

IBM SPSS Modeler 14.2 Nodi di input,
elaborazione e output



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

Il presente documento contiene informazioni di proprietà di SPSS Inc, una società IBM. Viene fornito su contratto di licenza ed è protetto dalle leggi sul copyright. Le informazioni contenute nella presente pubblicazione non includono alcuna garanzia di prodotto e qualsiasi dichiarazione inclusa nel presente manuale non deve essere interpretata come tale.

Al momento dell'invio di informazioni a IBM o SPSS, l'utente concede a IBM e SPSS un diritto non esclusivo a utilizzare o distribuire tali informazioni nel modo che ritiene appropriato senza incorrere in alcun obbligo nei confronti dell'utente.

© **Copyright IBM Corporation 1994, 2011..**

Prefazione

IBM® SPSS® Modeler è l'efficace workbench di data mining aziendale di IBM Corp.. SPSS Modeler consente alle organizzazioni di migliorare le relazioni con i clienti e con il pubblico grazie a un'analisi approfondita dei dati. Le organizzazioni potranno utilizzare le informazioni ottenute tramite SPSS Modeler per mantenere i clienti di valore, cogliere opportunità di vendite incrociate, attrarre nuovi clienti, individuare frodi, diminuire i rischi e migliorare l'offerta di servizi a livello statale.

L'interfaccia visiva di SPSS Modeler favorisce l'applicazione di una competenza aziendale specifica da parte degli utenti, grazie alla quale sarà possibile ottenere modelli di previsione più efficaci e una riduzione nei tempi di sviluppo delle soluzioni. SPSS Modeler offre una vasta gamma di tecniche di creazione di modelli, quali previsione, classificazione, segmentazione e algoritmi per l'individuazione delle associazioni. IBM® SPSS® Modeler Solution Publisher consente quindi di distribuire a livello aziendale i modelli creati in modo che vengano utilizzati dai responsabili dei processi decisionali oppure inseriti in un database.

Informazioni su IBM Business Analytics

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

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

Supporto tecnico

Il supporto tecnico è a disposizione dei clienti che dispongono di un contratto di manutenzione. I clienti possono contattare il supporto tecnico per richiedere assistenza per l'utilizzo di IBM Corp. o per l'installazione di uno degli ambienti hardware supportati. Per contattare il supporto tecnico, visitare il sito Web IBM Corp. all'indirizzo <http://www.ibm.com/support>. Ricordare che durante la richiesta di assistenza sarà necessario fornire i dati di identificazione personali, i dati relativi alla propria società e il numero del contratto di manutenzione.

Contenuto

1	Informazioni su IBM SPSS Modeler	1
	IBM SPSS Modeler Server	1
	Opzioni di IBM SPSS Modeler.	2
	IBM SPSS Text Analytics	2
	Documentazione di IBM SPSS Modeler	2
	Esempi di applicazioni.	4
	Cartella Demos.	5
2	Nodi Input	6
	Osservazioni generali	6
	Nodo Enterprise View	7
	Impostazione delle opzioni relative al nodo Enterprise View.	8
	Connessioni Enterprise View	10
	Scelta della Definizione dei provider di dati	11
	Scelta della tabella	12
	Nodo di input Database	13
	Impostazione delle opzioni relative al nodo Database.	14
	Aggiunta di una connessione a un database	16
	Specifica di valori preimpostati per una connessione a un database	17
	Selezione di una tabella da un database	18
	Invio di query al database	19
	Nodo Testo variabile.	20
	Impostazione delle opzioni relative al nodo Testo variabile.	22
	Nodo Testo fisso.	24
	Impostazione delle opzioni relative al nodo Testo fisso.	24
	Impostazione dell'archiviazione e della formattazione dei campi.	27
	Nodo Data Collection	31
	Opzioni scheda File del nodo Importazione di Data Collection.	31
	Proprietà dei metadati dell'importazione di IBM SPSS Data Collection.	35
	Stringa di connessione al database	36
	Proprietà avanzate	37
	Importazione di insiemi di risposte multiple	37
	Note sull'importazione di colonne in IBM SPSS Data Collection	38
	Nodo di input IBM Cognos BI	39
	Connessioni Cognos	41
	Selezione dei pacchetti Cognos.	41

Nodo di input SAS	42
Impostazione delle opzioni relative al nodo di input SAS.	43
Nodo di input Excel.	44
XML Source Node	45
Selezione da più elementi radice	47
Rimozione degli spazi indesiderati dai dati di origine XML	48
Nodo input utente.	50
Impostazioni relative al nodo Input utente	51
Schede comuni a tutti i nodi di input	56
Impostazione dei livelli di misurazione nel nodo di input	56
Filtro di campi dal nodo Input.	58

3 Nodi Operazioni su record 60

Panoramica sui nodi Operazioni su record	60
Nodo Seleziona	61
Nodo Campione	63
Opzioni del nodo Campione	64
Impostazioni di raggruppamento e stratificazione	68
Dimensioni dei campioni per gli strati.	69
Nodo Bilanciamento	71
Impostazione delle opzioni per il nodo Bilanciamento.	72
Nodo Aggregazione	73
Impostazione delle opzioni per il nodo Aggregazione	74
Nodo Aggregazione RFM	75
Impostazione delle opzioni per il nodo Aggregazione RFM	77
Nodo Ordina	78
Impostazioni di ottimizzazione dell'ordinamento	79
Nodo Unione	80
Tipi di join	81
Definizione di chiavi e metodo di unione.	83
Selezione di dati per join parziali	84
Filtro di campi dal nodo Unione	85
Impostazione di tag e ordine dei dati di input	86
Impostazioni di ottimizzazione dell'unione	88
Nodo Accodamento	90
Impostazione delle opzioni relative al nodo Accodamento	91
Nodo Elimina duplicati	92
Impostazioni di ottimizzazione del nodo Elimina duplicati	94

4 Nodi Operazioni su campi

96

Panoramica sui nodi Operazioni sui campi	96
Preparazione automatica dati	98
Scheda Campi	101
Scheda Impostazioni	101
Impostazioni campo	102
Prepara date e ore	103
Escludi campi	104
Preparazione degli input e degli obiettivi	105
Creazione e selezione delle funzioni	107
Nomi campi	109
Scheda Analisi	110
Riepilogo elaborazione campi	112
Campi	113
Riepilogo delle azioni	115
Potere predittivo	116
Tabella Campi	117
Dettagli campo	118
Dettagli dell'azione	120
Generazione di un nodo Nuovo campo	123
Nodo Tipo	125
Livelli di misurazione	127
Conversione di dati continui	129
Che cos'è l'istanziamento?	130
Valori dei dati	131
Definizione dei valori mancanti	137
Controllo dei valori del tipo	137
Impostazione del ruolo del campo	138
Copia degli attributi del tipo	140
Scheda per l'impostazione dei formati dei campi	141
Applicazione di filtri o ridenominazione di campi	143
Impostazione delle opzioni di filtro	144
Nodo Risultati classificatore binario	150
Impostazioni del nodo Risultati classificatore binario	151
Nuovo campo	154
Impostazione delle opzioni di base relative al nodo Crea variabile	155
Derivazione di campi multipli	156
Impostazione delle opzioni relative al nodo Crea variabile formula	158
Impostazione delle opzioni relative al nodo Nuovo campo flag	159
Impostazione delle opzioni relative al nodo Crea variabile insieme	161
Impostazione delle opzioni relative al nodo Nuovo campo stato	162

Impostazione delle opzioni relative al nodo Crea variabile conteggio	164
Impostazione delle opzioni relative al nodo Nuovo campo condizionale	165
Ricodifica di valori con il nodo Nuovo campo	165
Nodo Riempimento	166
Conversione della classe di archiviazione mediante il nodo Riempimento.	168
Nodo Anonimizza	169
Impostazione delle opzioni relative al nodo Anonimizza	170
Anonimizzazione dei valori dei campi	172
Nodo Ricodifica	174
Impostazione delle opzioni per il nodo Ricodifica	174
Ricodifica di più campi.	177
Classe di archiviazione e livello di misurazione per i campi ricodificati.	178
Nodo Discretizza	178
Impostazione delle opzioni relative al nodo Discretizza	179
Bin a larghezza fissa	180
Blocchi (conteggio o somma uguale)	181
Classifica casi	184
Deviazione standard/Media	185
Categorizzazione ottimale	186
Anteprima dei bin generati.	188
Nodo Analisi RFM.	189
Impostazioni del nodo Analisi RFM.	190
Discretizzazione con il nodo Analisi RFM	192
Nodo Partizione	193
Opzioni del nodo Partizione	194
Nodo Crea flag	196
Impostazione delle opzioni relative al nodo Crea flag	196
Nodo Riorganizza	197
Impostazione delle opzioni relative al nodo Riorganizza	199
Nodo Trasponi	200
Impostazione delle opzioni del nodo Trasponi.	200
Nodo Intervalli di tempo	204
Indicazione degli intervalli di tempo	205
Opzioni di creazione degli intervalli di tempo	207
Periodo di stima	209
Previsioni	210
Intervalli supportati	213
Nodo Cronologia.	225
Impostazione delle opzioni relative al nodo Cronologia.	226
Nodo Ordina campi.	227
Impostazione delle opzioni Ordina campi.	227

5 *Nodi Grafico*

231

Caratteristiche comuni dei nodi Grafico	231
Effetti estetici, sovrapposizioni, riquadri e animazioni.	232
Utilizzo della scheda Output.	237
Utilizzo della scheda Annotazioni.	237
Grafici 3-D.	238
Nodo Lavagna grafica	239
Scheda Di base del nodo Lavagna grafica	240
Lavagna grafica Scheda Dettagliato	245
Scheda Aspetto del nodo Lavagna grafica	247
Impostazione della posizione di modelli, fogli di stile e mappe	249
Gestione dei modelli, dei fogli di stile e dei file delle mappe	251
Nodo Plot	252
Scheda Nodo Plot	255
Scheda Opzioni del nodo Plot.	258
Scheda Aspetto del nodo Plot	259
Utilizzo del grafico di un nodo Plot	260
Nodo Distribuzione	261
Scheda Plot del nodo Distribuzione	262
Scheda Aspetto del nodo Distribuzione	263
Utilizzo di un nodo Distribuzione.	264
Nodo Istogramma.	266
Scheda Plot del nodo Istogramma	267
Scheda Opzioni del nodo Istogramma	268
Scheda Aspetto del nodo Istogramma	269
Utilizzo degli istogrammi	269
Nodo Raccolta	270
Scheda Plot del nodo Raccolta	271
Scheda Opzioni del nodo Raccolta	272
Scheda Aspetto del nodo Raccolta	273
Utilizzo del grafico di un nodo Raccolta	274
Nodo Plot multiplo	275
Scheda Plot di un nodo Plot multiplo	276
Scheda Aspetto del nodo Plot multiplo.	278
Utilizzo del grafico di un nodo Plot multiplo	279
Nodo Web	280
Scheda Plot del nodo Web.	281
Scheda Opzioni del nodo Web	283
Scheda Aspetto del nodo Web.	285
Utilizzo di un grafico Web.	286

Nodo Plot tempo	291
Scheda Plot del nodo Plot tempo	292
Scheda Aspetto del nodo Plot tempo	293
Utilizzo del grafico Plot tempo	294
Nodo Valutazione	295
Scheda Plot del nodo Valutazione	301
Scheda Opzioni del nodo Valutazione	302
Scheda Aspetto del nodo Valutazione	304
Lettura dei risultati di un grafico di valutazione relativo a un modello	306
Utilizzo di un grafico di valutazione	308
Esplorazione dei grafici.	309
Utilizzo delle sezioni.	310
Utilizzo delle aree	314
Utilizzo degli elementi contrassegnati	317
Generazioni di nodi da grafici.	319
Modifica delle visualizzazioni	321
Regole generali per la modifica delle visualizzazioni.	322
Modifica e formattazione del testo	323
Modifica dei colori, dei motivi, dei tratteggi e della trasparenza	324
Rotazione e modifica della forma e delle proporzioni degli elementi puntiformi.	326
Modifica delle dimensioni degli elementi grafici.	326
Impostazione dei margini e della spaziatura	327
Formattazione dei numeri.	328
Modifica delle impostazioni degli assi e delle scale	329
Modifica delle categorie	331
Modifica dei riquadri di orientamento	333
Trasformazione del sistema di coordinate	333
Modifica di statistiche ed elementi grafici	335
Modifica della posizione della legenda	338
Copia di una visualizzazione e dei dati delle visualizzazioni.	338
Tasti di scelta rapida	339
Aggiunta di titoli e annotazioni	339
Utilizzo dei fogli di stile dei grafici	341
Applicazione di fogli di stile	342
Stampa, salvataggio, copia ed esportazione di grafici.	344

6 Nodi Output

346

Panoramica sui nodi Output	346
Gestione dell'output	347

Visualizzazione dell'output	348
Pubblica sul Web.	348
Visualizzazione dell'output in un browser HTML.	351
Esportazione dell'output.	351
Selezione di celle e di colonne.	352
Nodo Tabella	353
Scheda Impostazioni del nodo Tabella	354
Scheda Formato del nodo Tabella	354
Scheda Output del nodo Output	355
Browser del nodo Tabella	357
Nodo Matrice	359
Scheda Impostazioni del nodo Matrice	359
Scheda Aspetto del nodo Matrice	360
Browser di output del nodo Matrice.	362
nodo Analisi	364
Scheda Analisi del nodo Analisi.	364
Browser di output del nodo Analisi	367
Nodo Esplora	369
Scheda Impostazioni del nodo Esplora.	370
Nodo Esplora, scheda Qualità	372
Browser di output del nodo Esplora	374
Nodo Trasformazioni.	383
Scheda Opzioni del nodo Trasformazioni	384
Scheda Output del nodo Trasformazioni.	385
Visualizzatore output del nodo Trasformazioni	385
Nodo Statistiche.	389
Scheda Impostazioni del nodo Statistiche	389
Browser di output del nodo Statistiche	391
Nodo Medie	393
Confronto di medie per gruppi indipendenti	394
Confronto delle medie tra campi accoppiati	395
Opzioni del nodo Medie	396
Browser di output del nodo Medie	397
Nodo Report.	400
Scheda Modello del nodo Report.	401
Browser di output del nodo Report.	403
Nodo Globali.	403
Scheda Impostazioni del nodo Calcola globali	404
IBM SPSS Statistics Applicazioni di supporto	405

7 Nodi di esportazione

407

Panoramica sui nodi di esportazione	407
Nodo di esportazione Database	408
Scheda di esportazione del nodo Database	408
Opzioni di unione esportazione database	410
Esportazione Database - Opzioni Schema	412
Esportazione database - Opzioni Indice	414
Esportazione database - Opzioni avanzate	417
Nodo di esportazione File piatto	419
Scheda di esportazione del nodo File piatto	419
Nodo Esporta IBM SPSS Data Collection	420
Nodo di esportazione IBM Cognos BI	422
Connessione Cognos	422
Connessione ODBC	424
Nodo Esporta SAS	426
Scheda Esporta del nodo Esporta SAS	426
Nodo Esportazione da Excel	427
Scheda di esportazione del nodo Excel	427
Nodo Esporta XML	428
Scrittura dei dati XML	429
Opzioni dei record di mappatura XML	430
Opzioni dei campi di mappatura XML	431
Anteprima mappatura XML	432

8 Nodi IBM SPSS Statistics

433

Nodi IBM SPSS Statistics - Panoramica	433
Nodo File Statistics	434
Nodo Trasformazioni Statistics	435
Scheda Sintassi del nodo Trasformazioni Statistics	436
Sintassi consentita	438
Nodo Modello Statistics	440
Scheda Modello del nodo Modello Statistics	441
Riepilogo dell'insieme di modelli del nodo Modello Statistics	442
Nodo Output Statistics	444
Scheda Sintassi del nodo Output Statistics	445
Scheda Output del nodo Output Statistics	448

Nodo Esporta Statistics	449
Scheda Esporta del nodo Esporta Statistics	450
Ridenominazione o filtro dei campi per IBM SPSS Statistics	451

9 Supernodi 453

Panoramica sui Supernodi	453
Tipi di Supernodi	453
Supernodi di input	454
Supernodi di elaborazione	454
Supernodo finale	455
Creazione di Supernodi	456
Nidificazione di Supernodi	458
Esempio di Supernodi validi	459
Esempi di Supernodi non validi	460
Blocco di Supernodi	461
Blocco e sblocco di Supernodi	462
Modifica di un Supernodo bloccato	463
Modifica Supernodo	464
Modifica dei tipi di Supernodo	464
Annotazione e ridenominazione dei Supernodi	465
Parametri dei Supernodi	466
Supernodi e cache	470
Supernodi e script	471
Salvataggio e caricamento dei Supernodi Supernodi	472

Appendice

A Note 474

Indice 477

Informazioni su IBM SPSS Modeler

IBM® SPSS® Modeler è un insieme di strumenti di data mining che consente di sviluppare rapidamente modelli predittivi con l'ausilio di competenze aziendali e di eseguirne il deployment nelle operazioni aziendali per migliorare i processi decisionali. Progettato secondo il modello CRISP-DM conforme agli standard di settore, SPSS Modeler supporta l'intero processo di data mining, dai dati a risultati aziendali migliori.

SPSS Modeler offre numerosi metodi di modellazione ricavati dall'apprendimento automatico, dall'intelligenza artificiale e dalla statistica. I metodi disponibili nella palette Modelli consentono di ricavare nuove informazioni dai dati e di sviluppare modelli predittivi. Ogni metodo ha determinati punti di forza e si presta meglio per particolari tipi di problemi.

SPSS Modeler può essere acquistato come prodotto autonomo oppure utilizzato in combinazione con SPSS Modeler Server. È inoltre disponibile una serie di opzioni, come illustrato nelle sezioni seguenti. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/modeler/>.

IBM SPSS Modeler Server

SPSS Modeler utilizza un'architettura client/server per distribuire le richieste di operazioni che utilizzano molte risorse a potenti componenti software server, con un conseguente miglioramento della performance su insiemi di dati di grandi dimensioni. Oltre a quelli elencati di seguito, potrebbero essere disponibili altri prodotti e aggiornamenti. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/modeler/>.

SPSS Modeler. SPSS Modeler è una versione del prodotto completa nelle funzioni, che viene installata ed eseguita sul desktop dell'utente. È possibile eseguirla in modalità locale come prodotto autonomo oppure in modalità distribuita assieme a IBM® SPSS® Modeler Server per ottenere una migliore performance su insiemi di dati di grandi dimensioni.

SPSS Modeler Server. SPSS Modeler Server viene eseguito continuamente in modalità di analisi distribuita assieme a una o più installazioni di IBM® SPSS® Modeler. Una configurazione di questo tipo consente di ottenere performance superiori quando si lavora su insiemi di dati di grandi dimensioni, in quanto le operazioni che richiedono un utilizzo consistente della memoria possono essere eseguite sul server senza necessità di scaricare i dati sul computer client. SPSS Modeler Server supporta inoltre funzionalità di ottimizzazione SQL e modellazione in-database, con ulteriori vantaggi in termini di prestazioni e grado di automazione. Per eseguire un'analisi deve essere disponibile almeno un'installazione SPSS Modeler.

Opzioni di IBM SPSS Modeler

I seguenti componenti e funzionalità possono essere acquistati separatamente e concessi in licenza per l'utilizzo con SPSS Modeler. Tenere presente che potrebbero rendersi disponibili anche prodotti e aggiornamenti aggiuntivi. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/modeler/>.

- Accesso a SPSS Modeler Server per una maggiore scalabilità e prestazioni migliori su insiemi di dati di grandi dimensioni, nonché per il supporto di funzionalità di ottimizzazione SQL e modellazione in-database.
- SPSS Modeler Solution Publisher, per il calcolo automatico o in tempo reale del punteggio al di fuori dall'ambiente di SPSS Modeler. Per ulteriori informazioni, vedere l'argomento [IBM SPSS Modeler Solution Publisher in il capitolo 2 in IBM SPSS Modeler 14.2 Solution Publisher](#).
- Adattatori per il deployment in IBM SPSS Collaboration and Deployment Services o nell'applicazione thin-client IBM SPSS Modeler Advantage. Per ulteriori informazioni, vedere l'argomento [Archiviazione e deployment di oggetti di IBM SPSS Collaboration and Deployment Services Repository in il capitolo 9 in Manuale dell'utente di IBM SPSS Modeler 14.2](#).

IBM SPSS Text Analytics

IBM® SPSS® Text Analytics è un componente aggiuntivo completamente integrato per SPSS Modeler che utilizza tecnologie linguistiche avanzate e di Natural Language Processing (NLP) per elaborare rapidamente una grande varietà di dati di testo non strutturati, estrarre e organizzare i concetti chiave e raggruppare questi concetti in categorie. È quindi possibile combinare i concetti e le categorie estratti con dati strutturati esistenti, per esempio dati demografici, e applicarli alla modellazione utilizzando la suite completa degli strumenti di data mining di IBM® SPSS® Modeler per prendere decisioni migliori e più mirate.

