

*IBM SPSS Missing Values 30*



**Nota**

Prima di utilizzare queste informazioni e il prodotto che supportano, leggere le informazioni in [“Informazioni particolari” a pagina 19](#).

**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. Valori mancanti.....</b>	<b>1</b>
Introduzione ai Valori mancanti.....	1
Analisi valori mancanti.....	1
Visualizzazione Modelli dei Valori mancanti.....	3
Visualizzazione Statistiche Descrittive per i Valori mancanti.....	4
Stima delle statistiche e Impatto dei valori mancanti.....	4
Funzioni Aggiuntive di comando MVA.....	7
Assegnazione multipla.....	7
Analizzare Modelli.....	8
Valori dati mancanti.....	9
MULTIPLE IMPUTAZIONE Comando Funzioni aggiuntive.....	13
Gestione dei dati di Impatto Multiplo.....	13
Analisi Dei Dati Di Imputazione Multipla.....	14
Opzioni Di Imputazione Multipla.....	17
<b>Informazioni particolari.....</b>	<b>19</b>
Marchi.....	20
<b>Indice analitico.....</b>	<b>23</b>



---

# Capitolo 1. Valori mancanti

Le seguenti funzioni per i valori mancanti sono incluse in SPSS Statistiche Professional Edition o nell'opzione Valori mancanti.

---

## Introduzione ai Valori mancanti

I casi con valori mancanti costituiscono una grande sfida, in quanto le procedure di modellazione tipiche eliminano questi casi dall'analisi. Quando ci sono pochi valori mancanti (molto grossolanamente, meno del 5% del totale dei casi) e quei valori possono essere considerati dispersi a caso; cioè se manca un valore non dipende da altri valori, allora il metodo tipico della cancellazione listwise è relativamente "sicuro". L'opzione Valori mancanti può aiutarti a stabilire se la cancellazione listwise è sufficiente e fornisce metodi per la gestione dei valori mancanti quando non lo è.

Analisi del valore mancante versus Procedure di Imputazione multipla

L'opzione Valori mancanti fornisce due serie di procedure per la gestione dei valori mancanti:

- Le procedure [Multiple Imputation](#) forniscono l'analisi di modelli di dati mancanti, orientati verso un'eventuale imputazione multipla dei valori mancanti. Ovvero, vengono prodotte più versioni del dataset, ognuna contenente la propria serie di valori imputati. Quando vengono eseguite analisi statistiche, le stime dei parametri per tutti i datasets impatti sono pooled, fornendo preventivi generalmente più accurati di quanto non sarebbero con una sola imputazione.
- [Missing Value Analysis](#) fornisce una serie leggermente diversa di strumenti descrittivi per l'analisi dei dati mancanti (più particolarmente Little's MCAR test), e include una varietà di metodi di imputazione unici. Si noti che l'imputazione multipla è generalmente considerata superiore alla singola imputazione.

Attività dei Valori mancanti

È possibile iniziare con l'analisi dei valori mancanti seguendo questi passaggi fondamentali:

1. **Esamina la missibilità.** Utilizzare Missing Value Analysis e Analizzare Patterns per esplorare modelli di valori mancanti nei dati e determinare se è necessaria una maggiore imputazione.
2. **Assegna i valori mancanti.** Utilizza Valori di Dati mancanti per moltiplicare i valori mancanti.
3. **Analizzare i dati "completi".** Utilizzare qualsiasi procedura che supporti più dati di imputazione. Consultare ["Analisi Dei Dati Di Imputazione Multipla"](#) a pagina 14 per informazioni sull'analisi di più datasets di imputazione e un elenco di procedure che supportano questi dati.

---

## Analisi valori mancanti

La procedura di Analisi del valore mancante svolge tre funzioni primarie:

- Descrivere il modello dei dati mancanti. Dove si trovano i valori mancanti? Quanto sono estesi? Le coppie di variabili tendono ad avere valori mancanti in più casi? I valori dei dati sono estremi? I valori mancanti sono casuali?
- Stima le medie, le deviazioni standard, le covarianze e le correlazioni per metodi di valori mancanti differenti: listwise, pairwise, regressione o EM (expectation-maximization). Con il metodo pairwise vengono inoltre visualizzati i conteggi di casi completi pairwise.
- Compilare (assegnare) i valori mancanti mediante dei valori stimati utilizzando i metodi di regressione o EM; tuttavia, l'assegnazione multipla è solitamente ritenuta più attendibile dal punto di vista dell'accuratezza dei risultati.

L'analisi dei dati mancanti consente di risolvere diversi problemi dovuti alla presenza di dati incompleti. Se i casi con valori mancanti risultano sistematicamente diversi dai casi privi di valori mancanti, i risultati possono essere fuorvianti. Inoltre, i dati mancanti possono ridurre la precisione delle statistiche calcolate perché ci sono meno informazioni di quanto originariamente previsto. Un'altra preoccupazione è che

le ipotesi dietro molte procedure statistiche si basano su casi completi, e i valori mancanti possono complicare la teoria richiesta.

**Esempio.** Nella valutazione di un trattamento per la leucemia si misurano diverse variabili. Tuttavia non tutte le misurazioni sono disponibili per ogni paziente. Gli schemi dei dati mancanti vengono visualizzati, tabellati e trovati a caso. Un'analisi EM viene utilizzata per stimare i mezzi, le correlazioni e le covarianze. Viene utilizzato anche per determinare che i dati sono mancanti completamente a caso. I valori mancanti vengono poi sostituiti da valori imputati e salvati in un nuovo file di dati per ulteriori analisi.

**Statistiche.** Statistiche univariate, incluso numero di valori non mancanti, media, deviazione standard, numero di valori mancanti e numero di valori estremi. Mezzo stimato, matrice di covarianza e matrice di correlazione, utilizzando metodi listwise, pairwise, EM o regressione. Il test MCAR di Little con risultati EM. Riassunto dei mezzi con vari metodi. Per gruppi definiti da valori mancanti versus mancanti: *t* test. Per tutte le variabili: modelli di valore mancanti visualizzati casi - per - variabili.

Considerazioni sui dati

**Dati.** I dati possono essere categoriali o quantitativi (scala o continua). Tuttavia, è possibile stimare le statistiche e imputare i dati mancanti solo per le variabili quantitative. Per ogni variabile, i valori mancanti che non sono codificati come system - missing devono essere definiti dall'utente - mancante. Ad esempio, se un articolo del questionario ha la risposta *Non conoscere* codificato come 5 e si desidera trattarla come mancante, l'articolo dovrebbe avere 5 codificati come valore mancante dall'utente.

**Peso della frequenza.** I pesi di frequenza (replica) sono onorati da questa procedura. I casi con valore di peso di replica negativo o zero vengono ignorati. I pesi non interi vengono troncati.

**Ipotesi.** La stima in listino, pairwise e regressione dipende dal presupposto che il modello dei valori mancanti non dipende dai valori dei dati. (Questa condizione è nota come **mancante completamente a caso**, o MCAR.) Pertanto, tutti i metodi (compreso il metodo EM) per la stima danno stime coerenti e imparziali delle correlazioni e delle covarianze quando i dati sono MCAR. La violazione dell'assunzione di MCAR può portare a stime biased prodotte dai metodi di listino, pairwise e regressione. Se i dati non sono MCAR, è necessario utilizzare la stima EM.

La stima EM dipende dal presupposto che il modello dei dati mancanti sia correlato solo ai dati osservati. (Questa condizione è chiamata **mancante a caso** MAR.) Questo presupposto consente di aggiustare le stime utilizzando le informazioni disponibili. Ad esempio, in uno studio di istruzione e reddito, i soggetti con bassa istruzione possono avere valori di reddito più mancanti. In questo caso, i dati sono MAR, non MCAR. In altre parole, per MAR, la probabilità che il reddito sia registrato dipende dal livello di istruzione del soggetto. La probabilità può variare per istruzione ma non per reddito *all'interno di quel livello di istruzione*. Se la probabilità che il reddito sia registrato varia anche per il valore del reddito all'interno di ogni livello di istruzione (ad esempio, le persone con redditi elevati non le segnalano), allora i dati non sono né MCAR né MAR. Non si tratta di una situazione non comune e, se si applica, nessuno dei metodi è adeguato.

