, NY 10504-1785

US

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

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

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

Le informazioni relative a prodotti non IBM sono ottenute dai fornitori di quei prodotti, dagli annunci pubblicati 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 di regressione logistica [2](#)  
analisi di sopravvivenza  
in Kernel Ridge Regression [28](#)

Analisi probit  
criteri [11](#)  
Definisci intervallo [11](#)  
funzioni aggiuntive del comando [11](#)  
Intervalli di confidenza fiduciari [11](#)  
iterazioni [11](#)  
Potenza mediana relativa [11](#)  
statistiche [11](#)  
Tasso di risposta naturale [11](#)  
test di parallelismo [11](#)

## B

bontà di adattamento  
in Regressione Logistica Multinomiale [7](#)

## C

categoria di riferimento  
in Regressione Logistica Multinomiale [7](#)  
celle con zero osservazioni  
in Regressione Logistica Multinomiale [8](#)  
Chi-quadrato di Pearson  
bontà di adattamento [7](#)  
per stimare il valore di scaling di dispersione [8](#)  
classificazione  
in Regressione Logistica Multinomiale [6](#)  
contrasti  
in regressione logistica [3](#)  
Cook's D  
in regressione logistica [4](#)  
covariate  
in regressione logistica [3](#)  
Covariate categoriali [3](#)  
covariate di stringa  
in regressione logistica [3](#)  
Cox e Snell R - quadrato  
in Regressione Logistica Multinomiale [7](#)  
criterio di convergenza  
in Regressione Logistica Multinomiale [8](#)  
cronologia delle iterazioni  
in Regressione Logistica Multinomiale [8](#)

## D

delta  
come correzione per le celle con zero osservazioni [8](#)  
DfBeta  
in regressione logistica [4](#)  
dimezzamenti  
in Regressione Logistica Multinomiale [8](#)

## E

eliminazione all'indietro  
in regressione logistica [3](#)

## F

funzione di devianza  
per stimare il valore di scaling di dispersione [8](#)

## I

intercettazione  
includere o escludere [6](#)  
interruzione del rapporto di lavoro  
in Regressione Logistica Multinomiale [8](#)  
intervalli di confidenza  
in Regressione Logistica Multinomiale [7](#)  
Intervalli di confidenza fiduciari  
in Analisi Probit [11](#)  
iterazioni  
in Analisi Probit [11](#)  
in regressione logistica [5](#)  
in Regressione Logistica Multinomiale [8](#)

## K

Kernel Ridge  
alfa [28](#)  
coef0 [28](#)  
gamma [28](#)  
laurea [28](#)  
Modello singolo [28](#)  
selezione modello [28](#)  
KRR (Kernel Ridge Regression)  
grafici [30](#)  
parametri [30](#)  
parametri della griglia [30](#)  
pieghe incrociate [30](#)  
salvare [30](#)  
visualizzazione [30](#)

## L

Legge Metcherlich dei rendimenti diminutivi  
in Regression non lineare [19](#)

## M

matrice di correlazione  
in Regressione Logistica Multinomiale [7](#)  
matrice di covarianza  
in Regressione Logistica Multinomiale [7](#)  
modelli di effetti principali  
in Regressione Logistica Multinomiale [6](#)  
Modelli di fragilità condivisa parametrici

## Modelli di fragilità condivisa parametrici (*Continua*)

- analisi [37](#)
- criteri [38](#)
- esportazione [43](#)
- modello [39](#)
- predire [41](#)
- stampare [41](#)
- stima [40](#)
- tracciare [42](#)

## Modelli di fragilità parametrica

- variabili di stato di sopravvivenza [43](#)

## modelli fattoriali completi

- in Regressione Logistica Multinomiale [6](#)

## modelli non lineari

- in Regression non lineare [19](#)

## modelli personalizzati

- in Regressione Logistica Multinomiale [6](#)

## Modelli Temporali Accelerati Parametrici accelerati

- analisi [31](#)
- criteri [32](#)
- esportazione [36](#)
- modello [32](#)
- predire [35](#)
- stampare [34](#)
- stima [33](#)
- tracciare [35](#)

## modello di densità

- in Regression non lineare [19](#)

## modello di densità delle

- in Regression non lineare [19](#)

## Modello di Gauss

- in Regression non lineare [19](#)

## modello di registrazione log

- in Regression non lineare [19](#)

## Modello Gompertz

- in Regression non lineare [19](#)

## Modello Johnson - Schumacher

- in Regression non lineare [19](#)

## Modello Michaelis Menten

- in Regression non lineare [19](#)

## Modello Peal - Reed

- in Regression non lineare [19](#)

## Modello Richards

- in Regression non lineare [19](#)

## Modello Verhulst

- in Regression non lineare [19](#)

## Modello von Bertalanffy

- in Regression non lineare [19](#)

## Modello Weibull

- in Regression non lineare [19](#)