- Il nodo Text Mining offre funzioni di modellazione di concetti e categorie nonché un workbench interattivo in cui è possibile eseguire esplorazioni avanzate di collegamenti e cluster di testo, creare categorie personalizzate e rifinire i modelli di risorse linguistiche.
- Sono supportati numerosi formati di importazione, inclusi blog e altre sorgenti basate sul Web.
- Sono inclusi inoltre modelli, librerie e dizionari personalizzati per domini specifici, quali CRM e genomica.

Nota: per l'accesso a questo componente, è necessaria una licenza distinta. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/modeler/>.

Documentazione di IBM SPSS Modeler

La documentazione completa nel formato guida in linea è disponibile nel menu Aiuto di SPSS Modeler. Sono incluse la documentazione per SPSS Modeler, SPSS Modeler Server e SPSS Modeler Solution Publisher, nonché la Guida alle applicazioni e altro materiale di supporto.

La documentazione completa in formato PDF dei singoli prodotti è disponibile nella cartella *Documentation* del DVD di ciascun prodotto.

- **Manuale dell'utente di IBM SPSS Modeler.** Introduzione generale all'utilizzo di SPSS Modeler che illustra come creare stream di dati, gestire valori mancanti, generare espressioni CLEM, utilizzare progetti e report e assemblare stream per il deployment tramite IBM SPSS Collaboration and Deployment Services, le applicazioni predittive o IBM SPSS Modeler Advantage.
- **IBM SPSS Modeler Source, Process, and Output Nodes.** Descrizioni di tutti i nodi utilizzati per leggere, elaborare e generare dati di output in vari formati, ovvero di nodi ad eccezione dei nodi Modelli.
- **IBM SPSS Modeler Nodi Modelli.** Descrizioni di tutti i nodi utilizzati per creare modelli di data mining. IBM® SPSS® Modeler offre numerosi metodi di modellazione ricavati dall'apprendimento automatico, dall'intelligenza artificiale e dalla statistica. [Per ulteriori informazioni, vedere l'argomento Panoramica sui nodi Modelli in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)
- **IBM SPSS Modeler Algorithms Guide.** Descrizione dei fondamenti di matematica per i metodi di modellazione utilizzati in SPSS Modeler.
- **IBM SPSS Modeler Guida alle applicazioni.** Gli esempi inclusi in questa guida forniscono indicazioni mirate e sintetiche su specifici metodi e tecniche di modellazione. Una versione in linea di questa guida è inoltre disponibile dal menu Aiuto. [Per ulteriori informazioni, vedere l'argomento Esempi di applicazioni in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **IBM SPSS Modeler Script e automazione.** Informazioni sulle modalità di automazione del sistema tramite script, incluse le proprietà che è possibile utilizzare per manipolare nodi e stream.
- **IBM SPSS Modeler Deployment Guide.** Informazioni sull'esecuzione di stream e scenari SPSS Modeler come fasi dell'elaborazione di lavori in IBM® SPSS® Collaboration and Deployment Services Deployment Manager.
- **IBM SPSS Modeler Guida per lo sviluppatore CLEF.** CLEF consente di integrare programmi di terze parti (quali routine di elaborazione di dati o algoritmi di modellazione) come nodi in SPSS Modeler.
- **IBM SPSS Modeler Guida alla modellazione in-database.** Informazioni sulle modalità per utilizzare al meglio la potenza del database in uso al fine di ottenere prestazioni migliori ed estendere la gamma di funzionalità analitiche tramite algoritmi di terze parti.
- **IBM SPSS Modeler Server e Guida della performance.** Informazioni su come configurare e amministrare IBM® SPSS® Modeler Server.
- **Manuale dell'utente di IBM SPSS Modeler Administration Console.** Informazioni sull'installazione e l'utilizzo dell'interfaccia utente della console per il monitoraggio e la configurazione di SPSS Modeler Server. La console viene implementata come plug-in dell'applicazione Deployment Manager.
- **IBM SPSS Modeler Solution Publisher Guide.** SPSS Modeler Solution Publisher è un componente aggiuntivo che consente di pubblicare gli stream al di fuori dell'ambiente SPSS Modeler standard.
- **Guida CRISP-DM di IBM SPSS Modeler.** Guida passo a passo al data mining tramite la metodologia CRISP-DM con SPSS Modeler.

Esempi di applicazioni

Mentre gli strumenti per il data mining di SPSS Modeler consentono di risolvere un'ampia gamma di problemi a livello aziendale e organizzativo, gli esempi di applicazioni forniscono indicazioni mirate e sintetiche su specifici metodi e tecniche di modellazione. Gli insiemi di dati utilizzati negli esempi hanno dimensioni molto più limitate rispetto agli enormi archivi di dati gestiti da alcuni data miner, ma i concetti e i metodi coinvolti sono rapportabili alle applicazioni del mondo reale.

È possibile accedere agli esempi facendo clic su Esempi di applicazioni nel menu Aiuto di SPSS Modeler. I file di dati e gli stream di esempio sono installati nella cartella *Demos* nella directory di installazione del prodotto. [Per ulteriori informazioni, vedere l'argomento Cartella Demos in *Manuale dell'utente di IBM SPSS Modeler 14.2*.](#)

Esempi di modellazione in-database. Vedere gli esempi nella *IBM SPSS Modeler Guida alla modellazione in-database*.

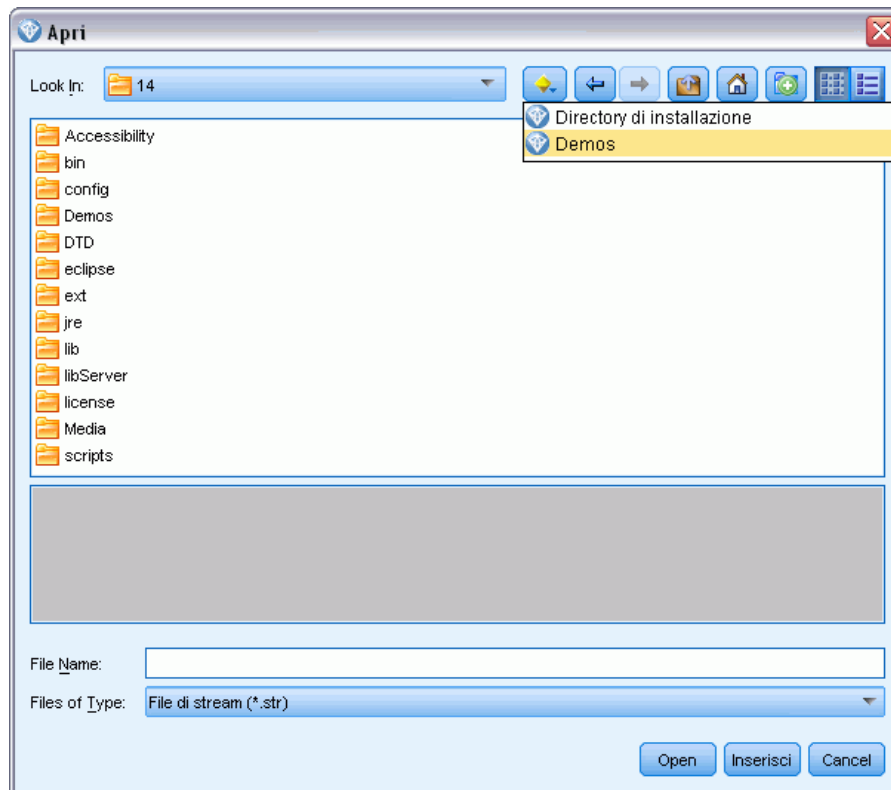
Esempi di script. Vedere gli esempi nella *IBM SPSS Modeler Guida per script e automazione*.

Cartella Demos

I file di dati e gli stream di esempio utilizzati negli esempi di applicazioni sono installati nella cartella *Demos* nella directory di installazione del prodotto. A questa cartella è possibile accedere anche dal gruppo di programmi IBM SPSS Modeler 14.2 nel menu Start di Windows oppure facendo clic su *Demos* nell'elenco delle directory recenti nella finestra di dialogo Apri file.

Figura 1-1

Selezione della cartella *Demos* dall'elenco delle directory utilizzate di recente



Nodi Input

Osservazioni generali

I nodi di input consentono di importare i dati archiviati in numerosi formati, inclusi file piatti, IBM® SPSS® Statistics (.sav), SAS, Microsoft Excel e database relazionali compatibili con ODBC. Tramite il nodo Input utente è inoltre possibile generare dati sintetici.

La palette Input contiene i seguenti nodi:



Il nodo Enterprise View crea una connessione a un IBM® SPSS® Collaboration and Deployment Services Repository, consentendo di leggere i dati di Enterprise View in uno stream e di assemblare un modello in uno scenario accessibile da altri utenti tramite il repository. [Per ulteriori informazioni, vedere l'argomento Nodo Enterprise View a pag. 7.](#)



Il nodo Database può essere utilizzato per importare dati da numerosi altri pacchetti che utilizzano ODBC (Open Database Connectivity), tra cui Microsoft SQL Server, DB2, Oracle e altri. [Per ulteriori informazioni, vedere l'argomento Nodo di input Database a pag. 13.](#)



Il nodo Testo variabile legge dati da file di testo a campi liberi, ovvero file i cui record contengono un numero costante di campi ma un numero variabile di caratteri. Questo nodo può essere utilizzato per file con testo di intestazione a lunghezza fissa e alcuni tipi di annotazioni. [Per ulteriori informazioni, vedere l'argomento Nodo Testo variabile a pag. 20.](#)



Il nodo Testo fisso importa dati da file di testo a campi fissi, ovvero file i cui campi non vengono delimitati ma iniziano nella stessa posizione e hanno una lunghezza fissa. Nel formato a campi fissi vengono in genere archiviati dati di versioni precedenti o generati dalla macchina. [Per ulteriori informazioni, vedere l'argomento Nodo Testo fisso a pag. 24.](#)



Il nodo File Statistics legge i dati dal formato di file .sav utilizzato da SPSS Statistics, nonché da file della cache salvati in IBM® SPSS® Modeler, che utilizzano lo stesso formato. [Per ulteriori informazioni, vedere l'argomento Nodo File Statistics in il capitolo 8 a pag. 434.](#)



Il nodo IBM® SPSS® Data Collection importa i dati dei sondaggi dai vari formati utilizzati dai software per ricerche di mercato conformemente al Modello dati di Data Collection. Per utilizzare questo nodo, è necessario che sia installata Data Collection Developer Library. [Per ulteriori informazioni, vedere l'argomento Nodo Data Collection a pag. 31.](#)



Il nodo di input IBM Cognos BI importa i dati dai database Cognos BI. [Per ulteriori informazioni, vedere l'argomento Nodo di input IBM Cognos BI a pag. 39.](#)



Il nodo File SAS importa dati SAS in SPSS Modeler. [Per ulteriori informazioni, vedere l'argomento Nodo di input SAS a pag. 42.](#)



Il nodo Excel importa dati da qualsiasi versione di Microsoft Excel. Non è richiesta alcuna sorgente dati ODBC. [Per ulteriori informazioni, vedere l'argomento Nodo di input Excel a pag. 44.](#)



Il nodo di input XML importa i dati in formato XML nello stream. È possibile importare un singolo file o tutti i file in una directory. Se lo si desidera, è possibile specificare un file di schema dal quale leggere la struttura XML. [Per ulteriori informazioni, vedere l'argomento XML Source Node a pag. 45.](#)



Il nodo Input utente consente di ottenere in modo semplice dati sintetici creandoli oppure modificando dati esistenti. È utile, per esempio, quando si desidera creare un insieme di dati di test per la modellazione. [Per ulteriori informazioni, vedere l'argomento Nodo input utente a pag. 50.](#)

Per iniziare uno stream, aggiungere un nodo di input nell'area di disegno dello stream. Quindi fare doppio clic sul nodo per visualizzare la relativa finestra di dialogo. Le varie schede della finestra di dialogo consentono di leggere dati, visualizzare campi e valori e impostare numerose opzioni relative per esempio a filtri, tipi di dati, ruoli dei campi e controllo dei valori mancanti.

Nodo Enterprise View

Il nodo Enterprise View consente di creare e di mantenere una connessione tra una sessione di IBM® SPSS® Modeler e una Enterprise View in un IBM® SPSS® Collaboration and Deployment Services Repository condiviso. Ciò consente di leggere dati da una Enterprise View in uno stream SPSS Modeler e di assemblare un modello SPSS Modeler in uno scenario a cui possono accedere altri utenti del repository condiviso.

Uno **scenario** è un file contenente uno stream di SPSS Modeler con determinati nodi, modelli e proprietà aggiuntive che ne consentono il deployment in IBM SPSS Collaboration and Deployment Services Repository ai fini del calcolo del punteggio o per l'aggiornamento automatico di modelli. L'uso di nodi Enterprise View con scenari garantisce che, in una situazione multi-utente, tutti gli utenti lavorino sugli stessi dati. Una **connessione** è un collegamento da una sessione di SPSS Modeler a una Enterprise View nel IBM SPSS Collaboration and Deployment Services Repository.

L'**Enterprise View** è l'insieme completo di dati appartenenti a un'organizzazione, indipendentemente dalla posizione fisica in cui risiedono i dati. Ogni connessione è costituita da una selezione specifica di un'unica **Visualizzazione applicazione** (sottoinsieme di Enterprise View su misura per una particolare applicazione), una **Definizione dei provider di dati** (DPD—collega le tabelle e le colonne della Visualizzazione applicazione con la sorgenti dati fisica) e un **ambiente** (identifica quali specifiche colonne devono essere associate ai segmenti aziendali definiti). Le definizioni Enterprise View, Visualizzazioni applicazione e DPD vengono archiviate nel repository, sebbene i dati effettivi risiedano in uno o più database o in altre sorgenti esterne.

Quando la connessione è stata stabilita, è necessario specificare una **tabella** di Visualizzazione applicazione con cui lavorare in SPSS Modeler. In una Visualizzazione applicazione, una tabella è una visualizzazione logica costituita da alcune o tutte le colonne di una o più tabelle fisiche in uno o più database. Pertanto il nodo Enterprise View consente di visualizzare in SPSS Modeler i record provenienti da più tabelle di database come un'unica tabella.

Requisiti

- Per utilizzare il nodo Enterprise View, è necessario dapprima installare e configurare nel sito un IBM SPSS Collaboration and Deployment Services Repository, con una Enterprise View, le Visualizzazioni applicazione e le DPD già definite.

Nota: per l'accesso a un repository IBM® SPSS® Collaboration and Deployment Services è necessaria una licenza distinta. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

- È inoltre necessario che IBM® SPSS® Collaboration and Deployment Services Enterprise View Driver sia installato su ogni computer utilizzato per modificare o eseguire lo stream. Per Windows, installare semplicemente il driver sul computer su cui è installato IBM® SPSS® Modeler o IBM® SPSS® Modeler Server. Non sono necessarie ulteriori configurazioni del driver. In UNIX, è necessario aggiungere allo script di avvio un riferimento allo script *pev.sh*. Per ulteriori informazioni, vedere l'argomento [Configurazione di un driver per il nodo Enterprise View in l'appendice B in IBM SPSS Modeler Server 14.2 Guida della performance e amministrazione](#). Per dettagli sull'installazione di IBM SPSS Collaboration and Deployment Services Enterprise View Driver, rivolgersi all'amministratore di sistema.
- Una Definizione dei provider di dati (DPD) viene definita rispetto a una determinata sorgente dati ODBC. Per utilizzare una DPD da SPSS Modeler, è necessario che sull'host di SPSS Modeler sia stata definita una sorgente dati ODBC con lo stesso nome e che si connette allo stesso archivio dati di quella a cui fa riferimento la Definizione dei provider di dati.

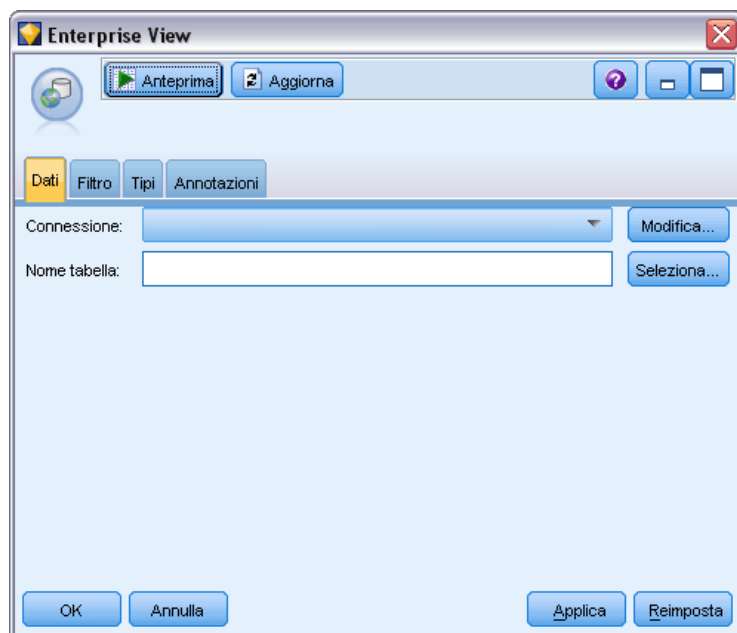
Impostazione delle opzioni relative al nodo Enterprise View

Le opzioni della scheda Dati della finestra di dialogo Enterprise View possono essere utilizzate per:

- selezionare la connessione a un repository esistente
- modificare la connessione a un repository esistente
- creare una nuova connessione a un repository
- selezionare una tabella di Visualizzazione applicazione

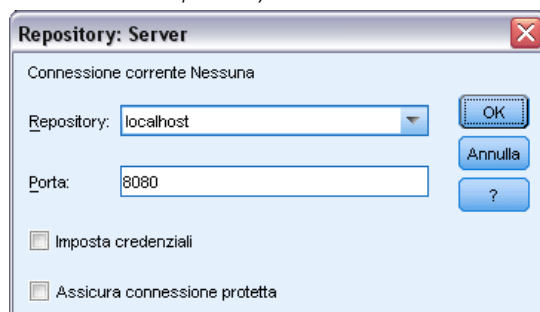
Per ulteriori informazioni su come utilizzare i repository, consultare *IBM® SPSS® Collaboration and Deployment Services Administrator's Guide*.

Figura 2-1
Aggiunta di una connessione a un IBM SPSS Collaboration and Deployment Services Repository



Connessione. L'elenco a discesa contiene le opzioni per la selezione della connessione a un repository esistente, la modifica di una connessione esistente o l'aggiunta di una connessione. Se si è già connessi a un repository tramite IBM® SPSS® Modeler, scegliendo l'opzione Aggiungi/modifica una connessione viene visualizzata la finestra di dialogo Connessioni Enterprise View che consente di definire o modificare i dettagli necessari per la connessione corrente. Se non si è connessi, questa opzione visualizza la finestra di dialogo Accesso del repository.

Figura 2-2
Accesso a un repository



Per informazioni su come accedere al repository, vedere il *Manuale dell'utente di SPSS Modeler*.

Dopo aver stabilito una connessione a un repository, tale connessione rimane attiva finché non si esce da SPSS Modeler. È possibile condividere una connessione con altri nodi all'interno dello stesso stream, ma è necessario creare una nuova connessione per ogni nuovo stream.

Dopo un accesso riuscito, viene visualizzata la finestra di dialogo Connessioni Enterprise View.

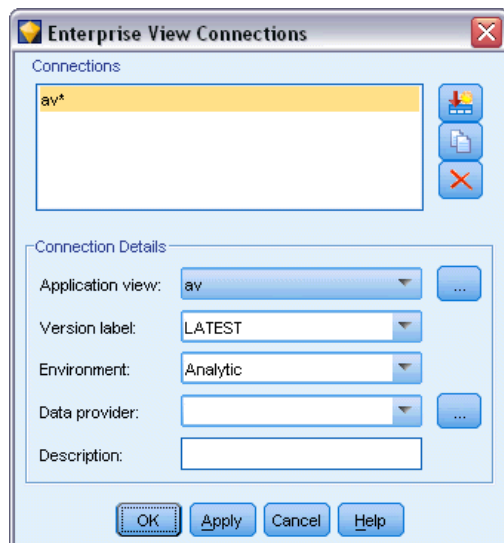
Nome tabella. All'inizio questo campo è vuoto e non è possibile compilarlo fino a quando non si crea una connessione. Se si conosce il nome della tabella di Visualizzazione applicazione a cui si desidera accedere, digitarlo nel campo Nome tabella. In alternativa, fare clic sul pulsante Seleziona per aprire una finestra di dialogo contenente l'elenco delle tabelle di Visualizzazione applicazione disponibili.