**Procedure correlate.** Molte procedure consentono di utilizzare la stima listwise o pairwise. Linear Regression and Factor Analysis consentono la sostituzione dei valori mancanti con i valori medi. Nel modulo aggiuntivo Forecasting sono disponibili diversi metodi per sostituire i valori mancanti nella serie temporale.

Per Ottenere L'Analisi Del Valore Mancante

1. Dai menu, scegliere:

**Analisi > Analisi del valore mancante ...**

2. Selezionare almeno una variabile (scala) quantitativa per stimare le statistiche e opzionalmente imputare i valori mancanti.

Facoltativamente, è possibile:

- Selezionare variabili categoriali (numeriche o stringa) ed inserire un limite sul numero di categorie (**Categorie massime**).
- Fare clic su **Patterns** per tabulare i modelli di dati mancanti. Per ulteriori informazioni, consultare l'argomento [“Visualizzazione Modelli dei Valori mancanti”](#) a pagina 3.

- Fare clic su **Descriptive** per visualizzare le statistiche descrittive dei valori mancanti. Per ulteriori informazioni, consultare l'argomento [“Visualizzazione Statistiche Descrittive per i Valori mancanti”](#) a pagina 4.
- Selezionare un metodo per stimare le statistiche (medie, covarianze e correlazioni) e possibilmente implicare valori mancanti. Per ulteriori informazioni, consultare l'argomento [“Stima delle statistiche e Impatto dei valori mancanti”](#) a pagina 4.
- Se si seleziona EM o Regression, fare clic su **Variabili** per specificare un sottoinsieme da utilizzare per la stima. Per ulteriori informazioni, consultare l'argomento [“Variabili previste e predittore”](#) a pagina 6.
- Selezionare una variabile di etichetta del caso. Questa variabile viene utilizzata per etichettare i casi in tabelle di modelli che visualizzano singoli casi.

## Visualizzazione Modelli dei Valori mancanti

È possibile scegliere di visualizzare varie tabelle che mostrano i modelli e l'estensione dei dati mancanti. Queste tabelle possono aiutare a identificare:

- Dove si trovano i valori mancanti
- Se le coppie di variabili tendono ad avere valori mancanti nei singoli casi
- Se i valori dei dati sono estremi

Visualizza

Sono disponibili tre tipi di tabelle per la visualizzazione dei modelli di dati mancanti.

**I casi Tabulati.** I modelli di valore mancanti nelle variabili di analisi sono tabulati, con frequenze mostrate per ogni modello. Utilizzare **Sort variabili per modello di valore mancante** per specificare se i conteggi e le variabili sono ordinati per similitudine di pattern. Utilizzare **Omit pattern con meno di n% dei casi** per eliminare i pattern che si verificano infrequentemente.

**Casi con valori mancanti.** Ogni caso con un valore mancante o estremo viene tabulato per ogni variabile di analisi. Utilizzare **Sort variabili per modello di valore mancante** per specificare se i conteggi e le variabili sono ordinati per similitudine di pattern.

**Tutti i casi.** Ogni caso è tabulato, e i valori mancanti e estremi sono indicati per ogni variabile. I casi sono elencati nell'ordine che appaiono nel file dei dati, a meno che una variabile non sia specificata in **Ordina per**.

Nelle tabelle che visualizzano singoli casi vengono utilizzati i seguenti simboli:

- + . Estremamente alto valore
- . Estremamente basso valore
- S.** Valore mancante di sistema
- A.** Primo tipo di utente - valore mancante
- B.** Secondo tipo di utente - valore mancante
- C.** Terzo tipo di valore mancante utente

Variabili

È possibile visualizzare ulteriori informazioni per le variabili incluse nell'analisi. Le variabili che si aggiungono a **Informazioni aggiuntive per** vengono visualizzate singolarmente nella tabella dei modelli mancanti. Per le variabili quantitative (scale) viene visualizzata la media; per le variabili categoriali viene visualizzato il numero di casi che presentano lo schema in ogni categoria.

- **Ordina per.** I casi sono elencati in base all'ordine crescente o decrescente dei valori della variabile specificata. Disponibile solo per **Tutti i casi**.

Per Visualizzare Modelli di valore mancanti

1. Nella finestra principale di dialogo Analisi del valore mancante, selezionare la variabile (s) per la quale si desidera visualizzare i modelli di valore mancanti.

2. Fare clic su **Modelli**.
3. Selezionare le tabelle di modello che si desidera visualizzare.

## Visualizzazione Statistiche Descrittive per i Valori mancanti

### Statistiche univariata

Le statistiche univariate possono aiutare a identificare l'estensione generale dei dati mancanti. Per ogni variabile, vengono visualizzati i seguenti:

- Numero di valori non mancanti
- Numero e percentuale dei valori mancanti

Per le variabili quantitative (scale) vengono visualizzate anche le seguenti:

- Media
- Deviazione standard
- Numero di valori estremamente alti e bassi

### Statistiche variabile indicatore

Per ogni variabile viene creata una variabile indicatore. Questa variabile categoriale indica se la variabile è presente o mancante per un singolo caso. Le variabili indicatore vengono utilizzate per creare le tabelle di mancata corrispondenza,  $t$  e tabelle di frequenza.

**Percentuale di mismatch.** Per ogni coppia di variabili, visualizza la percentuale di casi in cui una variabile ha un valore mancante e l'altra variabile ha un valore non mancante. Ogni elemento diagonale nella tabella contiene la percentuale di valori mancanti per una singola variabile.

**$t$  test con gruppi formati da variabili indicatore.** I mezzi di due gruppi sono confrontati per ogni variabile quantitativa, utilizzando la statistica  $t$  di Student. I gruppi specificano se una variabile è presente o mancante. Vengono visualizzate la statistica  $t$ , i gradi di libertà, i conteggi dei valori mancanti e non mancanti e i mezzi dei due gruppi. È anche possibile visualizzare eventuali probabilità a due code associate alla statistica  $t$ . Se la tua analisi si traduce in più di un test, non utilizzare queste probabilità per test di significatività. Le probabilità sono appropriate solo quando viene calcolato un singolo test.

**Crosstabulazioni di variabili categoriali e di indicatore.** Viene visualizzata una tabella per ogni variabile categoriale. Per ogni categoria, la tabella mostra la frequenza e la percentuale di valori non mancanti per le altre variabili. Vengono visualizzate anche le percentuali di ogni tipo di valore mancante.

**Le variabili Omit mancano meno del n% dei casi.** Per ridurre la dimensione della tabella è possibile omettere statistiche che vengono calcolate solo per un numero ridotto di casi.

Per Visualizzare Statistiche Descrittive

1. Nella finestra principale di dialogo Analisi del valore mancante, selezionare la variabile (s) per la quale si desidera visualizzare le statistiche descrittive del valore mancante.
2. Fare clic su **Descriptive**.
3. Scegliere le statistiche descrittive che si desidera visualizzare.

## Stima delle statistiche e Impatto dei valori mancanti

È possibile scegliere di stimare mezzi, deviazioni standard, covarianze e correlazioni che utilizzano listwise (solo casi completi), pairwise, EM (aspettative - massimizzazione) e / o metodi di regressione. È inoltre possibile scegliere di imputare i valori mancanti (valori di sostituzione stimati). Da notare che Più Imputazione è generalmente considerata superiore alla singola imputazione per risolvere il problema dei valori mancanti. Il test MCAR di Little è ancora utile per determinare se l'imputazione sia necessaria.

### Metodo listwise

Questo metodo utilizza solo casi completi. Se una qualsiasi delle variabili di analisi ha valori mancanti, il caso viene ommesso dai calcoli.