## Morgan - Mercer - Florin modello

- in Regression non lineare [19](#)

## P

### Potenza mediana relativa

- in Analisi Probit [11](#)

## R

### R-quadrato di McFadden

- in Regressione Logistica Multinomiale [7](#)

### R-quadrato di Nagelkerke

## R-quadrato di Nagelkerke (*Continua*)

- in Regressione Logistica Multinomiale [7](#)

## rapporto del modello quadrato

- in Regression non lineare [19](#)

## rapporto di modello cubico

- in Regression non lineare [19](#)

## rapporto di verosimiglianza

- bontà di adattamento [7](#)

- per stimare il valore di scaling di dispersione [8](#)

## regressione asintotica

- in Regression non lineare [19](#)

## Regressione lineare

- Regressione minimi quadrati a due fasi [23](#)

- stima del peso [22](#)

## regressione logistica

## Regressione logistica

- binario [1](#)

- coefficienti [2](#)

- contrasti [3](#)

- Covariate categoriali [3](#)

- covariate di stringa [3](#)

- Definisci regola di selezione [3](#)

- esempio [2](#)

- funzioni aggiuntive del comando [5](#)

- Imposta regola [3](#)

- iterazioni [5](#)

- metodi di selezione delle variabili [3](#)

- misure di influenza [4](#)

- opzioni di visualizzazione [5](#)

- Probabilità per stepwise [5](#)

- residui [4](#)

- salvataggio di nuove variabili [4](#)

- statistiche [2](#)

- statistiche e grafici [5](#)

- termine costante [5](#)

- Test di bontà dell'adattamento di Hosmer-Lemeshow [5](#)

- Valore di interruzione di classificazione [5](#)

- valori attesi [4](#)

## Regressione logistica binaria [1, 2](#)

## Regressione logistica multinomiale

- categoria di riferimento [7](#)

- criteri [8](#)

- esportazione di informazioni dei modelli [9](#)

- funzioni aggiuntive del comando [9](#)

- modelli [6](#)

- salvare [9](#)

- statistiche [7](#)

## Regressione logistica SPSS [2](#)

## Regressione minimi quadrati a due fasi

- covarianza dei parametri [24](#)

- esempio [23](#)

- funzioni aggiuntive del comando [25](#)

- salvataggio di nuove variabili [24](#)

- statistiche [23](#)

- Variabili strumentali [23](#)

## Regressione non lineare

- Algoritmo Levenberg-Marquardt [21](#)

- Derivate [21](#)

- esempio [17](#)

- funzione di perdita [20](#)

- funzioni aggiuntive del comando [22](#)

- interpretazione dei risultati [21](#)

- logica condizionale [18](#)

- metodi di stima [21](#)

Regressione non lineare (*Continua*)  
modelli non lineari comuni [19](#)  
modello segmentato [18](#)  
parametri [19](#)  
Programmazione quadratica sequenziale [21](#)  
residui [21](#)  
Salva nuove variabili [21](#)  
statistiche [17](#)  
Stime bootstrap [21](#)  
valori attesi [21](#)  
Valori di avvio [19](#)  
vincoli di parametro [20](#)

Regressione Probit

esempio [10](#)  
statistiche [10](#)

Regressione quantile

criteri [13](#)  
esempio [12](#)  
esportazione [17](#)  
modello [14](#)  
salvare [16](#)  
statistiche [12](#)  
visualizzazione [15](#)

regressione vincolata

in Regression non lineare [20](#)

## S

selezione in avanti

in regressione logistica [3](#)

selezione stepwise

in regressione logistica [3](#)  
in Regressione Logistica Multinomiale [6](#)

singularità

in Regressione Logistica Multinomiale [8](#)

Sopravvivenza AFT

Finestra di dialogo di sopravvivenza - Variabili categoria  
[36](#)

Stima del peso

cronologia delle iterazioni [23](#)  
esempio [22](#)  
funzioni aggiuntive del comando [23](#)  
salvare i pesi migliori come nuova variabile [23](#)  
statistiche [22](#)  
verosimiglianza logaritmica [22](#)  
Visualizza ANOVA e stime [23](#)

stime del parametro

in Regressione Logistica Multinomiale [7](#)

## T

tabelle di classificazione

in Regressione Logistica Multinomiale [7](#)

tabelle di probabilità delle celle

in Regressione Logistica Multinomiale [7](#)

Tabella vita

variabili di stato di sopravvivenza [36](#)

termine costante

in regressione lineare [5](#)

Test di bontà dell'adattamento di Hosmer-Lemeshow

in regressione logistica [5](#)

test di parallelismo

in Analisi Probit [11](#)

## V

valore scaling di dispersione

in Regressione Logistica Multinomiale [8](#)

valori di leva

in regressione logistica [4](#)

verosimiglianza logaritmica

in Peso Stima [22](#)

in Regressione Logistica Multinomiale [7](#)

vincoli di parametro

in Regression non lineare [20](#)