Connessioni Enterprise View

Questa finestra di dialogo consente di definire o di modificare i dati necessari per la connessione al repository. È possibile specificare i seguenti dati:

- Visualizzazione applicazione e versione
- Ambiente
- Definizione dei provider di dati (DPD)
- Descrizione della connessione

Figura 2-3
Scelta di una visualizzazione applicazione



Connessioni. Elenca le connessioni ai repository esistenti.

- **Aggiungi una nuova connessione.** Visualizza la finestra di dialogo Recupera oggetto che consente di cercare e selezionare una Visualizzazione applicazione nel repository.
- **Copia la connessione selezionata.** Crea una copia della connessione selezionata, evitando all'utente di dover nuovamente cercare la stessa Visualizzazione applicazione.
- **Elimina la connessione selezionata.** Elimina dall'elenco la connessione selezionata.

Dettagli di connessione. Per la connessione attualmente selezionata nel riquadro Connessioni, visualizza la Visualizzazione applicazione, l'etichetta della versione, l'ambiente, la Definizione dei provider di dati e un testo descrittivo.

- **Visualizzazione applicazione.** In questo elenco a discesa viene visualizzata l'eventuale Visualizzazione applicazione selezionata. Se nella sessione corrente sono state stabilite connessioni ad altre visualizzazioni applicazione, nell'elenco a discesa saranno visualizzate anche queste. Fare clic sul pulsante Visualizza adiacente per cercare altre Visualizzazioni applicazione nel repository.
- **Etichetta versione.** L'elenco a discesa mostra tutte le etichette delle versioni definite per la Visualizzazione applicazione specificata. Le etichette delle versioni facilitano l'individuazione di versioni specifiche di oggetti nel repository. Per esempio, potrebbero esistere due versioni di una determinata Visualizzazione applicazione. Utilizzando le etichette si potrebbe impostare l'etichetta TEST per la versione utilizzata nell'ambiente di sviluppo e l'etichetta PRODUZIONE per la versione utilizzata nell'ambiente di produzione. Selezionare un'etichetta appropriata.

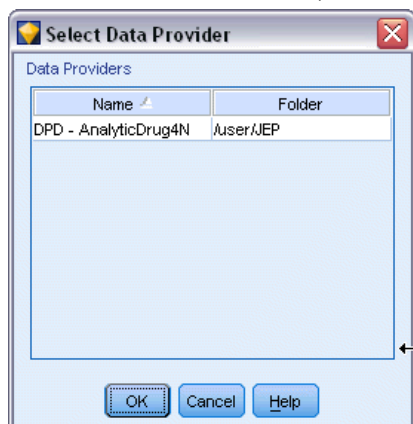
Nota: le etichette non devono includere il carattere “[”, altrimenti il nome della tabella non verrà visualizzato nella scheda Dati della finestra di dialogo Enterprise View.

- **Ambiente.** L'elenco a discesa mostra tutti gli ambienti validi. L'impostazione Ambiente determina quali DPD sono disponibili. Specifica pertanto quali colonne specifiche devono essere associate ai segmenti aziendali definiti. Per esempio, se si seleziona Analitico, vengono restituite solo le colonne di Visualizzazione applicazione definite come Analitiche. L'ambiente di default è Analitico; è possibile anche scegliere l'opzione Operativo.
- **Provider di dati.** Nell'elenco a discesa sono visualizzati fino a dieci nomi di Definizioni dei provider di dati per la Visualizzazione applicazione selezionata. Vengono visualizzate solo le Definizioni dei provider di dati che fanno riferimento alla Visualizzazione applicazione selezionata. Fare clic sul pulsante Visualizza adiacente per visualizzare il nome e il percorso di tutte le Definizioni dei provider di dati relative alla Visualizzazione applicazione corrente.
- **Descrizione.** Testo descrittivo relativo alla connessione con il repository. Il testo sarà utilizzato per il nome della connessione: se si fa clic su OK, il testo viene visualizzato nell'elenco a discesa Connessione e nella barra del titolo della finestra di dialogo Enterprise View, nonché come etichetta del nodo Enterprise View nell'area di disegno.

Scelta della Definizione dei provider di dati

La finestra di dialogo Seleziona provider di dati mostra il nome e il percorso di tutte le Definizioni dei provider di dati relative alla Visualizzazione applicazione corrente.

Figura 2-4
Scelta di una Definizione dei provider di dati



Le Visualizzazioni applicazione possono avere più Definizioni dei provider di dati per supportare varie fasi di un progetto. Per esempio, i dati storici utilizzati per generare un modello potrebbero provenire da un database, mentre i dati operativi potrebbero risiedere in un altro.

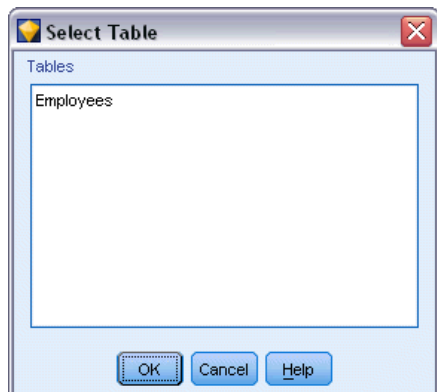
Una Definizione dei provider di dati (DPD) viene definita rispetto a una determinata sorgente dati ODBC. Per utilizzare una DPD da IBM® SPSS® Modeler, è necessario che sull'host di SPSS Modeler sia stata definita una sorgente dati ODBC con lo stesso nome e che si connette allo stesso archivio dati di quella a cui fa riferimento la Definizione dei provider di dati.

- Per scegliere una Definizione dei provider di dati con cui lavorare, selezionarne il nome nell'elenco, quindi fare clic su OK.

Scelta della tabella

La finestra di dialogo Seleziona tabella elenca tutte le tabelle a cui viene fatto riferimento nella Visualizzazione applicazione corrente. La finestra di dialogo è vuota se non è stata effettuata alcuna connessione a un IBM SPSS Collaboration and Deployment Services Repository.

Figura 2-5
Scelta di una tabella



- Per scegliere una tabella con cui lavorare, selezionarne il nome nell'elenco, quindi fare clic su OK.

Nodo di input Database

Il nodo di input Database può essere utilizzato per importare dati da numerosi altri pacchetti che utilizzano ODBC (Open Database Connectivity), tra cui Microsoft SQL Server, DB2, Oracle e altri.

Per leggere o scrivere su un database, occorre che un'origine dati ODBC sia installata e configurata per il database in questione, con le relative autorizzazioni di lettura e scrittura. IBM® SPSS® Data Access Pack contiene una serie di driver ODBC utili a questo scopo; i driver sono disponibili nel disco di installazione di IBM SPSS Data Access Pack fornito con questa versione. Per domande sulla creazione o l'impostazione di autorizzazioni per sorgenti dati ODBC, rivolgersi all'amministratore del database.

Il supporto dei database in IBM® SPSS® Modeler viene classificato in base a tre livelli, ciascuno dei quali rappresenta un diverso livello di supporto per il push back e l'ottimizzazione SQL, a seconda del produttore del database. I diversi livelli di supporto sono implementati tramite un numero di impostazioni di sistema, che possono essere personalizzate nell'ambito di un contratto di servizio.

I tre livelli di supporto database sono i seguenti:

Tabella 2-1

Livelli di supporto database

Livello di supporto	Descrizione
Livello 1	Push back SQL interamente disponibile con ottimizzazione SQL specifica per database.
Livello 2	Push back SQL per gran parte disponibile con ottimizzazione SQL non specifica per database.
Livello 3	Nessuna ottimizzazione o push back disponibile (solo supporto di lettura e scrittura database).

Driver ODBC supportati

Per le ultime informazioni sui database e i driver ODBC che sono supportati e testati per l'uso con SPSS Modeler 14.2, vedere le matrici di compatibilità del prodotto sul sito aziendale dedicato all'assistenza clienti (<http://www.ibm.com/support>).

Dove installare i driver

I driver ODBC devono essere installati e configurati su ogni computer utilizzato.

- Se IBM® SPSS® Modeler viene eseguito in modalità locale (standalone), i driver devono essere installati sul computer locale.
- Se SPSS Modeler viene eseguito in modalità distribuita su un IBM® SPSS® Modeler Server remoto, i driver ODBC devono essere installati nel computer in cui è installato SPSS Modeler Server.
- Se occorre accedere alle stesse origini dati sia da SPSS Modeler che da SPSS Modeler Server, i driver ODBC devono essere installati su entrambi i computer.

- Se SPSS Modeler è in esecuzione su Terminal Services, i driver ODBC devono essere installati sul server Terminal Services in cui è stato installato SPSS Modeler.
- Se si utilizza IBM® SPSS® Modeler Solution Publisher Runtime per eseguire flussi pubblicati su un computer a parte, occorre installare e configurare i driver ODBC anche su questo computer.

Nota: se si utilizza SPSS Modeler Server su UNIX per accedere a un database Teradata, è necessario utilizzare il gestore dei driver ODBC installato con il driver di ODBC Teradata. Per poter apportare questa modifica a SPSS Modeler Server, specificare un valore per ODBC_DRIVER_MANAGER_PATH all'inizio dello script modelersrv.sh dove indicato dai commenti. Questa variabile di ambiente deve essere impostata sul percorso del gestore dei driver ODBC, che viene fornito con il driver ODBC Teradata (/usr/odbc/lib in un'installazione predefinita del driver ODBC Teradata). Perché la modifica abbia effetto occorre riavviare SPSS Modeler Server. Per i dettagli relativi alle piattaforme SPSS Modeler Server che offrono supporto all'accesso Teradata, e la versione di driver ODBC Teradata supportata, vedere il sito aziendale dedicato al supporto tecnico all'indirizzo <http://www.ibm.com/support>.

Utilizzare i passaggi seguenti per accedere ai dati di un database:

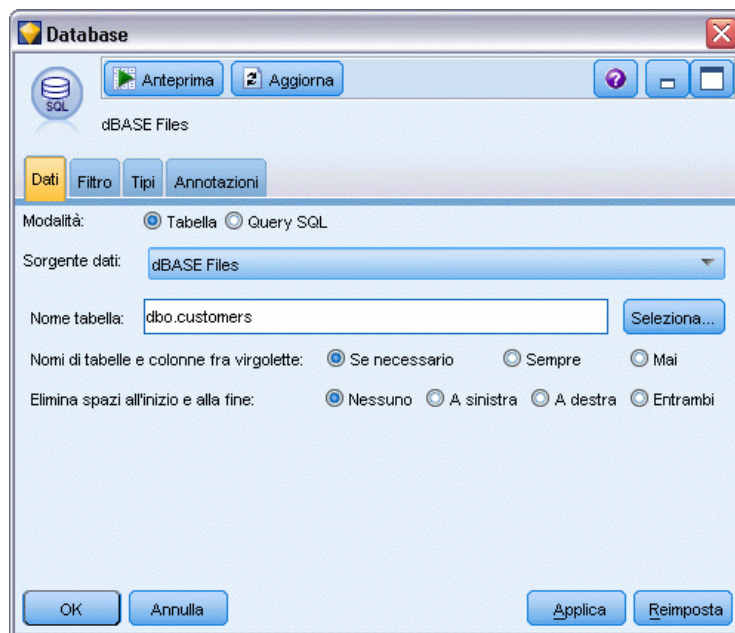
- ▶ Installare un driver ODBC e configurare una sorgente dati per il database che si desidera utilizzare.
- ▶ Nella finestra di dialogo del nodo Database eseguire la connessione a un database utilizzando la modalità Tabella o Query SQL.
- ▶ Selezionare una tabella del database.
- ▶ Tramite le schede della finestra di dialogo del nodo Database è possibile modificare i tipi di utilizzo e filtrare i campi dei dati.

Questi passaggi verranno descritti in dettaglio negli argomenti seguenti.

Impostazione delle opzioni relative al nodo Database

È possibile utilizzare le opzioni nella scheda Dati della finestra di dialogo del nodo di input Database per ottenere l'accesso a un database e leggere i dati dalla tabella selezionata.

Figura 2-6
Caricamento di dati tramite la selezione di una tabella



Modalità. Selezionare Tabella per eseguire la connessione a una tabella utilizzando i controlli della finestra di dialogo. Selezionare Query SQL per eseguire query nel database selezionato nel campo sottostante utilizzando SQL.

Sorgente dati. Sia per la modalità Tabella che per la modalità Query SQL è possibile immettere un nome nel campo Sorgente dati o selezionare Aggiungi nuova connessione database dall'elenco a discesa.

Le opzioni seguenti vengono utilizzate per eseguire la connessione a un database e selezionare una tabella utilizzando la finestra di dialogo:

Nome tabella. Se si conosce il nome della tabella a cui si desidera accedere, immettere tale nome nel campo Nome tabella. In alternativa, fare clic sul pulsante Seleziona per aprire una finestra di dialogo contenente l'elenco delle tabelle disponibili.

Nomi di tabelle e colonne fra virgolette. Specificare se si desidera che i nomi delle tabelle e delle colonne vengano racchiusi tra virgolette quando le query vengono inviate al database (per esempio nel caso in cui tali nomi contengano spazi o punteggiatura).

- L'opzione Se necessario consente di racchiudere nomi di tabelle e campi tra virgolette *solo* se contengono caratteri non standard. I caratteri non standard sono rappresentati da caratteri ASCII, spazi e qualsiasi carattere non alfanumerico diverso dal punto (.).
- Selezionare Mai se si desidera che i nomi di campi e tabelle non vengano *mai* racchiusi tra virgolette.
- Selezionare Sempre se si desidera che *tutti* i nomi di campi e tabelle vengano racchiusi tra virgolette.

Elimina spazi all'inizio e alla fine. Selezionare una delle opzioni per scartare gli spazi iniziali e finali nelle stringhe.

Nota: i confronti tra stringhe che utilizzano e non utilizzano il push back SQL potrebbero generare risultati diversi in presenza di spazi finali.

Letture di stringhe vuote da Oracle. Durante la lettura o la scrittura in un database Oracle, tenere presente che, a differenza di IBM® SPSS® Modeler e della maggioranza degli altri database, Oracle gestisce e archivia i valori di stringhe vuote come valori nulli. Questo significa che gli stessi dati estratti da un database Oracle potrebbero comportarsi in modo diverso rispetto a quando vengono estratti da un file o da un altro database e i dati potrebbero restituire risultati diversi.

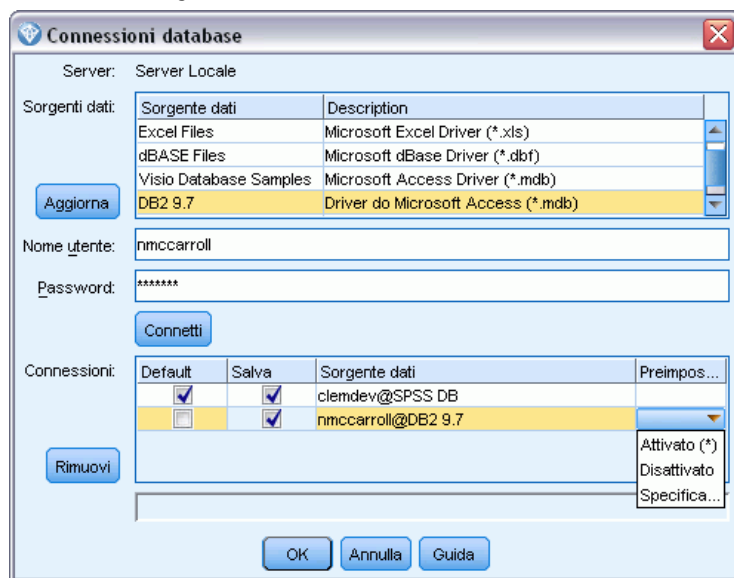
Aggiunta di una connessione a un database

Per aprire un database è innanzitutto necessario selezionare la sorgente dati a cui si desidera eseguire la connessione. Nella scheda Dati, selezionare Aggiungi nuova connessione database dall'elenco a discesa Sorgente dati.

In questo modo viene aperta la finestra di dialogo Connessioni database.

Nota: In alternativa, per aprire la finestra di dialogo, dal menu principale scegliere: Strumenti > Database...

Figura 2-7
Finestra di dialogo Connessioni database



Sorgenti dati. Elenca le sorgenti dati disponibili. Scorrere completamente l'elenco se non viene visualizzato il database desiderato. Dopo aver selezionato una sorgente dati e immesso eventuali password, fare clic su Connetti. Fare clic su Aggiorna per aggiornare l'elenco.

Nome utente. Se la sorgente dati è protetta da password, immettere il nome utente.

Password. Se la sorgente dati è protetta da password, immettere la password.

Connessioni. Visualizza i database attualmente connessi.

- **Default.** È possibile scegliere una connessione come valore di default. In questo caso, i nodi Database di input o di esportazione avranno come sorgente dati questa connessione predefinita, anche se l'impostazione si può modificare.
- **Salva.** Se lo si desidera, selezionare una o più connessioni da visualizzare nuovamente nelle sessioni successive.
- **Sorgente dati.** Le stringhe di connessione per i database connessi in quel momento.
- **Preimpostati.** Indica (con un carattere *) se sono stati specificati dei valori preimpostati per la connessione al database. Per specificare valori preimpostati, fare clic su questa colonna nella riga corrispondente alla connessione al database e scegliere Specifica dall'elenco. [Per ulteriori informazioni, vedere l'argomento Specifica di valori preimpostati per una connessione a un database a pag. 17.](#)

Per rimuovere una connessione esistente, selezionarla dall'elenco e fare clic su Rimuovi.

Una volta effettuate tutte le selezioni, fare clic su OK.

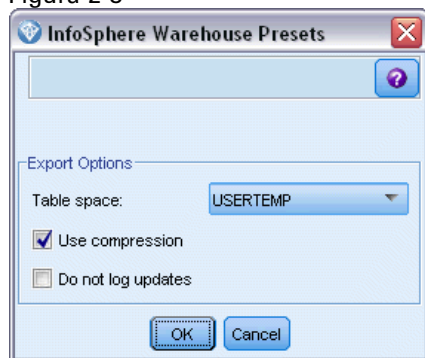
Specifica di valori preimpostati per una connessione a un database

È possibile specificare numerose impostazioni di default per una connessione a un database.

Impostazioni IBM InfoSphere Warehouse

Tutte le opzioni preimpostate di InfoSphere Warehouse sono valide per l'esportazione in un database.

Figura 2-8



Spazio di tabella. Lo spazio di tabella da utilizzare per l'esportazione. Gli amministratori di database possono creare o configurare spazi di tabella partizionati. Si consiglia di selezionare uno di questi spazi di tabella (anziché quello di default) da utilizzare per le esportazioni in database.

Utilizza compressione. Se è selezionata, questa opzione utilizza la compressione nella creazione delle tabelle per l'esportazione (per esempio, l'equivalente di CREATE TABLE MYTABLE(...) COMPRESS YES in SQL).

Non registrare gli aggiornamenti. Se è selezionata, questa opzione evita la registrazione delle operazioni quando si creano tabelle e si inseriscono i dati (l'equivalente di CREATE TABLE MYTABLE(...) NOT LOGGED INITIALLY in SQL).

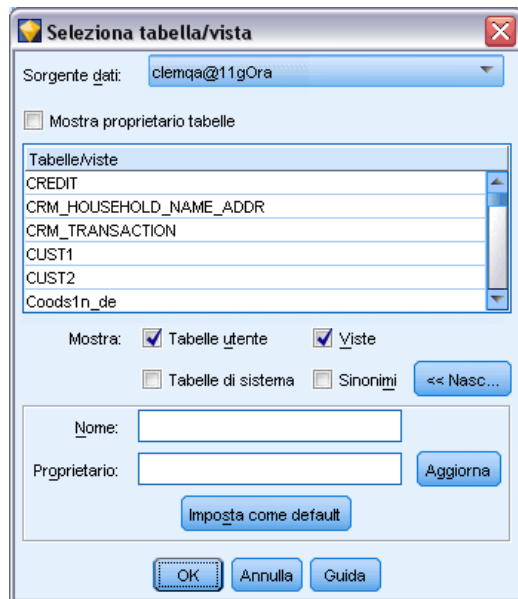
Impostazioni per altri database

Per gli altri database non è attualmente possibile definire valori preimpostati.

Selezione di una tabella da un database

Dopo aver eseguito la connessione a una sorgente dati, è possibile scegliere di importare campi da una tabella o una vista specifica. Nella scheda Dati della finestra di dialogo del nodo Database è possibile immettere il nome di una tabella nel campo Nome tabella oppure fare clic su **Seleziona** per aprire una finestra di dialogo contenente l'elenco delle tabelle e delle viste disponibili.

Figura 2-9
Selezione di una tabella dal database connesso



Mostra proprietario tabelle. Selezionare questa opzione se una sorgente dati richiede che venga specificato il nome del proprietario di una tabella prima di consentire l'accesso alla tabella. Deselezionare questa opzione per le sorgenti dati che non includono questo requisito.

Nota: in genere per i database SAS e Oracle è necessario specificare il proprietario della tabella.