## Metodo Pairwise

Questo metodo guarda a coppie di variabili di analisi e utilizza un caso solo se ha valori non mancanti per entrambe le variabili. Frequenze, mezzi e deviazioni standard vengono calcolate separatamente per ogni coppia. Poiché altri valori mancanti nel caso vengono ignorati, le correlazioni e le covarianze per due variabili non dipendono da valori mancanti in altre variabili.

## Metodo EM

Questo metodo presuppone una distribuzione per i dati parzialmente mancanti e le basi inferenze sulla probabilità sotto tale distribuzione. Ogni iterazione consiste in un passo E e un passo M. Il passo E trova l'aspettativa condizionale dei dati "mancanti", dati i valori osservati e le stime attuali dei parametri. Queste aspettative vengono poi sostituite per i dati "mancanti". Nella fase M, le stime di massima verosimiglianza dei parametri sono calcolate come se i dati mancanti fossero stati compilati. "Manca" è racchiuso tra virgolette perché i valori mancanti non vengono direttamente compilati. Le funzioni di essi invece sono utilizzate nel logaritmo.

La statistica di Roderick J. A. Little's chi - quadrato per verificare se i valori sono mancanti completamente a caso (MCAR) viene stampato come una nota a piè di pagina alle matrici EM. Per questo test, l'ipotesi null è che i dati siano mancanti completamente a caso e che il valore  $p$  sia significativo al livello 0.05 . Se il valore è inferiore a 0.05, i dati non mancano completamente a caso. I dati possono essere mancanti a caso (MAR) o non mancanti a caso (NMAR). Non è possibile assumere uno o l'altro e necessità di analizzare i dati per determinare come mancano i dati.

## Metodo di regressione

Questo metodo calcola più stime di regressione lineare e ha opzioni per ampliare le stime con componenti casuali. A ogni valore previsto, la procedura può aggiungere un residuo da un caso completo selezionato casualmente, un deviato normale casuale, oppure un deviato casuale (scalato dalla radice quadrata del quadrato medio residuo) dalla distribuzione  $t$ .

## Opzioni Di Stima EM

Utilizzando un processo iterativo, il metodo EM stima i mezzi, la matrice di covarianza e la correlazione di variabili quantitative (scale) con valori mancanti.

**Distribuzione.** EM fa inferenze in base alla probabilità sotto la distribuzione specificata. Per impostazione predefinita, viene assunta una distribuzione normale. Se si sa che le code della distribuzione sono più lunghe di quelle di una distribuzione normale, è possibile richiedere che la procedura costruisce la funzione di verosimiglianza da una distribuzione  $t$  di Student con  $n$  gradi di libertà. La distribuzione normale mista fornisce anche una distribuzione con code più lunghe. Specificare il rapporto tra le deviazioni standard della distribuzione normale mista e la proporzione di miscela delle due distribuzioni. La distribuzione normale mista presuppone che solo le deviazioni standard delle distribuzioni differiscano. I mezzi devono essere gli stessi.

**Numero massimo di iterazioni.** Imposta il numero massimo di iterazioni per stimare la vera covarianza. La procedura si interrompe quando si raggiunge questo numero di iterazioni, anche se le stime non sono convergenti.

**Salva dati completati.** È possibile salvare un dataset con i valori imputati al posto dei valori mancanti. Attenzione, però, che le statistiche basate sulla covarianza che utilizzano i valori imputati sottovaluteranno i rispettivi valori di parametro. Il grado di sottostima è proporzionale al numero di casi che sono congiuntamente inosservati.

Per specificare le opzioni EM

1. Nella finestra principale di dialogo Analisi del valore mancante, selezionare la variabile (s) per la quale si desidera stimare i valori mancanti utilizzando il metodo EM.
2. Selezionare **EM** nel gruppo di stima.
3. Per specificare le variabili previste e predittrici, fare clic su **Variabili**. Per ulteriori informazioni, consultare l'argomento [“Variabili previste e predittore”](#) a pagina 6.
4. Fare clic su **EM**.

5. Selezionare le opzioni EM desiderate.

## Opzioni Di Stima Regressione

Il metodo di regressione stima i valori mancanti utilizzando la regressione lineare multipla. Vengono visualizzati i mezzi, la matrice di covarianza e la matrice di correlazione delle variabili previste.

**Regolazione della stima.** Il metodo di regressione può aggiungere un componente casuale alle stime di regressione. È possibile selezionare i residui, le variate normali, le variate *t* di Student o nessuna regolazione.

- *Residui.* I termini di errore vengono scelti casualmente dai residui osservati dei casi completi da aggiungere alla stime di regressione.
- *Normovariati.* I termini di errore sono estratti casualmente da una distribuzione con valore atteso 0 e deviazione standard pari alla radice quadrata del quadrato dell'errore medio del modello di regressione.
- *Variate T di Student.* I termini di errore vengono estratti casualmente dalla distribuzione *t* con i gradi di libertà specificati e scalati in base all'errore quadratico medio.

**Numero massimo di predittori.** Imposta un limite massimo sul numero di variabili predittore (indipendenti) utilizzate nel processo di stima.

**Salva dati completati.** Scrive un dataset nella sessione corrente o un file di dati IBM SPSS Statistiche esterno, con valori mancanti sostituiti da valori stimati dal metodo di regressione.

Per Specificare Opzioni Di Regressione

1. Nella finestra principale di dialogo Analisi del valore mancante, selezionare la variabile (s) per la quale si desidera stimare i valori mancanti utilizzando il metodo di regressione.
2. Selezionare **Regressione** nel gruppo di stima.
3. Per specificare le variabili previste e predittrici, fare clic su **Variabili**. Per ulteriori informazioni, consultare l'argomento [“Variabili previste e predittore”](#) a pagina 6.
4. Fare clic su **Regressione**.
5. Selezionare le opzioni di regressione desiderate.

## Variabili previste e predittore

Per impostazione predefinita, tutte le variabili quantitative sono utilizzate per la stima EM e regressione. Se necessario, è possibile scegliere variabili specifiche come variabili previste e predittrici nella stima (s). Una determinata variabile può essere in entrambe le liste, ma ci sono situazioni in cui si potrebbe voler limitare l'uso di una variabile. Ad esempio, alcuni analisti sono a disagio stimando valori di variabili di risultato. Si può anche voler utilizzare diverse variabili per diverse stime ed eseguire più volte la procedura. Ad esempio, se hai una serie di articoli che sono gli ascolti degli infermieri e un altro set che sono i ascolti dei medici, puoi voler fare una sola corsa usando l'articolo degli infermieri per stimare gli articoli degli infermieri scomparsi e un'altra corsa per le stime degli articoli dei medici.

Un'altra considerazione nasce quando si utilizza il metodo di regressione. In più regressione, l'utilizzo di un grande sottoinsieme di variabili indipendenti può produrre valori previsti più poveri di un sottoinsieme più piccolo. Pertanto, una variabile deve raggiungere un limite *F-to-enter* di 4.0 da utilizzare. Questo limite può essere modificato con la sintassi.

Per Specificare Variabili e Predittore previste

1. Nella finestra principale di dialogo Analisi del valore mancante, selezionare la variabile (s) per la quale si desidera stimare i valori mancanti utilizzando il metodo di regressione.
2. Selezionare **EM** o **Regressione** nel gruppo di stima.
3. Fare clic su **Variabili**.
4. Se si desidera utilizzare specifiche piuttosto che tutte le variabili come variabili previste e predittrici, selezionare **Seleziona variabili** e spostare le variabili all'elenco appropriato (s).

## Funzioni Aggiuntive di comando MVA

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare variabili descrittive separate per modelli di valore mancanti, modelli di dati e modelli tabulati utilizzando la parola chiave DESCRIBE sui sottocomandi MPATTERN, DPATTERN o TPATTERN .
- Specificare più di una variabile di ordinamento per la tabella dei modelli di dati, utilizzando il comando DPATTERN .
- Specificare più di una variabile di ordinamento per i modelli di dati, utilizzando il sottocomando DPATTERN .
- Specificare tolleranza e convergenza, utilizzando il sottocomando EM .
- Specificare la tolleranza e  $F_{per}$  entrare, utilizzando il comando REGRESSION .
- Specificare diversi elenchi variabili per EM e Regression, utilizzando i sottocomandi EM e REGRESSION .
- Specificare percentuali diverse per la soppressione dei casi visualizzati, per ognuna di TTESTS, TABULATE e MISMATCH.

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

## Assegnazione multipla

Lo scopo di più imputazione è quello di generare valori possibili per i valori mancanti, creando così diverse serie di dati "completi". Le procedure analitiche che funzionano con più datasets di imputazione producono output per ogni dataset "completo", più output pooled che stima quali sarebbero stati i risultati se il dataset originale non avesse valori mancanti. Questi risultati pooled sono generalmente più accurati di quelli forniti da singoli metodi di imputazione.




Considerazioni Dati Imputazione multiple

**variabili di analisi.** Le variabili di analisi possono essere:









- *Nominale.* Una variabile può essere considerata nominale quando i suoi valori rappresentano categorie prive di classificazione intrinseca (ad esempio, il dipartimento della società in cui lavora un dipendente). Degli esempi di variabili nominali includono la regione, il codice postale e l'affiliazione religiosa.
- *Ordinale.* Una variabile può essere trattata come ordinale quando i suoi valori rappresentano categorie con una classificazione intrinseca (ad esempio, livelli di soddisfazione del servizio da molto insoddisfatti a molto soddisfatti). Degli esempi di variabili ordinali includono i punteggi di atteggiamento che rappresentano i gradi di soddisfazione o fiducia e i punteggi di classificazione delle preferenze.
- *Scala.* Una variabile può essere trattata come una scala (continua) quando i suoi valori rappresentano categorie ordinate con una metrica significativa, in modo che i confronti di distanza tra i valori siano appropriati. Esempi di variabili di scala sono l'età espressa in anni o il reddito espresso in migliaia di Euro.

La procedura presuppone che il livello di misurazione appropriato sia stato assegnato a tutte le variabili; tuttavia, è possibile modificare temporaneamente il livello di misurazione per una variabile facendo clic con il tasto destro del mouse sulla variabile nell'elenco delle variabili di origine e selezionando un livello di misurazione dal menu a comparsa. Per modificare in modo permanente il livello di misurazione per una variabile,

L'icona accanto a ciascuna variabile nell'elenco delle variabili identifica il livello di misurazione e il tipo di dati.

	Numerico	Stringa	Dati	Ora
Scala (continuo)		n/d		

*Tabella 1. Icone del livello di misurazione (Continua)*

	<b>Numerico</b>	<b>Stringa</b>	<b>Dati</b>	<b>Ora</b>
Ordinale				
Nominale				

**Peso della frequenza.** I pesi di frequenza (replica) sono onorati da questa procedura. I casi con valore di peso di replica negativo o zero vengono ignorati. I pesi non interi sono arrotondati al numero intero più vicino.

**Peso analisi.** I pesi di analisi (regressione o campionamento) sono incorporati in sintesi dei valori mancanti e nei modelli di imputazione di adattamento. Sono esclusi i casi con un peso di analisi negativo o zero.

**Campioni complessi.** La Procedura Di Imputazione Multipla non gestisce esplicitamente strati, cluster o altre strutture di campionamento complesse, anche se può accettare pesi di campionamento finali sotto forma di variabile peso di analisi. Inoltre, notare che le procedure di Complex Sampling attualmente non analizzano automaticamente i datasets moltiplicati. Per un elenco completo delle procedure che supportano il pool, consultare [“Analisi Dei Dati Di Imputazione Multipla”](#) a pagina 14.

**Valori mancanti.** Entrambi i valori mancanti dell'utente e del sistema sono trattati come valori non validi; cioè, entrambi i tipi di valori mancanti vengono sostituiti quando i valori sono imputati e entrambi sono trattati come valori non validi di variabili utilizzate come predittori nei modelli di imputazione. I valori mancanti dell'utente e del sistema sono trattati anche come mancanti nelle analisi dei valori mancanti.

**Risultati di replica (Impute Valori dati mancanti).** Se si desidera replicare esattamente i risultati di imputazione, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali, lo stesso ordine di dati e lo stesso ordine variabile, oltre ad utilizzare le stesse impostazioni di procedura.

- **Generazione numero casuale.** La procedura utilizza la generazione di numeri casuali durante il calcolo dei valori imputati. Per riprodurre gli stessi risultati randomizzati in futuro, utilizzare lo stesso valore di inizializzazione per il generatore di numeri casuali prima di ogni esecuzione della procedura Impatto Data Values.
- **Ordine dei casi.** I valori sono imputati nel caso ordine.
- **Ordine variabile.** Il metodo di imputazione della specifica condizionale (FCS) condiziona i valori nell'ordine specificato nell'elenco Variabili di analisi.

Ci sono due dialoghi dedicati a più imputazioni.

- [Analizzare Patterns](#) fornisce misure descrittive degli schemi dei valori mancanti nei dati, e può essere utile come passo esplorativo prima dell'imputazione.
- [Importa i valori dei dati mancanti](#) viene utilizzato per generare più imputazioni. I dataset completi possono essere analizzati mediante delle procedure che supportano i dataset ad assegnazione multipla. Consultare [Generatori di numeri casuali](#) per informazioni relative all'analisi dei dataset di assegnazione multipla e un elenco di procedure che supportano tali dati.

### Informazioni correlate

## Analizzare Modelli

[Analizzare Patterns](#) fornisce misure descrittive degli schemi dei valori mancanti nei dati, e può essere utile come passo esplorativo prima dell'imputazione.

**Esempio.** Un fornitore di telecomunicazioni vuole comprendere meglio i modelli di utilizzo del servizio nel proprio database clienti. Hanno dati completi per i servizi utilizzati dai loro clienti, ma le informazioni

demografiche raccolte dall'azienda hanno un certo numero di valori mancanti. Analizzare gli schemi dei valori mancanti può aiutare a determinare i prossimi passi per l'imputazione.

Dai menu, scegliere:

### **Analizzare > L'Imputazione Multipla > Analizzare Modelli ...**

1. Selezionare almeno due variabili di analisi. La procedura analizza i modelli di dati mancanti per queste variabili.

Impostazioni facoltative

**Peso analisi.** Questa variabile contiene pesi di analisi (regressione o campionamento). La procedura incorpora i pesi di analisi in sintesi dei valori mancanti. Sono esclusi i casi con un peso di analisi negativo o zero.

**Output.** È disponibile la seguente emissione opzionale:

- **Riepilogo dei valori mancanti.** Questo visualizza un grafico a torta panato che mostra il numero e la percentuale di variabili di analisi, casi o singoli valori dati che hanno uno o più valori mancanti.
- **Modelli dei valori mancanti.** Questo visualizza modelli tabulati dei valori mancanti. Ogni modello corrisponde a un gruppo di casi con lo stesso modello di dati incompleti e completi sulle variabili di analisi. È possibile utilizzare questo output per determinare se il metodo di imputazione monotono può essere utilizzato per i tuoi dati, o in caso contrario, in che modo i tuoi dati approssimano un modello monotono. La procedura ordina le variabili di analisi per rivelare o approssimare uno schema monotono. Se nessun modello nonmonotono esiste dopo il riordino si può concludere che i dati hanno uno schema monotono quando le variabili di analisi sono ordinate in quanto tali.
- **Variabili con la frequenza più elevata dei valori mancanti.** Questo visualizza una tabella di variabili di analisi ordinate per percentuale di valori mancanti in ordine decrescente. La tabella include le statistiche descrittive (media e deviazione standard) per le variabili di scala.

È possibile controllare il numero massimo di variabili da visualizzare e la percentuale minima mancante per una variabile da includere nel display. Viene visualizzata la serie di variabili che soddisfano entrambi i criteri. Ad esempio, impostando il numero massimo di variabili a 50 e la percentuale minima mancante a 25 richiede che la tabella visualizza fino a 50 variabili che hanno almeno 25% valori mancanti. Se ci sono 60 variabili di analisi ma solo 15 hanno 25% o più valori mancanti, l'output include solo 15 variabili.