Tabelle/viste. Selezionare la tabella o la vista da importare.

Mostra. Elenca le colonne della sorgente dati attualmente connesse. Fare clic su una delle opzioni seguenti per personalizzare la visualizzazione delle tabelle disponibili:

- Selezionare **Tabelle utente** per visualizzare le normali tabelle di database create dagli utenti del database.

- Selezionare Tabelle di sistema per visualizzare le tabelle del database il cui proprietario è il sistema, per esempio le tabelle che forniscono le informazioni relative al database, come i dettagli sugli indici. Questa opzione può essere usata per visualizzare le schede utilizzate in database di Excel (si noti che è disponibile anche un nodo di input Excel separato. [Per ulteriori informazioni, vedere l'argomento Nodo di input Excel a pag. 44.](#))
- Selezionare Viste per visualizzare tabelle virtuali risultanti da una query su una o più tabelle ordinarie.
- Selezionare Sinonimi per visualizzare i sinonimi creati nel database per tutte le tabelle esistenti.

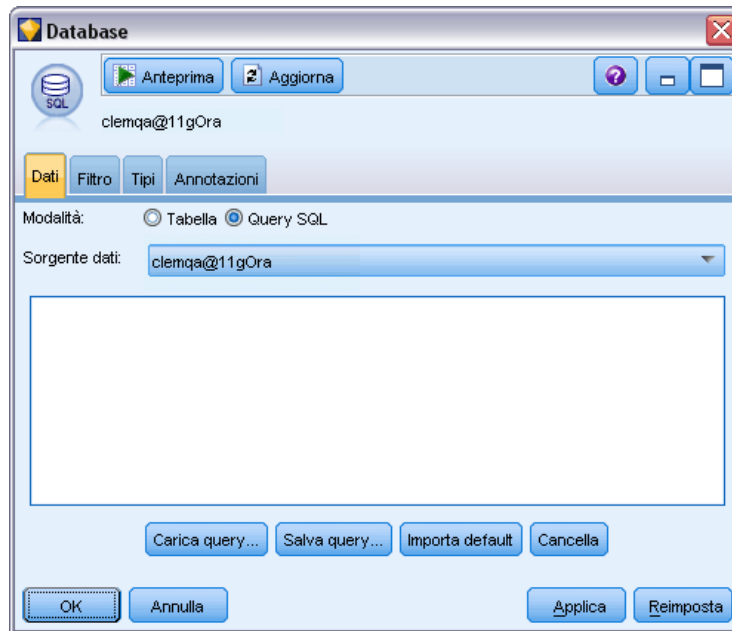
Filtri Nome/Proprietario. Questi campi consentono di filtrare l'elenco delle tabelle visualizzate in base al nome o al proprietario. Per esempio, digitare SYS per visualizzare l'elenco delle tabelle con il proprietario specificato. Per le ricerche tramite i caratteri jolly, è possibile utilizzare il carattere di sottolineatura (_) per rappresentare qualsiasi carattere singolo e il segno di percentuale (%) per rappresentare una sequenza di zero o più caratteri.

Imposta come default. Salva le impostazioni correnti come di default per l'utente corrente. Le impostazioni verranno ripristinate successivamente quando un utente apre una nuova finestra di selezione delle tabelle *per lo stesso nome di sorgente dati e login utente*.

Invio di query al database

Dopo aver eseguito una connessione a una sorgente dati è possibile scegliere di importare campi utilizzando una query SQL. Nella finestra di dialogo principale selezionare Query SQL come modalità di connessione. In questo modo verrà aggiunta una finestra dell'editor di query nella finestra di dialogo. Grazie all'editor di query è possibile creare o caricare una query SQL il cui insieme di risultati verrà letto nello stream di dati. Per chiudere la finestra dell'editor di query, selezionare Tabella come modalità di connessione.

Figura 2-10
Caricamento di dati tramite query SQL



Carica query. Selezionare questa opzione per aprire il browser dei file, che può essere utilizzato per caricare una query salvata in precedenza.

Salva query. Selezionare questa opzione per aprire la finestra di dialogo Salva query, che può essere utilizzata per salvare la query corrente.

Importa default. Selezionare questa opzione per importare un'istruzione SQL SELECT di esempio creata automaticamente in base alla tabella e alle colonne selezionate nella finestra di dialogo.

Cancella. Cancella il contenuto dell'area di lavoro. Utilizzare questa opzione se si desidera cominciare da capo.

Nodo Testo variabile

È possibile utilizzare i nodi Testo variabile per leggere dati da file di testo a campi liberi, ovvero file i cui record contengono un numero costante di campi e un numero variabile di caratteri, noti anche come file di testo delimitati. Questo tipo di nodi può essere utilizzato per file con testo di intestazione a lunghezza fissa e alcuni tipi di annotazioni. I record vengono letti uno alla volta e passati nello stream finché non viene letto l'intero file.

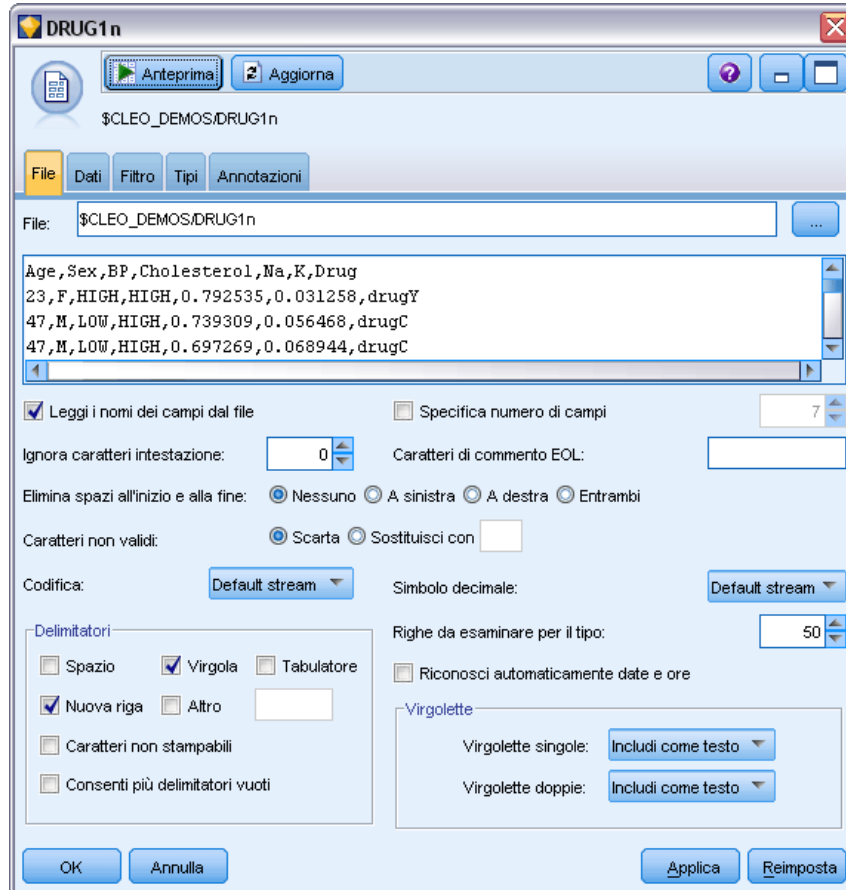
Note per la lettura dei dati di testo delimitati

- Alla fine di ogni riga, i record devono essere delimitati da un carattere di nuova riga. Il carattere di nuova riga non deve essere utilizzato per altri scopi (per esempio, all'interno di un nome di campo o di un valore). Gli spazi iniziali e finali andrebbero eliminati per risparmiare spazio, anche se questo non è fondamentale. Se necessario, tali spazi possono essere eliminati dal nodo.

- I campi devono essere delimitati da una virgola o da un altro carattere che andrebbe utilizzato solo come delimitatore, cioè non deve comparire all'interno di nomi di campo o di valori. Se ciò non è possibile, tutti i campi di testo possono essere racchiusi tra virgolette doppie, purché nessuno dei nomi dei campi o dei valori di testo contenga virgolette doppie. Se i nomi dei campi o i valori contengono virgolette doppie, in alternativa i campi di testo possono essere racchiusi tra virgolette singole, sempre a condizione che le virgolette singole non siano utilizzate in altri punti all'interno di valori. Se non è possibile utilizzare né virgolette singole, né virgolette doppie, i valori di testo devono essere modificati in modo da rimuovere o sostituire il carattere di delimitazione oppure le virgolette singole o doppie.
- Ogni riga, intestazione compresa, deve contenere lo stesso numero di campi.
- La prima riga deve contenere i nomi dei campi. In caso contrario, deselezionare l'opzione Leggi i nomi dei campi dal file per assegnare a ogni campo un nome generico come *Campo1*, *Campo2* e così via.
- La seconda riga deve contenere il primo record di dati. Non devono essere presenti righe vuote o commenti.
- I valori numerici non devono includere il separatore delle migliaia o simboli di raggruppamento, per esempio il valore 3.000,00 non deve contenere il punto. L'indicatore dei decimali (virgola o punti negli Stati Uniti e in Gran Bretagna) deve essere usato solo ove appropriato.
- I valori di data e ora devono essere in uno dei formati riconosciuti nella finestra di dialogo Opzioni stream, per esempio DD/MM/YYYY o HH:MM:SS. Tutti i campi data e ora nel file dovrebbero seguire lo stesso formato e i campi contenenti una data devono utilizzare lo stesso formato per tutti i valori all'interno di quel campo. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Impostazione delle opzioni relative al nodo Testo variabile

Figura 2-11
Finestra di dialogo del nodo Testo variabile



File. Specificare il nome del file. È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un file. Il percorso del file viene mostrato quando il file viene selezionato e i relativi contenuti separati da delimitatori vengono visualizzati nel riquadro sottostante.

Il testo di esempio visualizzato dalla sorgente dati può essere copiato e incollato nei controlli seguenti: caratteri di commento EOL e delimitatori specificati dall'utente. Utilizzare Ctrl-C e Ctrl-V per copiare e incollare.

Leggi i nomi dei campi dal file. Questa opzione è selezionata per default e consente di utilizzare i nomi dei campi inclusi nella prima riga nel file di dati come etichette per le colonne. Se la prima riga non è una riga di intestazione, deselezionare questa opzione per assegnare automaticamente a ogni campo un nome generico, per esempio *Campo1*, *Campo2*, in base al numero di campi nell'insieme di dati.

Specifica numero di campi. Specificare il numero di campi in ogni record. Il numero di campi può essere rilevato automaticamente a condizione che i record siano delimitati da un carattere di nuova riga. È anche possibile impostare manualmente un numero.

Ignora caratteri intestazione. Specificare il numero di caratteri che si desidera ignorare all'inizio del primo record.

Caratteri di commento EOL. Specificare caratteri quali # o ! per indicare annotazioni nei dati. Ogni volta che nel file di dati verrà rilevato uno di questi caratteri, tutti gli elementi precedenti il carattere di nuova riga (escluso tale carattere) verranno ignorati.

Elimina spazi all'inizio e alla fine. Selezionare una delle opzioni per scartare gli spazi iniziali e finali nelle stringhe durante l'importazione.

Nota: i confronti tra stringhe che utilizzano e non utilizzano il push back SQL potrebbero generare risultati diversi in presenza di spazi finali.

Caratteri non validi. Selezionare Scarta per rimuovere i caratteri non validi dall'origine dei dati. Selezionare Sostituisci con per sostituire i caratteri non validi con il simbolo specificato (un solo carattere). I caratteri non validi sono caratteri nulli o caratteri non esistenti nel metodo di codifica specificato.

Codifica. Specifica il metodo di codifica del testo utilizzato. È possibile scegliere tra l'impostazione di default del sistema, l'impostazione di default dello stream o UTF-8.

- L'impostazione di default del sistema è specificata nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. [Per ulteriori informazioni, vedere l'argomento Supporto Unicode in IBM SPSS Modeler in l'appendice B in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- L'impostazione di default dello stream è specificata nella finestra di dialogo Proprietà stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Simbolo decimale. Selezionare il tipo di separatore di decimali utilizzato nella sorgente dati. Default stream rappresenta il carattere selezionato nella scheda Opzioni della finestra di dialogo delle proprietà dello stream. In alternativa, selezionare Punto (.) o Virgola (,) per utilizzare il relativo simbolo come separatore di decimali per tutti i dati inclusi nella finestra di dialogo.

Delimitatori. Tramite le caselle di controllo incluse in questo controllo è possibile specificare i caratteri da utilizzare come delimitatori per i campi del file, per esempio la virgola (,). È possibile specificare più delimitatori, per esempio “, |”, per i record che utilizzano più delimitatori. Il delimitatore di default è la virgola.

Nota: se la virgola è definita anche come separatore di decimali, in questo caso le impostazioni di default non funzioneranno. Nei casi in cui la virgola è sia delimitatore di campo che separatore decimale, selezionare Altro nell'elenco Delimitatori, quindi specificare manualmente una virgola nel campo di immissione.

Selezionare Consenti più delimitatori vuoti per trattare più caratteri delimitatori vuoti adiacenti come un unico delimitatore. Per esempio, se un valore nei dati è seguito da quattro spazi e quindi da un altro valore, verranno considerati due campi anziché cinque.

Righe da esaminare per il tipo. Indicare il numero di righe da esaminare per i tipi di dati specificati.

Riconosci automaticamente date e ore. Per fare in modo che IBM® SPSS® Modeler cerchi automaticamente di riconoscere le voci di dati come date e ore, selezionare questa casella di controllo. Per esempio, questo significa che una voce come 07-11-1965 verrà identificata come

una data e 02:35:58 come ora. Tuttavia, le voci ambigue, come 07111965 o 023558, verranno identificate come numeri interi in quanto non ci sono delimitatori fra i numeri.

Nota: per evitare potenziali problemi di dati quando si utilizzano file di dati delle versioni precedenti di SPSS Modeler, questa casella, per default, è disattivata per le informazioni salvate nelle versioni precedenti alla versione 13.

Virgolette. Tramite gli elenchi a discesa è possibile specificare in che modo si desidera vengano utilizzate le virgolette singole e doppie durante l'importazione. È possibile selezionare Scarta per scartare tutti i tipi di virgolette, Includi come testo per includerle nei valori di campo o Accoppia e scarta per individuare coppie di virgolette corrispondenti e quindi scartarle. Se per una virgoletta non viene trovato il corrispondente, verrà visualizzato un messaggio di errore. Le opzioni Scarta e Accoppia e scarta archiviano il valore di campo (senza virgolette) come stringa.

Mentre si utilizza questa finestra di dialogo è sempre possibile fare clic su Aggiorna per ricaricare i campi dalla sorgente dati. Questa opzione risulta utile quando si modificano le connessioni dei dati al nodo di input o quando si passa da una scheda all'altra nella finestra di dialogo.

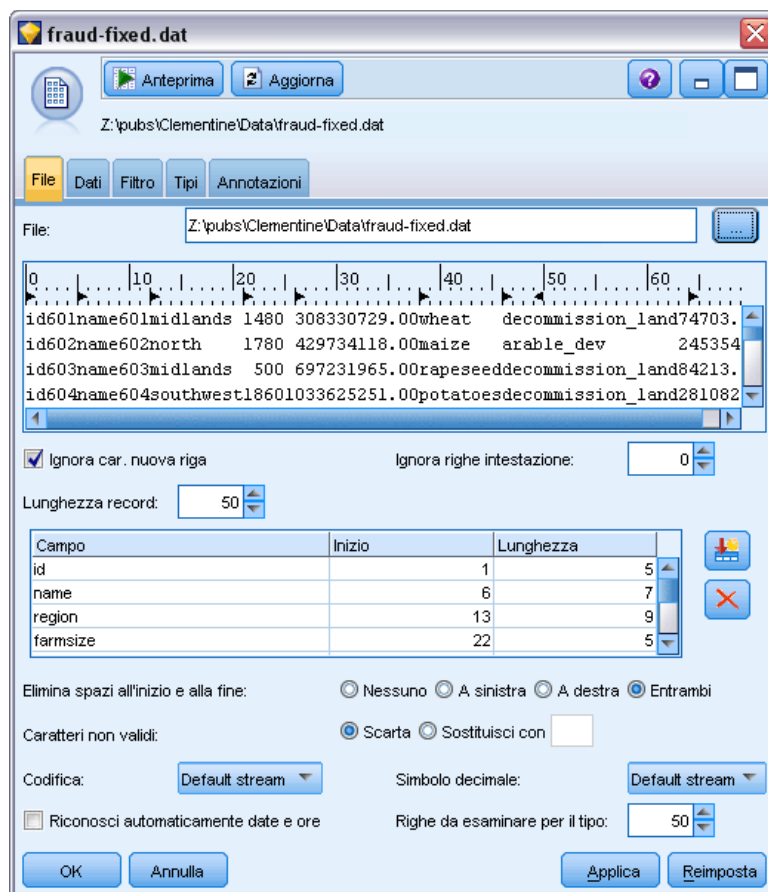
Nodo Testo fisso

È possibile utilizzare i nodi Testo fisso per importare dati da file di testo a campi fissi (ovvero file i cui campi non sono delimitati, iniziano nella stessa posizione e sono di lunghezza fissa). Nel formato a campi fissi vengono in genere archiviati dati di versioni precedenti o generati dalla macchina. Tramite la scheda File del nodo Testo fisso è possibile specificare la posizione e la lunghezza delle colonne nei dati.

Impostazione delle opzioni relative al nodo Testo fisso

La scheda File del nodo Testo fisso consente di importare dati in IBM® SPSS® Modeler e di specificare la posizione delle colonne e la lunghezza dei record. È possibile fare clic all'interno del riquadro di anteprima dei dati al centro della finestra di dialogo per aggiungere frecce indicanti i punti di separazione tra i campi.

Figura 2-12
Definizione di colonne in dati a campi fissi File.



File. Specificare il nome del file. È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un file. Dopo avere selezionato un file, viene visualizzato il relativo percorso e il contenuto con i delimitatori viene visualizzato nel riquadro sottostante.

Il riquadro di anteprima dei dati può essere utilizzato per specificare la lunghezza e la posizione delle colonne. Il righello nella parte superiore del riquadro di anteprima consente di misurare la lunghezza delle variabili e di specificare il punto di separazione tra le variabili. È possibile specificare le linee relative ai punti di separazione facendo clic nell'area del righello al di sopra dei campi. I punti di separazione possono essere spostati tramite trascinamento e scartati trascinandoli all'esterno dell'area di anteprima dei dati.

- Per ogni punto di separazione viene aggiunto automaticamente un nuovo campo ai campi della tabella sottostante.
- Le posizioni iniziali indicate dalle frecce vengono aggiunte automaticamente alla colonna Inizio nella tabella sottostante.

Ignora car. nuova riga. Selezionare questa opzione se si desidera ignorare il carattere di nuova riga alla fine di ogni record.

Ignora righe intestazione. Specificare il numero di righe che si desidera ignorare all'inizio del primo record. Questa opzione può essere utilizzata per ignorare le intestazioni delle colonne.

Lunghezza record. Specificare il numero di caratteri in ogni record.

Campo. Di seguito sono elencati tutti i campi definiti per questo file di dati. Esistono due modi per definire i campi:

- Definizione interattiva dei campi mediante il riquadro di anteprima dei dati.
- Definizione manuale dei campi mediante l'aggiunta di righe di campo vuote nella tabella. Fare clic sul pulsante a destra del riquadro dei campi per aggiungere nuovi campi. Quindi immettere nel campo vuoto un nome di campo, una posizione iniziale e una lunghezza. L'impostazione di queste opzioni determina l'aggiunta automatica delle frecce al riquadro di anteprima, con la possibilità di modifiche successive.

Per rimuovere un campo definito in precedenza, selezionarlo nell'elenco e fare clic sul pulsante con la X rossa.

Inizio. Specificare la posizione del primo carattere nel campo. Per esempio, se il secondo campo di un record inizia in corrispondenza del sedicesimo carattere, come posizione iniziale si specificherà 16.

Lunghezza. Specificare il numero di caratteri presenti nel valore più lungo per ogni campo. In questo modo viene definito il punto di interruzione per il campo successivo.

Elimina spazi all'inizio e alla fine. Selezionare una delle opzioni per scartare gli spazi iniziali e finali nelle stringhe durante l'importazione.

Nota: i confronti tra stringhe che utilizzano e non utilizzano il push back SQL potrebbero generare risultati diversi in presenza di spazi finali.

Caratteri non validi. Selezionare Scarta per rimuovere i caratteri non validi dall'input dei dati. Selezionare Sostituisci con per sostituire i caratteri non validi con il simbolo specificato (un solo carattere). I caratteri non validi sono caratteri nulli (0) o caratteri non esistenti nella codifica corrente.

Codifica. Specifica il metodo di codifica del testo utilizzato. È possibile scegliere tra l'impostazione di default del sistema, l'impostazione di default dello stream o UTF-8.