## **Valori dati mancanti**

I Valori Dati mancanti sono utilizzati per generare più imputazioni. I dataset completi possono essere analizzati mediante delle procedure che supportano i dataset ad assegnazione multipla. Consultare “Analisi Dei Dati Di Imputazione Multipla” a pagina 14 per informazioni sull'analisi di più datasets di imputazione e un elenco di procedure che supportano questi dati.

**Esempio.** Un fornitore di telecomunicazioni vuole comprendere meglio i modelli di utilizzo del servizio nel proprio database clienti. Hanno dati completi per i servizi utilizzati dai loro clienti, ma le informazioni demografiche raccolte dall'azienda hanno un certo numero di valori mancanti. Inoltre, questi valori non mancano completamente a caso, quindi l'imputazione multipla verrà utilizzata per completare il dataset.

Dai menu, scegliere:

### **Analizzare > Imputazione Multipla > Impatto Valori dati mancanti ...**

1. Selezionare almeno due variabili nel modello di imputazione. La procedura imputa più valori per i dati mancanti per queste variabili.
2. Specificare il numero di imputazioni da calcolare. Per impostazione predefinita, questo valore è 5.
3. Specificare un dataset o un file di dati di formato IBM SPSS Statistica a cui devono essere scritti i dati imputati.

Il dataset di output è costituito dai dati di caso originali con dati mancanti più una serie di casi con valori imputati per ogni imputazione. Ad esempio, se il dataset originale ha 100 casi e si hanno cinque imputazioni, il dataset di output avrà 600 casi. Tutte le variabili nel dataset di input sono incluse nel

dataset di output. Proprietà dizionari (nomi, etichette, ecc.) delle variabili esistenti vengono copiate nel nuovo dataset. Il file contiene anche una nuova variabile, *Imputation\_*, una variabile numerica che indica l'assegnazione (0 per i dati originali o 1 ..n per i casi con valori assegnati).

La procedura definisce automaticamente la variabile *Imputation\_* come variabile split quando viene creato il dataset di output. Se le suddivisioni sono in vigore quando la procedura viene eseguita, il dataset di output include una serie di imputazioni per ogni combinazione di valori di variabili split.

Impostazioni facoltative

**Peso analisi.** Questa variabile contiene pesi di analisi (regressione o campionamento). La procedura incorpora i pesi di analisi nei modelli di regressione e classificazione utilizzati per imputare i valori mancanti. I pesi di analisi vengono utilizzati anche in riepiloghi di valori imputati; ad esempio, media, deviazione standard e errore standard. Sono esclusi i casi con un peso di analisi negativo o zero.

Campi con livello di misurazione sconosciuto

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

**Esegui scansione dati.** Legge i dati del dataset attivo e assegna un livello di misurazione predefinito a tutti i campi con livello di misurazione sconosciuto. Con dataset di grandi dimensioni, questa operazione può richiedere del tempo.

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

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

## Metodo

La scheda Metodo specifica come saranno imputati i valori mancanti, compresi i tipi di modelli utilizzati. I predicatori categoriali sono indicatori (dummy) codificati.

### Metodo di assegnazione

Il metodo **Automatico** scandisce i dati e utilizza il metodo monotone se i dati mostrano un modello monotono di valori mancanti; altrimenti viene utilizzata la specifica completamente condizionale. Se si è certi di quale metodo si desidera utilizzare, è possibile specificarlo come metodo **Custom**.

### Specifica completamente condizionale

Si tratta di un metodo iterativo Markov catena Monte Carlo (MCMC) che può essere utilizzato quando lo schema dei dati mancanti è arbitrario (monotono o non monotono).

Per ogni iterazione e per ogni variabile nell'ordine specificato nell'elenco delle variabili, il metodo FCS (completamente condizionale) si adatta a un modello univariato (variabile singolo dipendente) utilizzando tutte le altre variabili disponibili nel modello come predittori, quindi imputa i valori mancanti per la variabile in dotazione. Il metodo continua fino al raggiungimento del numero massimo di iterazioni e i valori imputati alla massima iterazione vengono salvati nel dataset figurato.

### Numero massimo di iterazioni

Questo specifica il numero di iterazioni, o "passi", assunti dalla catena Markov utilizzata dal metodo FCS. Se il metodo FCS è stato scelto automaticamente, utilizza il numero predefinito di 10 iterazioni. Quando si sceglie esplicitamente FCS, è possibile specificare un numero personalizzato di iterazioni. Potrebbe essere necessario aumentare il numero di iterazioni se la catena Markov non si è convertita. Nella scheda Output è possibile salvare i dati della cronologia delle iterazioni FCS e trarlo per valutare la convergenza.

### Monotono

Si tratta di un metodo non iterativo che può essere utilizzato solo quando i dati hanno un modello monotono di valori mancanti. Un modello monotono esiste quando si possono ordinare le

variabili tali che, se una variabile ha un valore non mancante, tutte le variabili precedenti hanno anche valori non mancanti. Quando si specifica questo come un metodo **Custom**, assicurarsi di specificare le variabili nella lista in un ordine che mostra uno schema monotono.

Per ogni variabile nell'ordine monotono, il metodo monotone si adatta a un modello univariato (variabile dipendente singolo) utilizzando tutte le variabili precedenti nel modello come predittori, quindi imputa i valori mancanti per la variabile in adattamento. Questi valori imputati vengono salvati nel dataset implicito.

### Includere le interazioni a due vie

Quando il metodo di imputazione viene scelto automaticamente, il modello di imputazione per ogni variabile include un termine costante e gli effetti principali per le variabili predittori. Quando si sceglie un metodo specifico, si possono opzionalmente includere tutte le possibili interazioni a due vie tra le variabili predittive categoriali.

### Tipo di modello per le variabili di scala

#### Regressione lineare

Quando il metodo di imputazione viene scelto automaticamente, la regressione lineare viene utilizzata come modello univariato per le variabili di scala.

#### PMM (Predictive Mean Matching)

Quando si sceglie un metodo specifico, si può alternativamente scegliere la corrispondenza media predittiva (PMM) come modello per le variabili di scala. PMM è una variante di regressione lineare che assicura che i valori imputati siano plausibili. Per PMM, il valore imputato si basa sul valore definito per il valore **Randomly selezionare un caso completo dal valore delle previsioni più vicine (k)**, dove (k) è un intero positivo con un valore predefinito di 5.

La regressione logistica viene sempre utilizzata come modello univariato per variabili categoriali. Indipendentemente dal tipo di modello, i predittori categoriali vengono gestiti utilizzando la codifica dell'indicatore (dummy).

### Tolleranza della singolarità

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

## Vincoli

La Scheda Vincoli consente di limitare il ruolo di una variabile durante l'imputazione e limitare la gamma di valori imputati di una variabile di scala in modo che siano plausibili. Inoltre, è possibile limitare l'analisi a variabili con meno di una percentuale massima di valori mancanti.

**Scansione dei dati per riepilogo variabile.** Cliccando su **Data di scansione** è possibile che l'elenco mostri le variabili di analisi e la percentuale osservata mancante, minima e massima per ciascuno. I riepiloghi possono essere basati su tutti i casi o limitati ad una scansione dei primi casi  $n$ , come specificato nella casella di testo Casi. Cliccando su **Rescan Data** vengono aggiornati i riepiloghi di distribuzione.