- L'impostazione di default del sistema è specificata nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. [Per ulteriori informazioni, vedere l'argomento Supporto Unicode in IBM SPSS Modeler in l'appendice B in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- L'impostazione di default dello stream è specificata nella finestra di dialogo Proprietà stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Simbolo decimale. Selezionare il tipo di separatore di decimali utilizzato nella sorgente dati. Default stream rappresenta il carattere selezionato nella scheda Opzioni della finestra di dialogo delle proprietà dello stream. In alternativa, selezionare Punto (.) o Virgola (,) per utilizzare il relativo simbolo come separatore di decimali per tutti i dati inclusi nella finestra di dialogo.

Riconosci automaticamente date e ore. Per fare in modo che SPSS Modeler cerchi automaticamente di riconoscere le voci di dati come date e ore, selezionare questa casella di controllo. Per esempio, questo significa che una voce come 07-11-1965 verrà identificata come una data e 02:35:58 come ora. Tuttavia, le voci ambigue, come 07111965 o 023558, verranno identificate come numeri interi in quanto non ci sono delimitatori fra i numeri.

Nota: per evitare potenziali problemi di dati quando si utilizzano file di dati delle versioni precedenti di SPSS Modeler, questa casella, per default, è disattivata per le informazioni salvate nelle versioni precedenti alla versione 13.

Righe da esaminare per il tipo. Indicare il numero di righe da esaminare per i tipi di dati specificati.

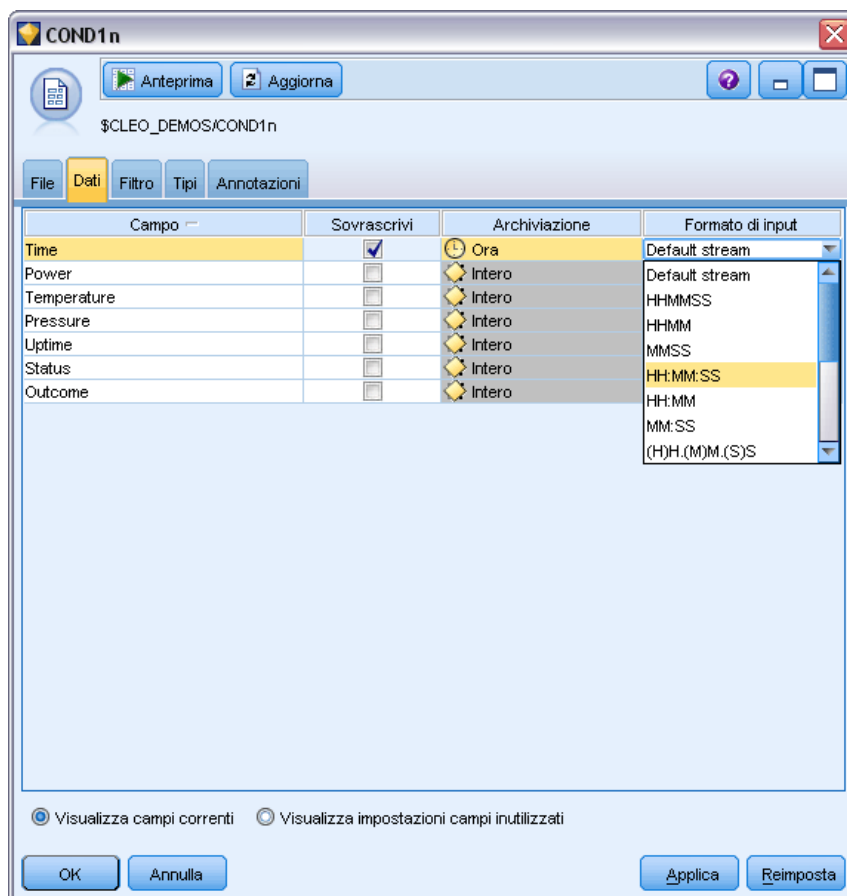
Mentre si utilizza questa finestra di dialogo è sempre possibile fare clic su **Aggiorna** per ricaricare i campi dalla sorgente dati. Questa opzione risulta utile quando si modificano le connessioni dei dati al nodo di input o quando si passa da una scheda all'altra nella finestra di dialogo.

Impostazione dell'archiviazione e della formattazione dei campi

Le opzioni nella scheda Dati, comuni ai nodi Testo fisso, Testo variabile, Input XML e Input utente, consentono di specificare il tipo di archiviazione dei campi importati o creati in IBM® SPSS® Modeler. Per i nodi Testo fisso, Testo variabile e Input utente è possibile specificare anche la formattazione dei campi e altri metadati.

Per i dati letti da altre sorgenti, l'archiviazione viene determinata automaticamente ma può essere modificata utilizzando una funzione di conversione, quale `to_integer`, in un nodo Riempimento o Nuovo campo.

Figura 2-13
Sovrascrittura del tipo di classe di archiviazione e del formato di campo durante l'importazione



Campo. La colonna *Campo* consente di visualizzare e selezionare i campi nell'insieme di dati corrente.

Sovrascrivi. Selezionare la casella di controllo nella colonna *Sovrascrivi* per attivare le opzioni delle colonne *Classe archiviazione* e *Formato di input*.

Archiviazione dei dati

L'archiviazione definisce le modalità di archiviazione dei dati in un campo. Per esempio, un campo con valori 1 e 0 è un campo in cui sono archiviati dati di tipo intero. Si distingue dal livello di misurazione, che descrive l'utilizzo dei dati e non influisce sull'archiviazione. Per esempio, è possibile impostare il livello di misurazione per un campo intero con valori 1 e 0 su *Flag*. In genere, questo indica che 1=*Vero* e 0=*Falso*. Mentre l'archiviazione deve essere stabilita all'origine, il livello di misurazione può essere modificato tramite un nodo *Tipo* in qualsiasi punto dello stream. [Per ulteriori informazioni, vedere l'argomento Livelli di misurazione in il capitolo 4 a pag. 127.](#)

I tipi di archiviazione disponibili sono:

- **Stringa.** Utilizzato per i campi che contengono dati non numerici, detti anche dati alfanumerici. Una stringa può includere qualsiasi sequenza di caratteri, quale *mario*, *Classe 2* o *1234*. Si noti che i numeri presenti nelle stringhe non possono essere utilizzati nei calcoli.
- **Intero.** Un campo i cui valori sono numeri interi.
- **Reale.** I valori sono numeri che possono includere decimali (non limitati ai numeri interi). Il formato di visualizzazione viene specificato nella finestra di dialogo Proprietà stream e può essere ignorato per singoli campi in un nodo Tipo (scheda Formato). [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Data.** Valori di data specificati in un formato standard quale anno, mese e giorno (per esempio, 26.09.07). Il formato specifico è indicato nella finestra di dialogo Proprietà stream.
- **Tempo.** Tempo misurato come durata. Per esempio, se una chiamata di assistenza è durata 1 ora, 26 minuti e 38 secondi, potrà essere rappresentata come 01:26:38, in base al formato dell'ora corrente specificato nella finestra di dialogo Proprietà stream.
- **Timestamp.** Valori che includono sia i componenti data e ora, per esempio 2007–09–26 09:04:00, anche questi in base al formato dell'ora corrente specificato nella finestra di dialogo Proprietà stream. Si noti che può essere necessario racchiudere i valori di timestamp tra virgolette per garantire che vengano interpretati come un unico valore anziché come valori di data e ora separati (per esempio quando si immettono valori in un nodo Input utente).

Conversioni della classe di archiviazione. È possibile convertire la classe di archiviazione di un campo utilizzando una vasta gamma di funzioni di conversione, per esempio `to_string` e `to_integer`, in un nodo Riempimento. [Per ulteriori informazioni, vedere l'argomento Conversione della classe di archiviazione mediante il nodo Riempimento in il capitolo 4 a pag. 168.](#) Si noti che le funzioni di conversione, e qualsiasi altra funzione che richieda un tipo di input specifico quale un valore di data o ora, dipendono dai formati correnti specificati nella finestra di dialogo Proprietà stream. Per esempio, se si desidera convertire un campo stringa con valori *Gen 2003*, *Feb 2003* e così via in campo di archiviazione della data, selezionare MES AAAA come formato di data di default per lo stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#) Le funzioni di conversione sono disponibili anche dal nodo Nuovo campo per la conversione temporanea durante un calcolo di derivazione. È inoltre possibile utilizzare il nodo Nuovo campo per eseguire altre manipolazioni, quali ricodifiche di campi stringa con valori categoriali. [Per ulteriori informazioni, vedere l'argomento Ricodifica di valori con il nodo Nuovo campo in il capitolo 4 a pag. 165.](#)

Lettura di dati misti. Quando si leggono campi con archiviazione numerica (numeri interi, reali, ora, timestamp o data), gli eventuali valori non numerici vengono impostati come valori nulli o valori mancanti di sistema poiché, a differenza di altre applicazioni, SPSS Modeler non consente di utilizzare tipi di archiviazione misti all'interno dello stesso campo. Per evitare questo inconveniente, è necessario modificare opportunamente il tipo di archiviazione nel nodo di input o nell'applicazione esterna per fare in modo che tutti i campi contenenti dati misti siano letti come stringhe.

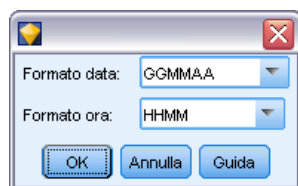
Formato di input del campo (solo nodi Testo fisso, Testo variabile e Input utente)

Per tutti i tipi di classe di archiviazione eccetto Stringa e Intero, è possibile specificare le opzioni di formattazione per il campo selezionato utilizzando l'elenco a discesa. Per esempio, quando si uniscono dati tra diverse opzioni internazionali, potrebbe essere necessario specificare il punto (.) come separatore decimale per un campo e la virgola per un altro campo.

Le opzioni di input definite nel nodo di input sovrascrivono le opzioni di formattazione specificate nella finestra di dialogo delle proprietà dello stream, ma non vengono mantenute successivamente nello stream. Queste opzioni sono state progettate per analizzare correttamente l'input in base alle informazioni note sui dati. I formati specificati risultano utili per analizzare i dati mentre vengono letti in SPSS Modeler e non per stabilire in quale modo devono essere formattati dopo essere stati letti in SPSS Modeler. Per definire il formato di ognuno dei campi in un'altra posizione dello stream, utilizzare la scheda Formato di un nodo Tipo. [Per ulteriori informazioni, vedere l'argomento Scheda per l'impostazione dei formati dei campi in il capitolo 4 a pag. 141.](#)

Figura 2-14

Definizione dei formati di data e ora per i campi Timestamp



Le opzioni variano in base al tipo di archiviazione. Per esempio, per il tipo Reale, è possibile selezionare Punto (.) o Virgola (,) come separatore decimale. Per i campi Timestamp, quando si seleziona Specifica dall'elenco a discesa viene aperta un'altra finestra di dialogo. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni relative al formato dei campi in il capitolo 4 a pag. 142.](#)

Per tutti i tipi di archiviazione, è inoltre possibile selezionare Default stream per utilizzare le impostazioni predefinite dello stream per l'importazione. Le impostazioni dello stream sono definite nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Opzioni aggiuntive

Nella scheda Dati è possibile definire molte altre opzioni:

- Per visualizzare le impostazioni di archiviazione per dati non più connessi al nodo corrente (per esempio dati di addestramento), selezionare Visualizza impostazioni campi inutilizzati. È possibile eliminare i campi precedenti facendo clic su Cancella.
- Mentre si utilizza questa finestra di dialogo è sempre possibile fare clic su Aggiorna per ricaricare i campi dalla sorgente dati. Questa opzione risulta utile quando si modificano le connessioni dei dati al nodo di input o quando si passa da una scheda all'altra nella finestra di dialogo.

Nodo Data Collection

I nodi di input Data Collection importano i dati di sondaggi in base al IBM® SPSS® Data Collection Survey Reporter Developer Kit utilizzato dal software di ricerca di mercato di IBM Corp.. Questo formato distingue i **dati del caso**— cioè le risposte effettive alle domande raccolte durante un sondaggio— dai **metadati**, che descrivono le modalità di raccolta e organizzazione dei dati del caso. I metadati sono costituiti da informazioni quali testi di domande, nomi di variabili e descrizioni, definizioni di variabili a risposta multipla, traduzioni delle stringhe di testo, nonché dalla definizione della struttura dei dati del caso.

Nota: Questo nodo richiede il Survey Reporter Developer Kit di Data Collection, distribuito da IBM Corp. insieme ai prodotti software di Data Collection. Per ulteriori informazioni, consultare la pagina Web di Data Collection all'indirizzo <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>. Fatta eccezione per l'installazione del Developer Kit, non è necessario eseguire altre operazioni di configurazione.

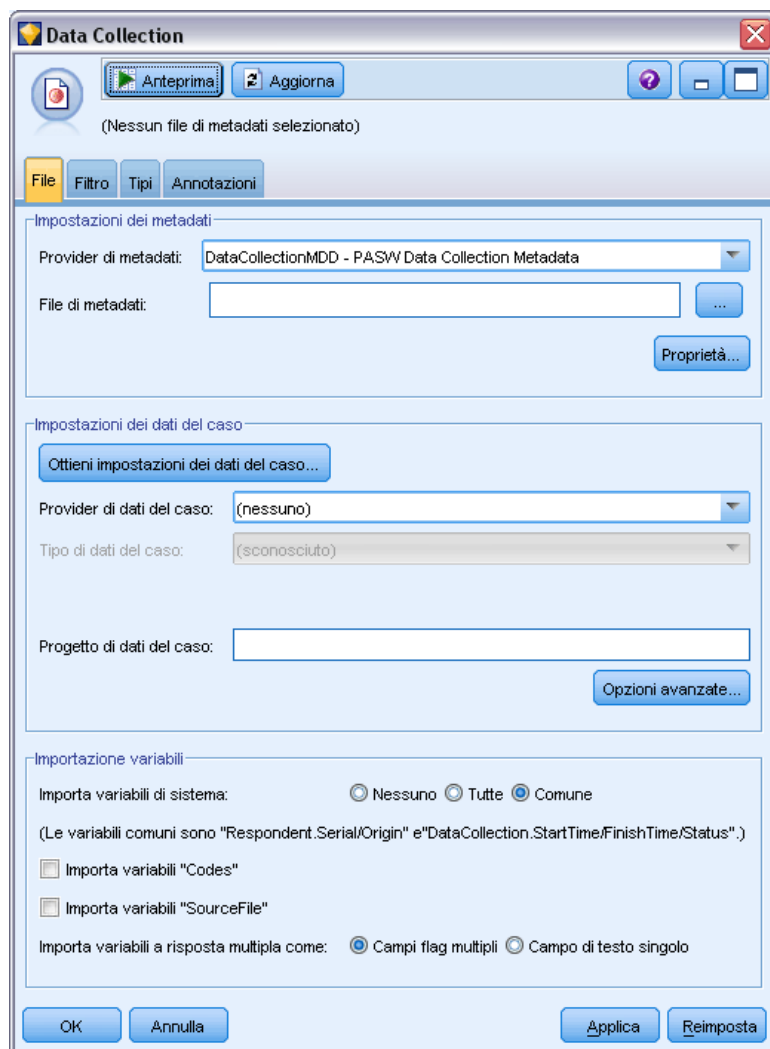
Commenti

- I dati dei sondaggi vengono letti dal formato piatto tabulare VDATA o da sorgenti nel formato gerarchico HDATA se includono una sorgente di metadati (è necessario Data Collection versione 4.5 o superiore).
- I tipi vengono istanziati automaticamente utilizzando le informazioni fornite dai metadati.
- Quando si importano in IBM® SPSS® Modeler dati di sondaggi, le domande vengono restituite come campi, con un record per ogni rispondente.

Opzioni scheda File del nodo Importazione di Data Collection

La scheda File del nodo Data Collection consente di specificare le opzioni relative ai metadati e ai dati del caso che si desidera importare.

Figura 2-15
Opzioni scheda File del nodo di input di Data Collection



Impostazioni dei metadati

Nota: per vedere l'elenco completo dei tipi di file di provider disponibili, è necessario installare IBM® SPSS® Data Collection Survey Reporter Developer Kit, disponibile con il software Data Collection. Per ulteriori informazioni, vedere la pagina Web di Data Collection all'indirizzo <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>

Provider di metadati. I dati di sondaggi possono essere importati da numerosi formati supportati dal software Survey Reporter Developer Kit di Data Collection. I tipi di provider disponibili includono:

- DataCollectionMDD. Legge i metadati da un file di definizione del questionario (.mdd). Questo è il formato standard del Modello dati di Data Collection.

- Database ADO. Legge i dati e i metadati del caso dai file ADO. Specificare il nome e la posizione del file *.adoinfo* che contiene i metadati. Il nome interno di questo DSC è *mrADODsc*.
- Database In2data. Legge i dati e i metadati del caso In2data. Il nome interno di questo DSC è *mrI2dDsc*.
- File di registro Data Collection. Legge i metadati da un file di registro Data Collection standard. In genere, i file di registro presentano un'estensione *.tmp*. Tuttavia, alcuni file di registro possono avere un'estensione diversa. Se necessario, è possibile rinominare il file in modo che presenti un'estensione *.tmp*. Il nome interno di questo DSC è *mrLogDsc*.
- File di definizione Quancept. Converte i metadati in uno script Quancept. Specificare il nome del file *.qdi* Quancept. Il nome interno di questo DSC è *mrQdiDrsDsc*.
- Database Quanvert. Legge i dati e i metadati del caso Quanvert. Specificare il nome e la posizione del file *.qvinfo* o *.pkd*. Il nome interno di questo DSC è *mrQvDsc*.
- Data Collection Participation Database. Legge le tabelle Campione e Tabella cronologia di un progetto e crea le variabili categoriali derivate corrispondenti alle colonne di queste tabelle. Il nome interno di questo DSC è *mrSampleReportingMDSC*.
- File Statistics. Legge i dati e i metadati del caso da un file *.sav* IBM® SPSS® Statistics. Scrive i dati del caso in un file SPSS Statistics *.sav* per l'analisi in SPSS Statistics. Scrive i metadati da un file *.sav* SPSS Statistics in un file *.mdd*. Il nome interno di questo DSC è *mrSavDsc*.
- File Surveycraft. Legge i dati e i metadati del caso SurveyCraft. Specificare il nome del file *.vq* SurveyCraft. Il nome interno di questo DSC è *mrSCDsc*.
- File di script Data Collection. Legge dai metadati in un file *mrScriptMetadata*. In genere, questi file presentano un'estensione *.mdd* o *.dms*. Il nome interno di questo DSC è *mrScriptMDSC*.
- File XML Triple-S. Legge i metadati da un file Triple-S in formato XML. Il nome interno di questo DSC è *mrTripleSDsc*.

Proprietà dei metadati. Se lo si desidera, selezionare Proprietà per specificare la versione del sondaggio da importare e la lingua, il contesto e il tipo di etichetta da utilizzare. [Per ulteriori informazioni, vedere l'argomento Proprietà dei metadati dell'importazione di IBM SPSS Data Collection a pag. 35.](#)

Impostazione dei dati del caso

Nota: per vedere l'elenco completo dei tipi di file di provider disponibili, è necessario installare Data Collection Survey Reporter Developer Kit, disponibile con il software Data Collection. Per ulteriori informazioni, vedere la pagina Web di Data Collection all'indirizzo <http://www.ibm.com/software/analytics/spss/products/data-collection/survey-reporter-dev-kit/>

Ottieni impostazione dei dati del caso. Solo in caso di lettura di metadati da file *.mdd*, fare clic su Ottieni impostazione dei dati del caso per determinare quali sorgenti dei dati del caso sono associate ai metadati selezionati, oltre alle impostazioni specifiche necessarie per accedere a una determinata sorgente. Questa opzione è disponibile solo per i file *.mdd*.

Provider di dati del caso. Sono supportati i seguenti tipi di provider:

- Database ADO. Legge i dati del caso utilizzando l'interfaccia Microsoft ADO. Selezionare UDL OLE-DB per il tipo di dati del caso e specificare una stringa di connessione nel campo UDL di dati del caso. [Per ulteriori informazioni, vedere l'argomento Stringa di connessione al database a pag. 36.](#) Il nome interno di questo componente è *mrADODsc*.
- File di testo delimitato (Excel). Legge i dati del caso da un file delimitato da virgola (.CSV), quale un file generato da Excel. Il nome interno è *mrCsvDsc*.
- File di dati Data Collection. Legge i dati del caso da un file in formato dati di Data Collection nativo (a partire da Data Collection versione 4.5). Il nome interno è *mrDataFileDsc*.
- Database In2data. Legge i dati e i metadati del caso da un file di database (.i2d) In2data. Il nome interno è *mrI2dDsc*.
- File di registro Data Collection. Legge i dati del caso da un file di registro Data Collection standard. In genere, i file di registro presentano un'estensione *.tmp*. Tuttavia, alcuni file di registro possono avere un'estensione diversa. Se necessario, è possibile rinominare il file in modo che presenti un'estensione *.tmp*. Il nome interno è *mrLogDsc*.
- File di dati Quantum. Legge i dati del caso da un file ASCII (.dat) formato Quantum. Il nome interno è *mrPunchDsc*.
- File di dati Quancept. Legge i dati del caso da un file *.drs*, *.drz* o *.dru* Quancept. Il nome interno è *mrQdiDrsDsc*.
- Database Quanvert. Legge i dati del caso da un file *qvinfo* o *.pkd* Quanvert. Il nome interno è *mrQvDsc*.
- Database Data Collection (MS SQL Server). Legge i dati del caso da un database Microsoft SQL Server relazionale. [Per ulteriori informazioni, vedere l'argomento Stringa di connessione al database a pag. 36.](#) Il nome interno è *mrRdbDsc2*.
- File Statistics. Legge i dati del caso da un file *.sav* SPSS Statistics. Il nome interno è *mrSavDsc*.
- File Surveycraft. Legge i dati del caso da un file *.qdt* SurveyCraft. I file *.vq* e *.qdt* devono trovarsi nella stessa directory e consentire entrambi l'accesso in lettura e in scrittura. Questo non è il modo di default in cui vengono creati quando si usa SurveyCraft, pertanto per importare i dati SurveyCraft è necessario spostare uno dei file. Il nome interno è *mrScDsc*.
- File di dati Triple-S. Legge i dati del caso da un file di dati Triple-S, sia nel formato a lunghezza fissa che nel formato delimitato da virgola. Il nome interno è *mrTripleDsc*.
- Data Collection XML. Legge i dati del caso da un file di dati XML Data Collection. In genere, questo formato può essere utilizzato per trasferire i dati del caso da una posizione a un'altra. Il nome interno è *mrXmlDsc*.

Tipo di dati del caso. Specifica se il caso viene letto da un file, cartella, UDL OLE-DB o DSN ODBC e aggiorna di conseguenza le opzioni della finestra di dialogo. Le opzioni valide dipendono dal tipo di provider. Per i provider di database, è possibile specificare le opzioni per la connessione OLE-DB o ODBC. [Per ulteriori informazioni, vedere l'argomento Stringa di connessione al database a pag. 36.](#)

Progetto di dati del caso. Durante la lettura dei dati del caso da un database di Data Collection, è possibile immettere il nome del progetto. Per tutti gli altri tipi di dati del caso, questa impostazione deve essere lasciata vuota.

Importazione variabili

Importa variabili di sistema. Specifica se vengono importate le variabili di sistema, incluse le variabili che indicano lo stato dell'intervista (in corso, completata, data fine e così via). È possibile scegliere Nessuna, Tutte o Comuni.

Importa variabili "Codes". Controlla l'importazione delle variabili che rappresentano i codici utilizzati per le risposte aperte "Altro" per le variabili categoriali.

Importa variabili "SourceFile". Controlla l'importazione di variabili che contengono nomi file o immagini delle risposte analizzate.

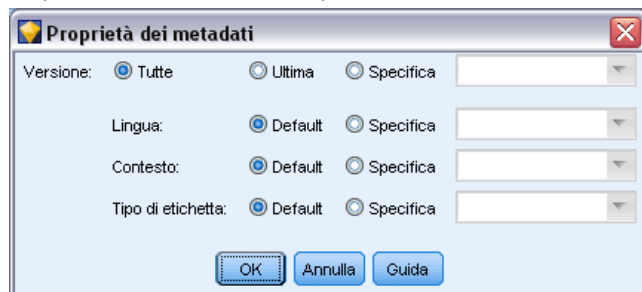
Importa variabili a risposta multipla come. Le variabili a risposta multipla possono essere importate come campi flag multipli (un insieme dicotomico multiplo), che è il metodo di default per i nuovi stream. Gli stream creati nelle versioni di IBM® SPSS® Modeler precedenti alla 12.0 importavano le risposte multiple in un unico campo, con i valori separati da virgole. Il metodo precedente è ancora supportato per consentire l'esecuzione degli stream esistenti come in precedenza, ma si consiglia di aggiornare i vecchi stream per l'utilizzo del nuovo metodo. [Per ulteriori informazioni, vedere l'argomento Importazione di insiemi di risposte multiple a pag. 37.](#)

Proprietà dei metadati dell'importazione di IBM SPSS Data Collection

Quando si importano dati di sondaggi di IBM® SPSS® Data Collection, è possibile specificare la versione del sondaggio da importare e la lingua, il contesto e il tipo di etichetta da utilizzare. Si noti che è possibile importare solo una lingua, un contesto e un tipo di etichetta alla volta.

Figura 2-16

Proprietà dei metadati dell'importazione di IBM SPSS Data Collection



Versione. Ogni versione del sondaggio può essere considerata come un'istantanea dei metadati utilizzati per raccogliere un particolare insieme di dati del caso. Via via che un questionario viene modificato, possono essere create più versioni. È possibile importare la versione più recente, tutte le versioni o una versione specifica.

- Tutte le versioni.** Selezionare questa opzione se si desidera utilizzare una combinazione (superinsieme) di tutte le versioni disponibili (denominata, talvolta, superversione). In caso di conflitti tra le versioni, le più recenti prevalgono generalmente sulle meno recenti. Per esempio, se l'etichetta di una categoria è diversa in una delle versioni, verrà utilizzato il testo dell'ultima versione.

- **Ultima versione.** Selezionare questa opzione se si desidera utilizzare la versione più recente.
- **Specifica versione.** Selezionare questa opzione se si desidera utilizzare una versione specifica del sondaggio.

L'utilizzo di tutte le versioni è utile quando, per esempio, si desidera esportare i dati del caso per più di una versione e sono state apportate modifiche alle definizioni di categorie e variabili, cioè i dati del caso raccolti con una versione non sono validi in un'altra. Selezionando tutte le versioni per le quali si desidera esportare i dati del caso, è possibile generalmente esportare i dati del caso raccolti con le diverse versioni contemporaneamente senza incorrere in errori di validità dovuti alle differenze tra le versioni. Tuttavia, a seconda delle modifiche nelle versioni, è possibile che si verifichino ancora errori di validità.

Lingua. Le domande e il testo correlato possono essere archiviati in più lingue nei metadati. È possibile utilizzare la lingua di default del sondaggio o specificare una lingua particolare. Se un elemento non è disponibile nella lingua specificata, verrà utilizzata quella di default.

Contesto. Selezionare il contesto dell'utente che si desidera utilizzare. Il contesto dell'utente determina quali testi vengono visualizzati. Per esempio, selezionare Domanda per visualizzare i testi delle domande o Analisi per visualizzare testi più brevi adatti alla visualizzazione durante l'analisi dei dati.

Tipo di etichetta. Elenca i tipi di etichette che sono stati definiti. L'impostazione di default è etichetta, che viene utilizzata per i testi delle domande nel contesto dell'utente Domanda e le descrizioni variabili nel contesto dell'utente Analisi. È possibile definire altri tipi di etichette per istruzioni, descrizioni e così via.

Stringa di connessione al database

Quando si utilizza il nodo IBM® SPSS® Data Collection per importare i dati del caso da un database tramite una connessione OLE-DB o ODBC, selezionare Modifica dalla scheda File per accedere alla finestra di dialogo Stringa di connessione che consente di personalizzare la stringa di connessione passata al provider per impostare con precisione la connessione.

Figura 2-17

Stringa di connessione importazione di IBM SPSS Data Collection

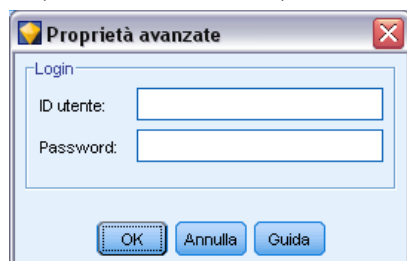


Proprietà avanzate

Quando si utilizza il nodo IBM® SPSS® Data Collection per importare i dati del caso da un database che richiede un login esplicito, selezionare Avanzate per fornire un ID utente e una password per accedere alla sorgente dati.

Figura 2-18

Proprietà avanzate dell'importazione di IBM SPSS Data Collection



Importazione di insiemi di risposte multiple

Le variabili a risposta multipla possono essere importate da IBM® SPSS® Data Collection come insiemi dicotomici multipli, con un campo flag separato per ogni valore possibile della variabile. Per esempio, se agli intervistati viene chiesto di selezionare da un elenco i musei che hanno visitato, l'insieme includerà un campo flag separato per ogni museo elencato.

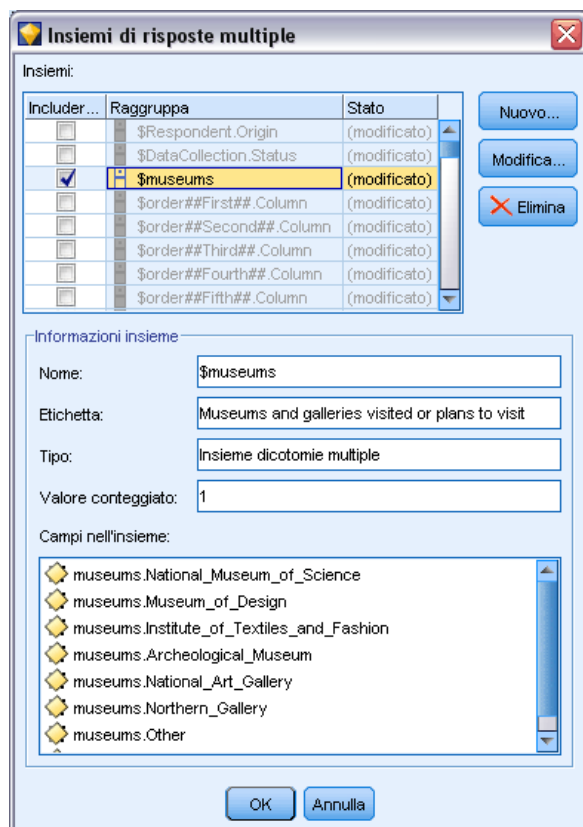
Figura 2-19

Domanda a risposta multipla

Q14	Quali musei o gallerie d'arte ha visitato o intende visitare?	
	SELEZIONARE TUTTE LE RISPOSTE APPROPRIATE.	
	Museo della Scienza	<input type="checkbox"/>
	Museo del Design.....	<input type="checkbox"/>
	Museo del Tessuto	<input type="checkbox"/>
	Museo Archeologico	<input type="checkbox"/>
	Galleria d'arte nazionale.....	<input type="checkbox"/>
	Galleria d'arte contemporanea.....	<input type="checkbox"/>
	Altro (specificare).....	<input type="checkbox"/>
	Non rispondo	<input type="checkbox"/>

Dopo l'importazione dei dati, è possibile aggiungere o modificare gli insiemi di risposte multiple di qualsiasi nodo che include una scheda Filtro. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple in il capitolo 4 a pag. 147.](#)

Figura 2-20
Finestra di dialogo *Insiemi di risposte multiple*



Importazione di risposte multiple in un unico campo (per gli stream creati nelle versioni precedenti)

Nelle precedenti versioni di IBM® SPSS® Modeler, l'importazione delle risposte multiple non si effettuava come descritto sopra ma in un unico campo, con i valori separati da virgole. Questo metodo è tuttora supportato per gli stream esistenti, ma si consiglia di aggiornare tali stream per l'utilizzo del nuovo metodo.

Note sull'importazione di colonne in IBM SPSS Data Collection

Le colonne di dati di IBM® SPSS® Data Collection vengono lette in IBM® SPSS® Modeler come riepilogato nella tabella seguente.

Tipo di colonna Data Collection	Archiviazione in SPSS Modeler	Livello di misurazione
Flag booleano (si/no)	Stringa	Flag (valori 0 e 1)
Categoriale	Stringa	Nominale
Data o timestamp	Timestamp	Continuo
Doppio (valore a virgola mobile entro un intervallo specificato)	Reale	Continuo
Lungo (valore intero entro un intervallo specificato)	Intero	Continuo

Tipo di colonna Data Collection	Archiviazione in SPSS Modeler	Livello di misurazione
Testo (descrizione testo libero)	Stringa	Typeless
Livello (indica griglie o cicli all'interno di una domanda)	Non si verifica in VDATA e non è importato in SPSS Modeler	
Oggetto (dati binari quali fax con testo scritto a mano o registrazioni vocali)	Non importato in SPSS Modeler	
Nessuno (tipo sconosciuto)	Non importato in SPSS Modeler	
Colonna Respondent.Serial (associa un ID univoco a ogni rispondente)	Intero	Typeless

Per evitare possibili incoerenze tra le etichette di valori lette dai metadati e i valori effettivi, tutti i valori dei metadati vengono convertiti in caratteri minuscoli. Per esempio, l'etichetta di valore *E1720_years* viene convertita in *e1720_years*.

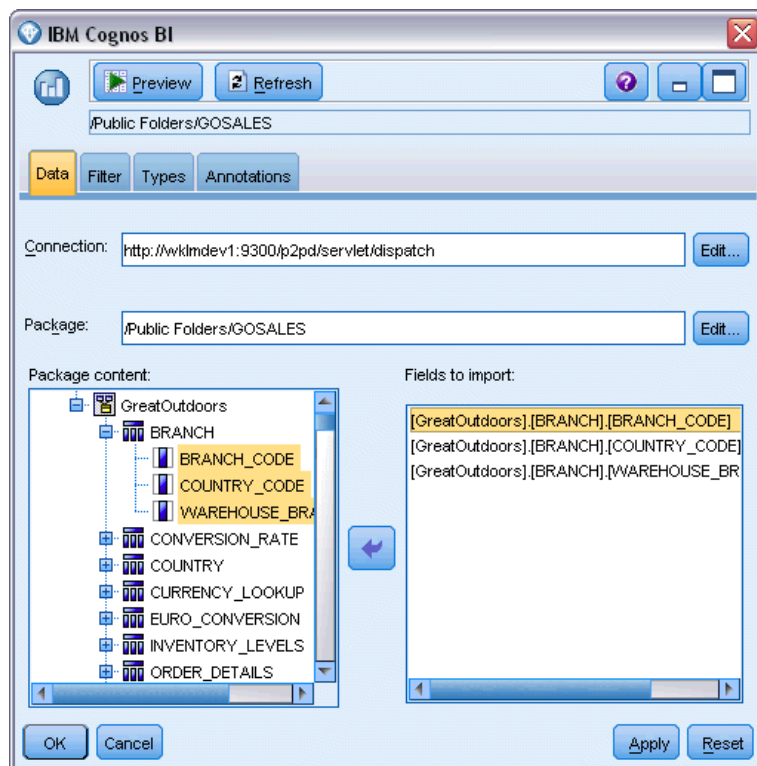
Nodo di input IBM Cognos BI

Il nodo di input IBM Cognos BI consente di importare dati del database Cognos BI nella sessione di data mining. In questo modo è possibile combinare le funzioni di business intelligence di Cognos con le funzioni di analisi predittiva di IBM® SPSS® Modeler. È possibile importare dati relazionali, DMR (dimensionally-modeled relational) e OLAP.

Da una connessione a un server Cognos si seleziona innanzitutto un pacchetto da cui importare i dati. I pacchetti contengono un modello Cognos e tutte le cartelle, le query, i report, le visualizzazioni, i collegamenti, gli URL e le definizioni dei lavori associati al modello. Un modello Cognos definisce le regole aziendali, le descrizioni dei dati, le relazioni tra i dati, le dimensioni e le gerarchie aziendali e altre attività amministrative.

Si passa quindi a selezionare gli oggetti delle query (che rappresentano le tabelle del database) o i singoli elementi delle query (che rappresentano le colonne delle tabelle) da importare dal pacchetto selezionato. *Nota:* I dati da importare devono essere in formato UTF-8.

Figura 2-21
Importazione di dati Cognos



Connessione. Fare clic sul pulsante Modifica per visualizzare una finestra di dialogo che permette di definire i dettagli di una nuova connessione Cognos da cui importare i dati. Se si è già connessi a un server Cognos mediante SPSS Modeler è possibile anche modificare i dettagli della connessione corrente. [Per ulteriori informazioni, vedere l'argomento Connessioni Cognos a pag. 41.](#)

Pacchetto. Dopo avere stabilito la connessione al server Cognos, fare clic sul pulsante Modifica accanto a questo campo per visualizzare un elenco dei pacchetti da cui è possibile importare contenuti. [Per ulteriori informazioni, vedere l'argomento Selezione dei pacchetti Cognos a pag. 41.](#)

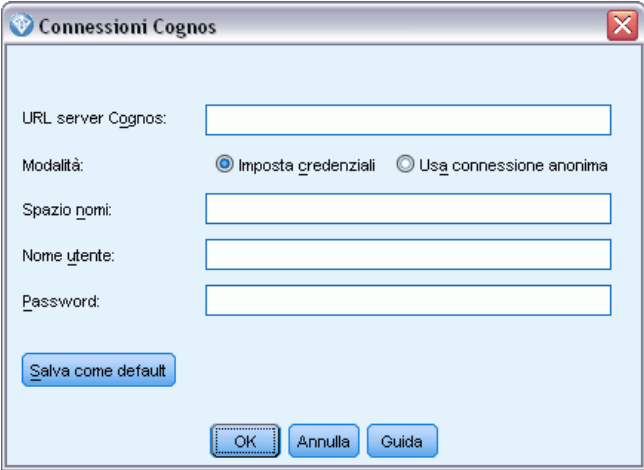
Contenuto pacchetto. Mostra il nome del pacchetto selezionato, insieme agli spazi dei nomi associati al pacchetto. Fare doppio clic su uno spazio dei nomi per visualizzare gli oggetti che è possibile importare. Selezionare gli oggetti da importare e fare clic sulla freccia destra per spostarli nella colonna Campi da importare. Se si seleziona l'oggetto di una query, vengono importati tutti gli elementi della query. Fare doppio clic sull'oggetto di una query per espanderla in modo da poter scegliere uno o più dei suoi elementi. Per selezionare più elementi, utilizzare la combinazione Ctrl-clic (selezione di singoli elementi), Maiusc-clic (selezione di un blocco di elementi) e Ctrl-A (selezione di tutti gli elementi).

Campi da importare. Elenca gli oggetti del database che verranno importati in SPSS Modeler per l'elaborazione. Se un determinato oggetto non è più necessario, selezionarlo e fare clic sulla freccia sinistra per inserirlo nuovamente nel riquadro Contenuto pacchetto. Per selezionare più oggetti, utilizzare le stesse combinazioni di tasti descritte per Contenuto pacchetto.

Connessioni Cognos

La finestra di dialogo Connessioni Cognos consente di selezionare il server Cognos da utilizzare per l'importazione o l'esportazione dei dati.

Figura 2-22
Selezione del server Cognos



URL server Cognos. Digitare l'URL del server Cognos da utilizzare per importare o esportare i dati. Se non si sa quale URL utilizzare, rivolgersi all'amministratore del sistema Cognos.

Modalità. Selezionare Imposta credenziali per accedere con uno spazio dei nomi, un nome utente e una password Cognos specifici (per esempio come amministratore). Selezionare Utilizza connessione Anonima per accedere senza credenziali utente, nel qual caso gli altri campi non vengono compilati.

Spazio dei nomi. Specificare il provider di protezione per l'autenticazione Cognos utilizzato per accedere al server. Il provider per l'autenticazione si utilizza per definire e gestire utenti, gruppi e ruoli e per controllare il processo di autenticazione.

Nome utente. Immettere il nome utente Cognos con il quale accedere al server.

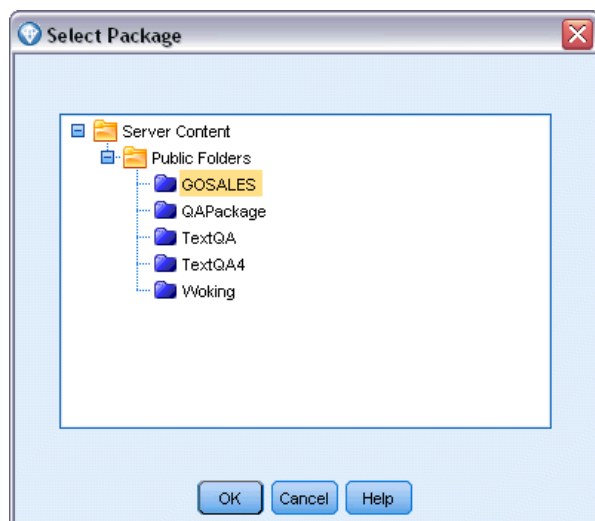
Password. Immettere la password associata al nome utente specificato.

Salva come default. Fare clic su questo pulsante per memorizzare i valori come impostazioni di default ed evitare di doverli inserire nuovamente ogni volta che si apre il nodo.