Definisci vincoli

- **Ruolo.** Ciò consente di personalizzare l'insieme di variabili da imputare e / o trattare come predittori. Di solito, ogni variabile di analisi è considerata sia come dipendente che predittivo nel modello di imputazione. Il **Ruolo** può essere utilizzato per disattivare l'imputazione per le variabili che si desidera **Utilizza come predittore solo** o per escludere le variabili dall'essere utilizzate come predittori (**Impatto solo**) e quindi rendere più compatto il modello di previsione. Questo è l'unico vincolo che può essere specificato per le variabili categoriali, o per le variabili utilizzate solo come predittori.
- **Min e Max.** Queste colonne consentono di specificare i valori minimi e massimi consentiti per le variabili di scala. Se un valore imputato ricade al di fuori di questo intervallo, la procedura richiama un altro valore fino a quando non trova uno all'interno dell'intervallo o si raggiunge il numero massimo di attentioni (vedi **Massimo attico** sotto). Queste colonne sono disponibili solo se **Regressione Lineare** viene selezionato come tipo di modello variabile sulla scheda Metodo.

- **Arrotondamento.** Alcune variabili possono essere utilizzate come scala, ma hanno valori che sono naturalmente ulteriormente limitati; per esempio, il numero di persone in un nucleo familiare deve essere intero e l'importo speso durante una visita al negozio di alimentari non può avere centesimi frazionali. Questa colonna consente di specificare la denominazione più piccola da accettare. Ad esempio, per ottenere valori interi, specificare 1 come denominazione di arrotondamento; per ottenere valori arrotondati al centesimo più vicino, specificare 0.01. In generale, i valori sono arrotondati al numero intero più vicino della denominazione di arrotondamento. La seguente tabella mostra come i diversi valori di arrotondamento agiscono su un valore assegnato di 6.64823 (prima dell'arrotondamento).

<i>Tabella 2. Risultati di arrotondamento</i>	
<b>Arrotondamento</b>	<b>Denominazione</b>
<b>Valore a cui 6.64832 è arrotondato</b>	
10	10
1	7
0.25	6.75
0.1	6.6
0.01	6.65

**Escludere le variabili con grandi quantità di dati mancanti.** In genere le variabili di analisi sono imputate e utilizzate come predittori senza considerare quanti valori mancanti hanno, a patto di avere dati sufficienti per stimare un modello di imputazione. È possibile scegliere di escludere variabili che hanno una percentuale elevata di valori mancanti. Ad esempio, se si specifica 50 come **Massimo percentuale mancante**, le variabili di analisi che hanno più di 50% valori mancanti non sono imputate, né sono utilizzate come predittori nei modelli di imputazione.

**Massimo massimo.** Se vengono specificati valori minimi o massimi per valori impliciti di variabili di scala (vedi **Mina e Max** sopra), la procedura tenta di disegnare valori per un caso fino a quando non trova una serie di valori che rientrano nelle gamme specificate. Se una serie di valori non viene ottenuta all'interno del numero specificato di disegnati per caso, la procedura richiama un'altra serie di parametri del modello e ripete il processo di disegno del caso. Si verifica un errore se un insieme di valori all'interno degli intervalli non viene ottenuto all'interno del numero specificato di case e parametri.

Notare che aumentare questi valori può aumentare il tempo di elaborazione. Se la procedura si prende a lungo, o non è in grado di trovare attenti adatti, controllare i valori minimi e massimi specificati per garantirne l'adeguatezza.

## Output

**Visualizza.** Controlla la visualizzazione dell'output. Viene sempre visualizzato un riepilogo di imputazione generale, che include tabelle relative alle specifiche di imputazione, alle iterazioni (per il metodo di specifica completamente condizionale), alle variabili dipendenti imputate, alle variabili dipendenti escluse dall'imputazione, e alla sequenza di imputazione. Se specificato, vengono mostrate anche costine per le variabili di analisi.

- **Modello Imputazione.** Questo visualizza il modello di imputazione per variabili e predittori dipendenti e include tipo di modello univariato, effetti modello e numero di valori imputati.
- **Statistiche descrittive.** In questo modo vengono visualizzate le statistiche descrittive per le variabili dipendenti per le quali i valori sono imputati. Per le variabili di scala le statistiche descrittive includono media, conteggio, deviazione standard, min e max per i dati di input originali (prima dell'imputazione), valori imputati (per imputazione), e dati completi (valori originali e imputati insieme - per imputazione). Per le variabili categoriali le statistiche descrittive includono il conteggio e la percentuale per categoria per i dati di input originali (prima dell'imputazione), i valori imputati (per imputazione), e i dati completi (valori originali e imputati insieme - per imputazione).

**Cronologia iterazioni.** Quando viene utilizzato il metodo di imputazione delle specifiche completamente condizionale, è possibile richiedere un dataset che contenga dati di cronologia delle iterazioni per



l'imputazione FCS. Il dataset contiene medie e deviazioni standard per iterazione e imputazione per ogni variabile dipendente di scala per cui i valori sono imputati. È possibile tracciare i dati per aiutare a valutare la convergenza del modello.

## MULTIPLE IMPUTAZIONE Comando Funzioni aggiuntive

Il linguaggio della sintassi dei comandi consente inoltre di:

- Specificare un sottoinsieme di variabili per le quali vengono mostrate statistiche descrittive (IMPUTATIONSUMMARIES secondario).
- Specificare sia un'analisi dei modelli mancanti e l'imputazione in un'unica esecuzione della procedura.
- Specificare il numero massimo di parametri di modello consentiti quando si impatta qualsiasi variabile (MAXMODELPARAM keyword).

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

## Gestione dei dati di Impatto Multiplo

Quando viene creato un dataset di imputazione multipla (MI), viene aggiunta una variabile chiamata *Imputation\_*, con etichetta variabile *Numero di Imputazione*, e il dataset viene ordinato da esso in ordine crescente. I casi dal dataset originale hanno un valore di 0. I casi per i valori imputati sono numerati da 1 tramite *M*, dove *M* è il numero di imputazioni.

Quando si apre un dataset, la presenza di *Imputation\_* identifica il dataset come un possibile dataset MI.

Attivazione di un dataset di Assegnazione Multipla per l'analisi

Il dataset deve essere suddiviso utilizzando l'opzione **Confronta gruppi**, con *Imputation\_* come variabile di raggruppamento, per essere trattata come un dataset MI in analisi. È anche possibile definire suddivise su altre variabili.

Dai menu, scegliere:

### Dati > File suddiviso...

1. Selezionare **Confronta gruppi**.
2. Selezionare *Numero di imputazione [imputation\_]* come variabile su casi di gruppo su.

In alternativa, quando si attivano le marcature accese (vedi sotto), il file viene suddiviso su *Numero di Impatto [Imputation\_]*.

Distinguo Valori Imputati da Valori Osservati

È possibile distinguere i valori imputati dai valori osservati per colore di sfondo cellulare, il tipo di carattere e il tipo di grassetto (per valori imputati). Quando si crea un nuovo dataset nella sessione corrente con Impute Valori mancanti, le marcature vengono attivate per impostazione predefinita. Quando si apre un file di dati salvato che include le imputazioni, le marcature vengono disattivate.

Per attivare le marcature, dai menu dell'Editor dei dati scegliere:

### Visualizza > Segnala dati Imputati ...

In alternativa, è possibile accendere contrassegni cliccando sul pulsante di marcatura dell'imputazione al margine destro della barra di modifica in Data View dell'Editor dei dati.

Spostamento Tra Imputazioni

1. Dai menu, scegliere:

#### Modifica > Vai all'Imputazione ...

2. Selezionare l'assegnazione (o Dati originali) dall'elenco a discesa.

In alternativa, è possibile selezionare l'assegnazione dall'elenco a discesa nella barra di modifica nella Vista dati dell'Editor dei dati.

La posizione relativa dei casi viene mantenuta quando si selezionano le assegnazioni. Ad esempio, se ci sono 1000 casi nel dataset originale, il caso 1034, il trentaquattresimo nella prima assegnazione, viene visualizzato nella parte superiore della griglia. Se si seleziona l'assegnazione **2** nell'elenco a discesa, il caso 2034, il trentaquattresimo dell'assegnazione 2, viene visualizzato nella parte superiore della griglia. Se si seleziona **Dati originali** nell'elenco a discesa, il caso 34 viene visualizzato nella parte superiore della griglia. Quando si passa da un'assegnazione all'altra, viene mantenuta anche la posizione delle colonne, in modo da agevolare il confronto dei valori tra le assegnazioni.

#### Trasformazione e Editing Valori Imputati

A volte sarà necessario eseguire trasformazioni sui dati imputati. Ad esempio, si potrebbe voler prendere il registro di tutti i valori di una variabile di stipendio e salvare il risultato in una nuova variabile. Un valore calcolato utilizzando i dati imputati verrà trattato come imputato se differisce dal valore calcolato utilizzando i dati originali.

Se si modifica un valore imputato in una cella dell'Editor dei dati, tale cella viene comunque trattata come imputato. Non è consigliabile modificare valori imputati in questo modo.

## Analisi Dei Dati Di Imputazione Multipla

Molte procedure supportano la messa in comune dei risultati dell'analisi di insiemi di dati a più imputazioni. Quando le marcature di imputazione sono attivate, viene visualizzata un'icona speciale accanto alle procedure che supportano il pooling. Sul sottomenu Statistiche descrittive del menu Analisi, ad esempio, Frequenze, Descrittive, Esplorazioni e Crosstabs tutti supportano il pooling, mentre Ratio, P-P Plotti e Q-Q Plotti no.

Sia la produzione tabellare che il modello PMML possono essere pooled. Non esiste una nuova procedura per la richiesta di risultati in pool; invece, una nuova scheda della finestra di dialogo Opzioni consente di controllare globalmente i risultati di imputazioni multiple.

- **Pooling di Output Tabular.** Per impostazione predefinita, quando si esegue una procedura supportata su un dataset di assegnazione multipla (MI), i risultati vengono prodotti automaticamente per ciascuna assegnazione, i dati originali (non assegnati) e i risultati raggruppati (finali) che tengono conto delle variazioni tra le assegnazioni. Le statistiche che vengono raggruppate variano per procedura.
- **Pooling di PMML.** È inoltre possibile ottenere PMML in pool da procedure supportate che esportano PMML. Il PMML raggruppato è richiesto nello stesso modo e viene salvato invece del PMML non raggruppato.

Le procedure non supportate non producono né output pooled né file PMML pooled.

#### Livelli di Pooling

L'output viene raggruppato utilizzando uno dei seguenti due livelli:

- **Combinata Naïve.** È disponibile solo il parametro pooled.
- **Combinazione univariata.** Il parametro pooled, il suo errore standard, la prova statistica ed i gradi di libertà efficaci,  $p$ -value, intervallo di confidenza e la diagnostica di pooling (frazione di informazioni mancanti, efficienza relativa, aumento relativo della varianza) vengono mostrate quando disponibili.

I coefficienti (regressione e correlazione), i mezzi (e le differenze medie), e i conteggi sono tipicamente pooled. Quando è disponibile l'errore standard della statistica, viene utilizzato il pooling univariato; altrimenti, viene utilizzato il pooling naïve.

#### Procedure That Support Pooling

Le seguenti procedure supportano i dataset MI, ai livelli di pool specificati per ogni parte di output.

**Frequenze.** Sono supportate le seguenti funzioni:

- La tabella Statistiche supporta le medie nel pool Univariate (se S.E. è richiesta anche la media) e gli N validi e N mancanti nel pooling naïve.
- La Tabella Frequenze supporta Frequenza a pooling Naïve.

**Descrittive.** Sono supportate le seguenti funzioni:

- La tabella Statistiche descrittive supporta le medie nel pool Univariate (se S.E. si richiede anche la media) e N al raggruppamento naïve.

**Crosstabs.** Sono supportate le seguenti funzioni:

- Il tavolo di Crosstabula supporta il conte al Naïve pooling.

**Mezzi.** Sono supportate le seguenti funzioni:

- La tabella Report supporta la media del pooling Univariate (se S.E. si richiede anche la media) e N al raggruppamento naïve.

**Test T: campione unico.** Sono supportate le seguenti funzioni:

- La tabella Statistiche supporta Mean a Univariate pooling e N a Naïve pooling.
- La tabella Test supporta Mean Difference al pooling Univariate.

**Test di T Independent - Samples T.** Sono supportate le seguenti funzioni:

- La tabella Statistiche del Gruppo supporta Mezzi a Univariate pooling e N a Naïve pooling.
- La tabella Test supporta Mean Difference al pooling Univariate.

**Test dei campioni in onda - Samples T.** Sono supportate le seguenti funzioni:

- La tabella Statistiche supporta i Mezzi a Univariate pooling e N a Naïve pooling.
- La tabella Correlazioni supporta Correlazioni e N a pooling Naïve.
- Il tavolo di prova supporta Mean a Univariate pooling.

**One-Way ANOVA.** Sono supportate le seguenti funzioni:

- La Tabella Statistica Descrittiva supporta Mean a Univariate pooling e N a Naïve pooling.
- Il Tavolo Contrasto di contrasto supporta Valore di contrasto alla pooling Univariate.

**GLM univariato e multivariato.** Sono supportate le seguenti funzioni:

- La tabella Fattori tra soggetti supporta N nel pool Naive.
- La tabella Statistiche descrittive supporta la media e N nel pool Naive.
- La tabella delle stime dei parametri supporta il coefficiente, B, al pooling Univariate.
- Il Significato marginale stimato: tabella di stime supporta Mean a pooling Univariate.
- Il Significato Marginale Stimato: Pairwise Confronti table supporta Mean Difference al pooling Univariate.

**Modelli misti lineari.** Sono supportate le seguenti funzioni:

- La Tabella Statistica Descrittiva supporta Mean e N a Naïve pooling.
- Le stime di Fixed Effects table supporta Stima al pooling Univariate.
- Le stime di Covariance Parameters table supporta Stima a pooling Univariate.
- Il Significato marginale stimato: tabella di stime supporta Mean a pooling Univariate.
- Il Significato Marginale Stimato: Pairwise Confronti table supporta Mean Difference al pooling Univariate.

**Modelli lineari generalizzati e Equazioni di stima generalizzate.** Queste procedure supportano PMML in pool.

- La tabella Informazioni variabili categoriali supporta N e Percent a pooling Naïve.
- La Tabella Informazioni Variabili continua supporta N e Mean a pooling Naïve.
- La tabella delle stime dei parametri supporta il coefficiente, B, al pooling Univariate.
- Il Significato marginale stimato: la tabella dei coefficienti di stima supporta Mean a pooling Naïve.
- Il Significato marginale stimato: tabella di stime supporta Mean a pooling Univariate.
- Il Significato Marginale Stimato: Pairwise Confronti table supporta Mean Difference al pooling Univariate.

**Correlazioni Bivariate.** Sono supportate le seguenti funzioni:

- La Tabella Statistica Descrittiva supporta Mean e N a Naïve pooling.
- La tabella Correlazioni supporta Correlazioni e N al pooling Univariate. Si noti che le correlazioni vengono trasformate utilizzando la trasformazione z di Fisher prima del raggruppamento e quindi nuovamente trasformate dopo il raggruppamento.

**Correlazioni Parziali.** Sono supportate le seguenti funzioni:

- La Tabella Statistica Descrittiva supporta Mean e N a Naïve pooling.
- La tabella Correlazioni supporta le Correlazioni a Pooling Naïve.

**Regressione Lineare.** Questa procedura supporta PMML di pooled.

- La Tabella Statistica Descrittiva supporta Mean e N a Naïve pooling.
- La tabella Correlazioni supporta Correlazioni e N a pooling Naïve.
- La tabella dei coefficienti supporta B a Univariate pooling e Correlazioni a pooling Naïve.
- La Tabella Correlazione Coefficienti supporta Correlazioni a Pooling Naïve.
- La Tabella Statistiche Residui supporta Mean e N a Naïve pooling.

**Regressione logistica binaria.** Questa procedura supporta PMML di pooled.

- Le Variabili nella tabella Equitazione supporta B a Univariate pooling.