Selezione dei pacchetti Cognos

La finestra di dialogo Seleziona pacchetto consente di selezionare un pacchetto Cognos da cui importare i dati.

Figura 2-23
Selezione dei pacchetti Cognos



Oggetti Cognos. Elenco dei pacchetti disponibili nel server selezionato. Selezionare il pacchetto da utilizzare e fare clic su OK. Per ogni nodo di input Cognos BI è possibile selezionare un solo pacchetto.

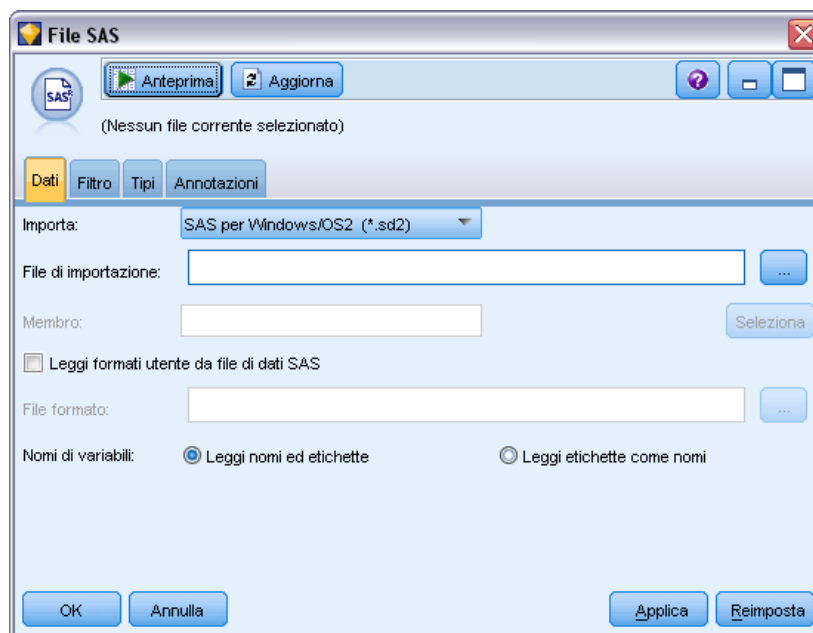
Nodo di input SAS

Il nodo di input SAS consente di importare dati in formato SAS nella sessione di data mining. È possibile importare quattro tipi di file:

- SAS per Windows/OS2 (.sd2)
- SAS per UNIX (.ssd)
- File di trasporto SAS (.tpt)
- SAS Versione 7/8/9 (.sas7bdat)

Quando i dati vengono importati tutte le variabili vengono mantenute e non viene modificato alcun tipo di variabile. Vengono selezionati tutti i casi.

Figura 2-24
Importazione di un file SAS



Impostazione delle opzioni relative al nodo di input SAS

Importa. Selezionare il tipo di file SAS da importare. È possibile selezionare SAS per Windows/OS2 (.sd2), SAS per UNIX (.SSD), File di trasporto SAS (.tpt) o SAS Versione 7/8/9 (.sas7bdat).

File di importazione. Specificare il nome del file. È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per passare alla posizione del file.

Membro. Selezionare un membro da importare dal file di trasporto SAS sopra specificato. È possibile immettere un nome per il membro oppure fare clic su **Seleziona** per scorrere l'elenco di tutti i membri nel file.

Leggi formati utente da file di dati SAS. Selezionare questa opzione per leggere i formati utente. I file SAS archiviano dati e formati di dati (per esempio etichette di variabili) in file diversi. Nella maggior parte dei casi, sarà necessario importare anche i formati. Se invece si utilizza un insieme di dati di grandi dimensioni, può essere utile deselezionare questa opzione per risparmiare memoria.

File formato. Se è richiesto un file di formato, questa casella di testo risulta attivata. È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per passare alla posizione del file.

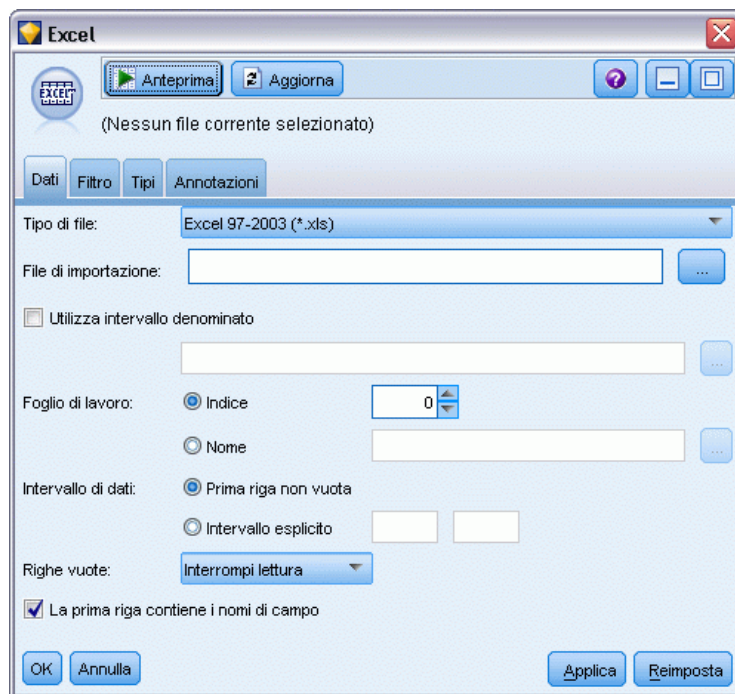
Nomi delle variabili. Selezionare un metodo per la gestione dei nomi e delle etichette di variabile durante l'importazione da un file SAS. I metadati che si sceglie di includere vengono mantenuti per la sessione di IBM® SPSS® Modeler e possono essere esportati di nuovo per essere utilizzati in SAS.

- **Leggi nomi ed etichette.** Selezionare questa opzione per leggere sia i nomi che le etichette di variabile in SPSS Modeler. Questa opzione è selezionata per default e i nomi di variabile vengono visualizzati nel nodo Tipo. Le etichette possono essere visualizzate nel generatore di espressioni, nei grafici, nei browser dei modelli e in altri tipi di output, in base alle opzioni specificate nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Leggi etichette come nomi.** Selezionare questa opzione per leggere dal file SAS le etichette di variabili descrittive anziché i nomi di campo brevi e utilizzare queste etichette come nomi di variabili in SPSS Modeler.

Nodo di input Excel

Il nodo di input Excel consente di importare dati da qualsiasi versione di Microsoft Excel.

Figura 2-25
nodo di input Excel



Tipo di file. Selezionare il tipo di file Excel che si sta importando.

File di importazione. Specifica il nome e la posizione del file del foglio di calcolo da importare.

Utilizza intervallo denominato. Consente di specificare un intervallo denominato di celle, come definito nel foglio di lavoro Excel. Fare clic sul pulsante con i puntini di sospensione (...) per scegliere dall'elenco di intervalli disponibili. Se viene utilizzato un intervallo denominato, le altre impostazioni del foglio di lavoro e dell'intervallo di dati non sono più applicabili e vengono pertanto disattivate.

Scegli foglio di lavoro. Specifica il foglio di lavoro da importare, in base all'indice o al nome.

- **Tramite indice.** Specifica il valore dell'indice del foglio di lavoro che si desidera importare, iniziando con 0 per il primo foglio di lavoro, 1 per il secondo e così via.
- **Per nome.** Specificare il nome del foglio di lavoro che si desidera importare. Fare clic sul pulsante con i puntini di sospensione (...) per scegliere dall'elenco di fogli di lavoro disponibili.

Intervallo su foglio di lavoro. È possibile importare i dati iniziando dalla prima riga non vuota o da un intervallo esplicito di celle.

- **L'intervallo inizia alla prima riga non vuota.** Individua la prima cella non vuota e la utilizza come angolo superiore sinistro dell'intervallo di dati.
- **Intervallo esplicito di celle.** Consente di specificare un intervallo esplicito in base alla riga e colonna. Per esempio, per specificare l'intervallo Excel A1:D5, è possibile inserire A1 nel primo campo e D5 nel secondo (o in alternativa, R1C1 e R5C4). Vengono restituite tutte le righe dell'intervallo specificato, incluse quelle vuote.

Su righe vuote. Se si incontra più di una riga vuota, è possibile decidere se interrompere la lettura oppure selezionare *Restituisci righe vuote* per continuare a leggere tutti i dati fino alla fine del foglio di lavoro, incluse le righe vuote.

La prima riga ha nomi di colonna. Indica che la prima riga dell'intervallo specificato deve essere utilizzata come nome di campo (colonna). Se l'opzione non è selezionata, i nomi di campo vengono generati automaticamente.

Archiviazione e livello di misurazione dei campi

Quando si leggono i valori da Excel, per default i campi con archiviazione numerica vengono letti con un livello di misurazione *Continuo* e i campi stringa come *Nominali*. Nella scheda Tipo è possibile modificare manualmente il livello di misurazione (continuo o nominale), ma l'archiviazione viene determinata automaticamente (tuttavia, se necessario, è possibile modificarla utilizzando una funzione di conversione, quale `to_integer`, in un nodo Riempimento o Nuovo campo). [Per ulteriori informazioni, vedere l'argomento Impostazione dell'archiviazione e della formattazione dei campi a pag. 27.](#)

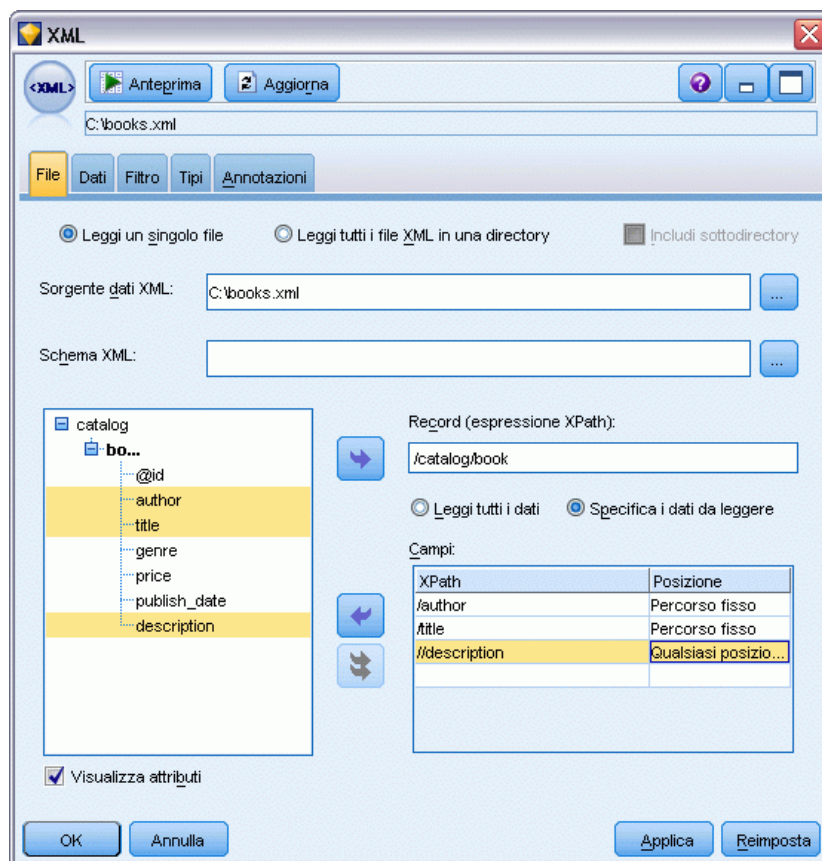
Per default, i campi contenenti sia valori numerici che valori stringa vengono letti come numeri, cioè i valori stringa vengono impostati su valori nulli (valori mancanti di sistema) in IBM® SPSS® Modeler, poiché, a differenza di Excel, SPSS Modeler non consente di utilizzare tipi di archiviazione misti all'interno dello stesso campo. Per evitare questa situazione, è possibile impostare il formato delle celle su Testo nel foglio di lavoro Excel, in modo che tutti i valori, inclusi i numeri, vengano letti come stringhe.

XML Source Node

Il nodo di input XML consente di importare i dati da un file in formato XML in uno stream IBM® SPSS® Modeler. XML è un linguaggio standard per lo scambio di dati e per molte organizzazioni rappresenta il formato preferito per questo scopo. Per esempio, l'agenzia delle entrate desidera analizzare i dati delle entrate fiscali inviate online i cui dati sono in formato XML.

L'importazione dei dati XML in uno stream SPSS Modeler consente di eseguire una vasta gamma di funzioni di analisi predittive sui dati di input. I dati XML vengono analizzati in un formato tabulare nel quale le colonne corrispondono ai diversi livelli di nidificazione degli elementi e degli attributi XML. Gli elementi XML vengono visualizzati in formato XPath (vedere <http://www.w3.org/TR/xpath20/>).

Figura 2-26
Importazione di dati XML



Leggi un singolo file. Per default, SPSS Modeler legge un singolo file, specificato nel campo Sorgente dati XML.

Leggi tutti i file XML in una directory. Scegliere questa opzione se si desidera leggere tutti i file XML in una particolare directory. Specificare la posizione nel campo Directory visualizzato. Selezionare la casella di controllo Includi sottodirectory per leggere anche i file XML di tutte le sottodirectory della directory specificata.

Sorgente dati XML. Immettere il percorso e il nome file completi del file sorgente XML che si desidera importare, oppure utilizzare il pulsante Visualizza per trovare il file.

Schema XML. (Opzionale) Specificare il percorso e il nome file completi di un file XSD o DTD dal quale si desidera leggere la struttura XML, oppure utilizzare il pulsante Visualizza per trovare questo file. Se si lascia vuoto questo campo, la struttura viene letta dal file sorgente XML. Un file

XSD o DTD può includere più elementi radice. In tal caso, quando viene selezionato un campo diverso, viene visualizzata una finestra di dialogo nella quale è possibile selezionare l'elemento radice che si desidera utilizzare. [Per ulteriori informazioni, vedere l'argomento Selezione da più elementi radice a pag. 47.](#)

Struttura XML. Un albero gerarchico che mostra la struttura del file sorgente XML (o dello schema, se specificato nel campo Schema XML. Per definire un limite del record, selezionare un elemento e fare clic sul pulsante con la freccia destra per copiarlo nel campo Record.

Visualizza attributi. Visualizza o nasconde gli attributi degli elementi XML nel campo Struttura XML.

Record (espressione XPath). Mostra la sintassi XPath per un elemento copiato dal campo della struttura XML. Questo elemento viene quindi evidenziato nella struttura XML e definisce il limite del record. Ogni volta che si incontra questo elemento nel file di input, viene creato un nuovo record. Se questo campo è vuoto, come limite del record viene utilizzato il primo elemento figlio nella radice.

Leggi tutti i dati. Per default, tutti i dati nel file di origine vengono letti nello stream.

Specifica i dati da leggere. Selezionare questa opzione se si desidera importare singoli elementi o attributi o entrambi. Selezionando questa opzione si attiva la tabella Campi, nella quale è possibile specificare i dati che si desidera importare.

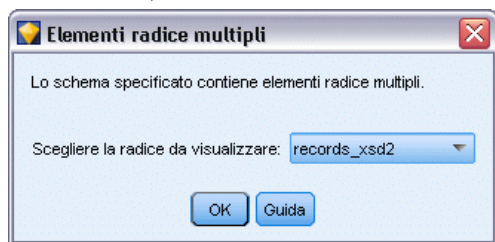
Campi. Questa tabella elenca gli elementi e gli attributi selezionati per l'importazione, se è stata selezionata l'opzione Specifica i dati da leggere. È possibile immettere la sintassi XPath di un elemento o di un attributo direttamente nella colonna XPath, oppure selezionare un elemento o un attributo nella struttura XML e fare clic sul pulsante con la freccia destra per copiare l'elemento nella tabella. Per copiare tutti gli elementi e gli attributi figlio di un elemento, selezionare l'elemento nella struttura XML e fare clic sul pulsante a doppia freccia.

- **XPath.** La sintassi XPath degli elementi da importare.
- **Posizione.** La posizione nella struttura XML degli elementi da importare. Percorso fisso mostra il percorso dell'elemento relativo all'elemento evidenziato nella struttura XML (o al primo elemento figlio nella radice se non è evidenziato alcun elemento). Qualsiasi posizione indica un elemento con il nome dato in qualsiasi posizione nella struttura XML. Personalizzato viene visualizzato se si immette una posizione direttamente nella colonna XPath.

Selezione da più elementi radice

A differenza dei file XML che possono includere un unico elemento radice, i file XSD o DTD possono contenere più elementi radice. Se presente, viene utilizzata la radice corrispondente alla radice del file di origine XML. In caso contrario, è necessario selezionare la radice da utilizzare.

Figura 2-27
Selezione da più elementi radice



Selezionare la radice da visualizzare. Selezionare l'elemento radice che si desidera utilizzare. La radice di default corrisponde al primo elemento radice della struttura XSD o DTD.

Rimozione degli spazi indesiderati dai dati di origine XML

È possibile implementare le interruzioni di linea nei dati di origine XML utilizzando la combinazione di caratteri [CR][LF]. In alcuni casi le interruzioni di linea vengono inserite al centro di una stringa di testo, per esempio:

```
<description>An in-depth look at creating applications[CR][LF]
with XML.</description>
```

Le interruzioni di linea potrebbero non essere visibili quando il file viene aperto in alcune applicazioni, ad esempio un browser Web. In ogni caso, durante la lettura dei dati nello stream attraverso il nodo di input XML, le interruzioni di linea vengono convertite in una serie di spazi, per esempio:

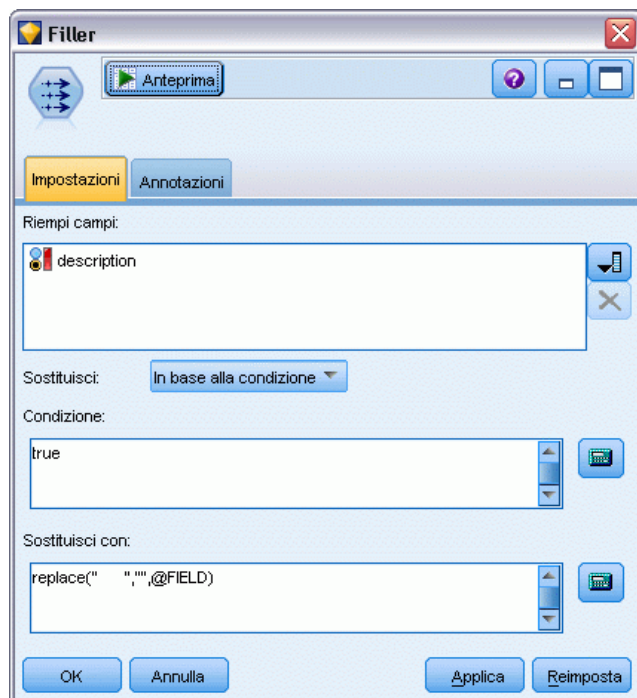
Figura 2-28
Record XML con interruzioni di linea convertite in spazi

```
description
An in-depth look at creating applications  with XML.
```


Il problema può essere risolto utilizzando un nodo Riempimento per la rimozione degli spazi indesiderati:

Figura 2-29

Nodo Riempimento con impostazioni per la rimozione degli spazi



Di seguito è riportato un esempio di come ottenere il risultato desiderato:

- ▶ Collegare un nodo Riempimento al nodo di input XML.
- ▶ Aprire il nodo Riempimento e utilizzare il pulsante di selezione dei campi per selezionare il campo con gli spazi indesiderati.
- ▶ Impostare la condizione Sostituisci su In base alla condizione e impostare Condizione su vera.
- ▶ Nel campo Sostituisci con, immettere `replace(" ", "", @FIELD)` e fare clic su OK.
- ▶ Collegare un nodo Tabella al nodo Riempimento ed eseguire lo stream.

Nell'output del nodo Tabella, il testo sarà:

Figura 2-30

Record XML con gli spazi indesiderati rimossi

```

description
An in-depth look at creating applications with XML.

```

Nodo input utente

Il nodo Input utente consente di ottenere in modo semplice dati sintetici creandoli oppure modificando dati esistenti. È utile, per esempio, quando si desidera creare un insieme di dati di test per la modellazione.

Creazione di dati

Il nodo Input utente è disponibile nella palette Input e può essere aggiunto direttamente all'area di disegno dello stream.