**Regressione logistica multinomiale.** Questa procedura supporta PMML di pooled.

- La tabella delle stime dei parametri supporta il coefficiente, B, al pooling Univariate.

**Regressione Ordinale.** Sono supportate le seguenti funzioni:

- La tabella delle stime dei parametri supporta il coefficiente, B, al pooling Univariate.

**Analisi Discriminante.** Questa procedura supporta XML modello pooled.

- La tabella Statistiche del Gruppo supporta Mean e Valido N a Naïve pooling.
- La tabella Pooled Within - Gruppi Matrices supporta Correlazioni a Pooling Naïve.
- Il Canonical Discriminant Function Coefficients table supporta i coefficienti Unstandardized a Naïve pooling.
- Le Funzioni alla tabella Centroidi Group supporta Coefficienti non standardizzati a Pooling Naïve.
- La tabella Coefficienti di classificazione delle funzioni supporta i coefficienti a pooling Naïve.

**Test del chi-quadrato.** Sono supportate le seguenti funzioni:

- La Tabella Descrittiva supporta Mean e N a Naïve pooling.
- Il Tavolo Frequenze supporta l'Observato N al Naïve pooling.

**Test binomiale.** Sono supportate le seguenti funzioni:

- La Tabella Descrittiva supporta Mezzi e N a Naïve pooling.
- La tabella Test supporta N, Proporzione Osservata e Prova Proporzionale a pooling Naïve.

**Test delle successioni.** Sono supportate le seguenti funzioni:

- La Tabella Descrittiva supporta Mezzi e N a Naïve pooling.

**Test di Kolmogorov-Smirnov per un campione.** Sono supportate le seguenti funzioni:

- La Tabella Descrittiva supporta Mezzi e N a Naïve pooling.

**Test di due campioni indipendenti.** Sono supportate le seguenti funzioni:

- Il Tavolo Classifica supporta Mean Rank e N al Naïve pooling.
- Il Tavolo Frequenze supporta N a Naïve pooling.

**Test per diversi campioni indipendenti.** Sono supportate le seguenti funzioni:

- Il Tavolo Classifica supporta Mean Rank e N al Naïve pooling.
- Il Tavolo Frequenze supporta i Conti al Naïve pooling.

**Test per due campioni correlati.** Sono supportate le seguenti funzioni:

- Il Tavolo Classifica supporta Mean Rank e N al Naïve pooling.
- Il Tavolo Frequenze supporta N a Naïve pooling.

**Test per diversi campioni correlati.** Sono supportate le seguenti funzioni:

- Il Tavolo Classifica supporta Mean Rank a Naïve pooling.

**Cox Regression.** Questa procedura supporta PMML di pooled.

- Le Variabili nella tabella Equitazione supporta B a Univariate pooling.
- Il Covariato Means table supporta Mean a Naïve pooling.

**Proporzioni di un campione.** Sono supportate le seguenti funzioni:

- La Tabella One - Sample Proporzioni Confidence Intervalli supporta Proporzione al pooling Univariate.
- The One - Sample Proporzioni Test table supporta Observato - Test Value al pooling Univariate.

**Proporzioni di Campioni Indipendenti.** Sono supportate le seguenti funzioni:

- The Independent - Samples Proporzioni Group Statistics table supporta Proporzione a Univariate pooling.
- Il Tavolo Indipendente - Campioni Confidence Intervalli supporta Differenza in Proporzioni a Naïve pooling.
- La Tabella Campioni Indipendenti - Campioni Proporzioni Test supporta Differenza in Proporzioni al pooling Univariate.

**Proporzioni dei campioni accoppiati.** Sono supportate le seguenti funzioni:

- La Tavola di Campioni Accoppiati Statistiche supporta Proporzione a pooling Univariate.
- La Tabella Intervalli di confidenza dei campioni accoppiati supporta Differenza in Proporzioni a Naïve pooling.
- Il Pandato - Campioni Proporzioni Test da tavolo supporta Differenza in Proporzioni al pooling Univariate.

## Opzioni Di Imputazione Multipla

La scheda Assegnazioni multiple consente di controllare due tipi di preferenze relative alle assegnazioni multiple:

**Marcatura dei Dati indicizzati.** Per impostazione predefinita, le celle che contengono i dati assegnati hanno uno sfondo di colore diverso rispetto alle celle contenenti dati non assegnati. L'aspetto facilmente distinguibile dei dati assegnati dovrebbe facilitare la consultazione di un dataset e l'individuazione di tali celle. È possibile cambiare il colore predefinito dello sfondo della cella e il carattere, nonché visualizzare i dati assegnati in grassetto.

**Output analisi.** Questo gruppo controlla il tipo di output del Visualizzatore prodotto ogni volta che viene analizzato un dataset ad assegnazione multipla. Per impostazione predefinita, verrà prodotto output per il dataset originale (precedente all'assegnazione) e per ogni dataset assegnato. Inoltre, nel caso delle procedure che supportano il raggruppamento in pool dei dati assegnati, verranno generati risultati raggruppati finali. Quando viene eseguito il raggruppamento univariato, viene visualizzata anche la diagnostica del raggruppamento. Tuttavia, è possibile sopprimere qualsiasi output che non si desidera produrre.

Per specificare le impostazioni di assegnazione multipla

Dai menu, scegliere:

**Modifica > Opzioni**

Fare clic sulla scheda Assegnazione multipla.

## 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 valori mancanti
  - aspettativa - massimizzazione [6](#)
  - Assegnazione dei valori mancanti [4](#)
  - EM [5](#)
  - funzioni aggiuntive del comando [7](#)
  - metodi [4](#)
  - modelli [3](#)
  - regressione [6](#)
  - statistica descrittiva [4](#)
  - stima delle statistiche [4](#)
  - Test MCAR [4](#)
- Analizza modelli [8](#)
- Assegna valori dei dati mancanti
  - dei vincoli [11](#)
  - metodo di assegnazione [10](#)
  - output [12](#)
- assegnazione multipla
  - Analizza pattern [8](#)
  - Assegna valori dei dati mancanti [9](#)
- Assegnazione multipla [13](#), [14](#)

## C

- cancellazione a coppia
  - in Analisi del valore mancante [1](#)
- cancellazione a livello di lista
  - in Analisi del valore mancante [1](#)
- casi di tabulazione
  - in Analisi del valore mancante [3](#)
- categorie di tabulazione
  - in Analisi del valore mancante [4](#)
- conta valori estremi
  - in Analisi del valore mancante [4](#)
- correlazioni
  - in Analisi del valore mancante [5](#), [6](#)
- covarianza
  - in Analisi del valore mancante [5](#), [6](#)
- cronologia delle iterazioni
  - in Più Imputazione [12](#)

## D

- Dati non completi
  - consultare Analisi del valore mancante [1](#)
- deviazione standard
  - in Analisi del valore mancante [4](#)

## E

- EM
  - in Analisi del valore mancante [5](#)

## I

- imputazione monotone
  - in Più Imputazione [10](#)

## M

- mancata corrispondenza
  - in Analisi del valore mancante [4](#)
- media
  - in Analisi del valore mancante [4-6](#)

## N

- Normovariati
  - in Analisi del valore mancante [6](#)

## O

- ordinamento dei casi
  - in Analisi del valore mancante [3](#)

## R

- regressione
  - in Analisi del valore mancante [6](#)
- residui
  - in Analisi del valore mancante [6](#)

## S

- Specifica completamente condizionale
  - in Più Imputazione [10](#)

## T

- tabella delle frequenze
  - in Analisi del valore mancante [4](#)
- Test MCAR
  - in Analisi del valore mancante [1](#)
- Test MCAR di Little
  - in Analisi del valore mancante [1](#)
- test t
  - in Analisi del valore mancante [4](#)
- test t di Student
  - in Analisi del valore mancante [6](#)

## V

- valori mancanti
  - statistica univariata [4](#)
- Variabili indicatore
  - in Analisi del valore mancante [4](#)
- variabili indicatore mancanti
  - in Analisi del valore mancante [4](#)