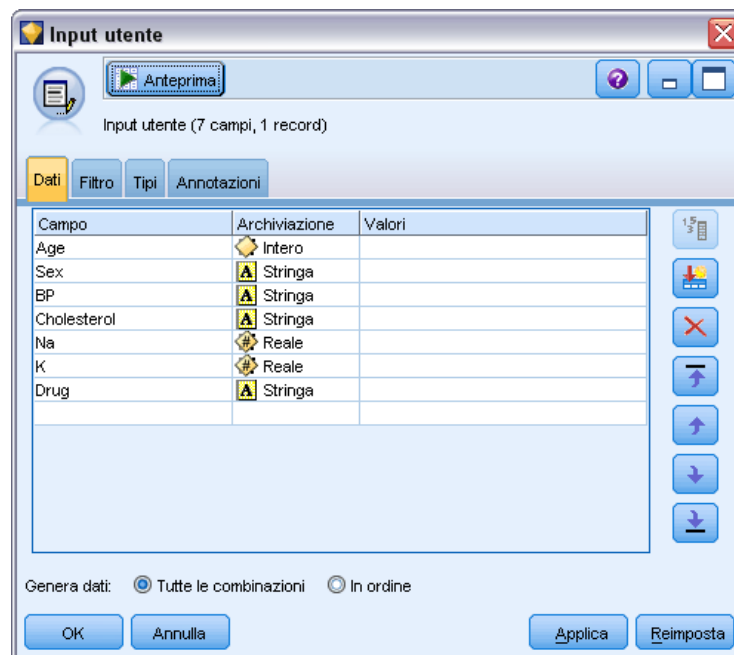
- ▶ Fare clic sulla scheda Input della palette dei nodi.
- ▶ Fare doppio clic sul nodo Input utente o trascinarlo per aggiungerlo nell'area di disegno dello stream.
- ▶ Fare doppio clic sul nodo aggiunto per aprirne la relativa finestra di dialogo e specificare campi e valori.

Nota: i nodi Input utente selezionati dalla palette Input risulteranno completamente vuoti, senza campi né informazioni sui dati. Ciò consente di creare dati sintetici completamente nuovi.

Creazione di dati da una sorgente dati esistente

Figura 2-31

Nodo Input utente generato da un nodo stream



È possibile anche generare un nodo Input utente da qualsiasi nodo non terminale dello stream:

- ▶ Stabilire il nodo dello stream che si desidera sostituire.

- ▶ Fare clic con il pulsante destro del mouse sul nodo i cui dati si desidera utilizzare per il nodo Input utente e scegliere Genera nodo Input utente dal menu di scelta rapida.
- ▶ Verrà visualizzato il nodo Input utente con tutti i processi di downstream collegati. Tale nodo sostituisce il nodo esistente nel punto stabilito all'interno dello stream di dati. Il nodo Input utente creato eredita dai metadati l'intera struttura di dati e tutte le informazioni sui tipi di campi, se disponibili.

Nota: se i dati non sono stati eseguiti in tutti i nodi dello stream, i nodi risultano non completamente istanziati, pertanto è possibile che l'archiviazione e i valori dei dati non siano disponibili durante la sostituzione di un nodo esistente con un nodo Input utente.

Impostazioni relative al nodo Input utente

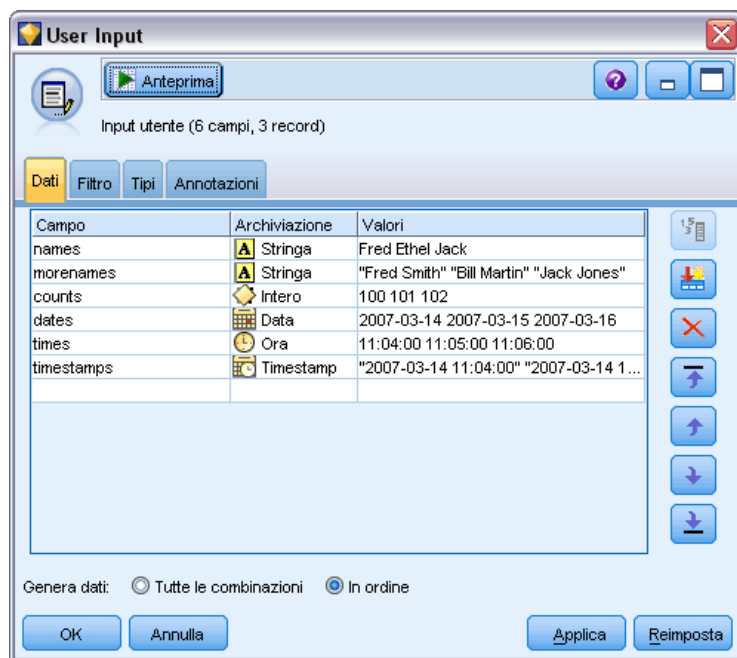
La finestra di dialogo di un nodo Input utente contiene diversi strumenti che è possibile utilizzare per immettere valori e definire la struttura dei dati per i dati sintetici. Per un nodo generato, la tabella nella scheda Dati contiene nomi di campi della sorgente dati originale. Per un nodo aggiunto dalla palette Input, questa tabella risulta invece vuota. Tramite le opzioni della tabella è possibile eseguire le operazioni seguenti:

- Aggiungere nuovi campi utilizzando il pulsante di aggiunta di un nuovo campo a destra della tabella.
- Rinominare campi esistenti.
- Specificare un tipo di archiviazione per ogni campo.
- Specificare valori.
- Modificare l'ordine dei campi visualizzati.

Inserimento di dati

Per ogni campo è possibile specificare valori o inserire valori dall'insieme di dati originale utilizzando il pulsante di selezione dei valori a destra della tabella. Per ulteriori informazioni sulla definizione di valori, vedere le regole descritte di seguito. È inoltre possibile scegliere di lasciare il campo vuoto. Ai campi vuoti verrà assegnato il valore nullo di sistema (\$null\$).

Figura 2-32
Definizione del tipo di archiviazione in un nodo Input utente generato



Per specificare i valori stringa, immetterli nella colonna dei valori, separati da spazi:

Fred Ethel Martin

Le stringhe che includono spazi possono essere racchiuse tra virgolette doppie:

"Bill Smith" "Fred Martin" "Jack Jones"

Per i campi numerici, è possibile immettere più valori nello stesso modo (ovvero separati da spazi):

10 12 14 16 18 20

oppure specificare la serie di numeri immettendo i numeri più basso e più alto (10, 20) e quindi il numero che rappresenta il criterio utilizzato per la serie (2). In base a questo metodo, si specificherà:

10,20,2

È possibile combinare questi metodi utilizzandone uno all'interno dell'altro, per esempio:

15 7 10,20,2 21 23

In questo modo si otterranno i valori seguenti:

15 7 10 12 14 16 18 20 21 23

I valori di data e ora possono essere immessi utilizzando il formato di default corrente selezionato nella finestra di dialogo Proprietà stream, per esempio:

11:04:00 11:05:00 11:06:00

2007-03-14 2007-03-15 2007-03-16

Per i valori di timestamp, costituiti da una componente di data e di ora, è necessario utilizzare le virgolette doppie:

"2007-03-14 11:04:00" "2007-03-14 11:05:00" "2007-03-14 11:06:00"

Per ulteriori informazioni, vedere i commenti indicati di seguito relativi all'archiviazione di dati.

Genera dati. Consente di specificare come vengono generati i record quando viene eseguito lo stream.

- **Tutte le combinazioni.** Genera record contenenti ogni possibile combinazione di valori dei campi; in questo modo, il valore di ogni campo comparirà in diversi record. Talvolta questo può generare una quantità di dati eccessiva, pertanto può essere opportuno fare seguire a questo nodo un nodo Campione.
- **In ordine.** Genera i record nell'ordine in cui vengono specificati i valori dei campi di dati. Il valore di ogni campo compare in un solo record. Il numero totale dei record è uguale al numero massimo di valori per un singolo campo. Se i campi hanno un numero di valori inferiore a quello massimo vengono inseriti valori non definiti (\$null\$).

Per esempio, le immissioni seguenti determineranno la creazione dei record elencati nelle tabelle riportate di seguito.

- **Età.** 30,60,10
- **Pressione.** BASSO
- **Colesterolo** NORMALE ALTO
- **Cura.** (vuoto)

Genera dati impostato su Tutte le combinazioni:

Età	Pressione	Colesterolo	Cura
30	BASSO	NORMAL	\$null\$
30	BASSO	ALTO	\$null\$
40	BASSO	NORMAL	\$null\$
40	BASSO	ALTO	\$null\$
50	BASSO	NORMAL	\$null\$
50	BASSO	ALTO	\$null\$
60	BASSO	NORMAL	\$null\$
60	BASSO	ALTO	\$null\$

Genera dati impostato su In ordine:

Età	Pressione	Colesterolo	Cura
30	BASSO	NORMAL	\$null\$
40	\$null\$	ALTO	\$null\$
50	\$null\$	\$null\$	\$null\$
60	\$null\$	\$null\$	\$null\$

Archiviazione dei dati

L'archiviazione definisce le modalità di archiviazione dei dati in un campo. Per esempio, un campo con valori 1 e 0 è un campo in cui sono archiviati dati di tipo intero. Si distingue dal livello di misurazione, che descrive l'utilizzo dei dati e non influisce sull'archiviazione. Per esempio, è possibile impostare il livello di misurazione per un campo intero con valori 1 e 0 su *Flag*. In genere, questo indica che 1=*Vero* e 0=*Falso*. Mentre l'archiviazione deve essere stabilita all'origine, il livello di misurazione può essere modificato tramite un nodo Tipo in qualsiasi punto dello stream. [Per ulteriori informazioni, vedere l'argomento Livelli di misurazione in il capitolo 4 a pag. 127.](#)

I tipi di archiviazione disponibili sono:

- **Stringa.** Utilizzato per i campi che contengono dati non numerici, detti anche dati alfanumerici. Una stringa può includere qualsiasi sequenza di caratteri, quale *mario*, *Classe 2* o *1234*. Si noti che i numeri presenti nelle stringhe non possono essere utilizzati nei calcoli.
- **Intero.** Un campo i cui valori sono numeri interi.
- **Reale.** I valori sono numeri che possono includere decimali (non limitati ai numeri interi). Il formato di visualizzazione viene specificato nella finestra di dialogo Proprietà stream e può essere ignorato per singoli campi in un nodo Tipo (scheda Formato). [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Data.** Valori di data specificati in un formato standard quale anno, mese e giorno (per esempio, 26.09.07). Il formato specifico è indicato nella finestra di dialogo Proprietà stream.
- **Tempo.** Tempo misurato come durata. Per esempio, se una chiamata di assistenza è durata 1 ora, 26 minuti e 38 secondi, potrà essere rappresentata come 01:26:38, in base al formato dell'ora corrente specificato nella finestra di dialogo Proprietà stream.
- **Timestamp.** Valori che includono sia i componenti data e ora, per esempio 2007-09-26 09:04:00, anche questi in base al formato dell'ora corrente specificato nella finestra di dialogo Proprietà stream. Si noti che può essere necessario racchiudere i valori di timestamp tra virgolette per garantire che vengano interpretati come un unico valore anziché come valori di data e ora separati (per esempio quando si immettono valori in un nodo Input utente).

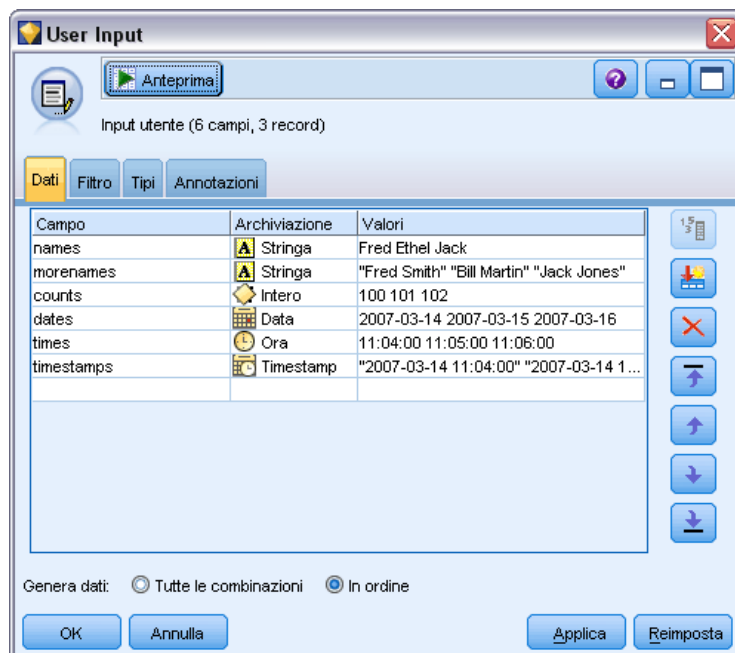
Conversioni della classe di archiviazione. È possibile convertire la classe di archiviazione di un campo utilizzando una vasta gamma di funzioni di conversione, per esempio `to_string` e `to_integer`, in un nodo Riempimento. [Per ulteriori informazioni, vedere l'argomento Conversione della classe di archiviazione mediante il nodo Riempimento in il capitolo 4 a pag. 168.](#) Si noti che le funzioni di conversione, e qualsiasi altra funzione che richieda un tipo di input specifico quale un valore di data o ora, dipendono dai formati correnti specificati nella finestra di dialogo Proprietà stream. Per esempio, se si desidera convertire un campo stringa con valori *Gen 2003*, *Feb 2003* e così via in campo di archiviazione della data, selezionare MES AAAA come formato di data di default per lo stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#) Le funzioni di conversione sono disponibili anche dal nodo Nuovo campo per la conversione temporanea durante un calcolo di derivazione. È inoltre possibile utilizzare il nodo Nuovo campo per eseguire altre manipolazioni, quali ricodifiche di campi stringa con valori categoriali. [Per ulteriori informazioni, vedere l'argomento Ricodifica di valori con il nodo Nuovo campo in il capitolo 4 a pag. 165.](#)

Letture di dati misti. Quando si leggono campi con archiviazione numerica (numeri interi, reali, ora, timestamp o data), gli eventuali valori non numerici vengono impostati come valori nulli o valori mancanti di sistema poiché, a differenza di altre applicazioni, IBM® SPSS® Modeler non consente di utilizzare tipi di archiviazione misti all'interno dello stesso campo. Per evitare questo inconveniente, è necessario modificare opportunamente il tipo di archiviazione nel nodo di input o nell'applicazione esterna per fare in modo che tutti i campi contenenti dati misti siano letti come stringhe.

Nota: i nodi Input utente generati possono includere informazioni raccolte dal nodo di input, se tale nodo è istanziato. Invece, un nodo non istanziato non conterrà alcuna informazione su archiviazione o tipo di utilizzo.

Figura 2-33

Definizione del tipo di archiviazione in un nodo Input utente generato



Regole per la definizione dei valori

Per i campi simbolici, è necessario lasciare uno spazio tra i valori, per esempio:

ALTO MEDIO BASSO

Per i campi numerici, è possibile immettere più valori nello stesso modo (ovvero separati da spazi):

10 12 14 16 18 20

oppure specificare la serie di numeri immettendo i numeri più basso e più alto (10, 20) e quindi il numero che rappresenta il criterio utilizzato per la serie (2). In base a questo metodo, si specificherà:

10,20,2

È possibile combinare questi metodi utilizzandone uno all'interno dell'altro, per esempio:

1 5 7 10,20,2 21 23

In questo modo si otterranno i valori seguenti:

1 5 7 10 12 14 16 18 20 21 23

Schede comuni a tutti i nodi di input

Le opzioni seguenti possono essere specificate per tutti i nodi Input facendo clic sulla scheda corrispondente:

- **Scheda dati.** Utilizzata per modificare il tipo di archiviazione di default.
- **Scheda Filtro.** Utilizzata per eliminare o rinominare campi di dati. Questa scheda include le stesse opzioni disponibili nel nodo Filtro. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di filtro in il capitolo 4 a pag. 144.](#)
- **Scheda tipi.** Utilizzata per impostare i livelli di misurazione. Questa scheda include le stesse opzioni disponibili nel nodo Tipo.
- **Scheda Annotazioni.** Utilizzata per tutti i nodi, questa scheda fornisce opzioni per rinominare i nodi, uno strumento personalizzato e la possibilità di archiviare annotazioni estese. [Per ulteriori informazioni, vedere l'argomento Annotazioni in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Impostazione dei livelli di misurazione nel nodo di input

È necessario specificare le proprietà dei campi in un nodo di input o in un nodo Tipo distinto. Questa funzione è analoga in entrambi i nodi. Sono disponibili le seguenti proprietà:

- **Campo.** Fare doppio clic sul nome di un campo per specificare valore ed etichette di campo per i dati in IBM® SPSS® Modeler. Per esempio, qui è possibile visualizzare o modificare i metadati dei campi importati da IBM® SPSS® Statistics. Analogamente, è possibile creare nuove etichette per campi e loro valori. Le etichette specificate qui vengono visualizzate in SPSS Modeler in base alle selezioni effettuate nella finestra di dialogo Proprietà stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Misurazione.** Livello di misurazione utilizzato per descrivere le caratteristiche dei dati in un campo specifico. Se tutti i dettagli di un campo sono noti, il campo viene definito **completamente istanziato**. [Per ulteriori informazioni, vedere l'argomento Livelli di misurazione in il capitolo 4 a pag. 127.](#)

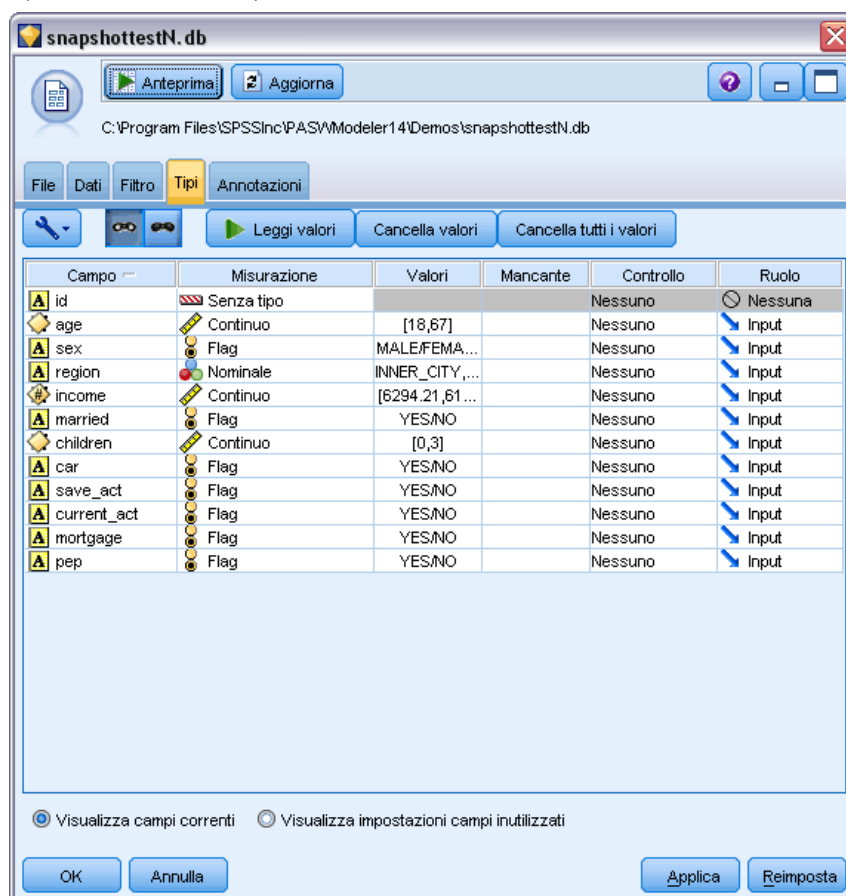
Nota: Il livello di misurazione è diverso dalla classe di archiviazione di un campo, che indica se i dati sono archiviati come stringhe, numeri interi, numeri reali, date, ore o timestamp.

- **Valori.** Questa colonna consente di specificare le opzioni di lettura dei valori dei dati dall'insieme di dati. L'opzione Specifica permette invece di specificare i livelli di misurazione e i valori in una finestra di dialogo a parte. È anche possibile scegliere di passare i campi senza leggerne i valori. [Per ulteriori informazioni, vedere l'argomento Valori dei dati in il capitolo 4 a pag. 131.](#)

- **Mancante/i.** Utilizzata per specificare come vengono gestiti i valori mancanti del campo. [Per ulteriori informazioni, vedere l'argomento Definizione dei valori mancanti in il capitolo 4 a pag. 137.](#)
- **Controllo.** In questa colonna è possibile impostare opzioni per assicurarsi che i valori dei campi siano inclusi nei valori o nell'intervallo specificato. [Per ulteriori informazioni, vedere l'argomento Controllo dei valori del tipo in il capitolo 4 a pag. 137.](#)
- **Ruolo.** Proprietà utilizzata per indicare ai nodi Modelli se i campi verranno trattati come campi di Input (campi predittori) o Obiettivo (campi previsti) nel processo di apprendimento automatico. Entrambi e Nessuno sono altri ruoli possibili, oltre a Partizione, che indica un campo utilizzato per eseguire la partizione dei record in campioni separati per addestramento, test e validazione. Il valore Suddivisione specifica che i modelli separati verranno costruiti per ciascun valore possibile del campo. [Per ulteriori informazioni, vedere l'argomento Impostazione del ruolo del campo in il capitolo 4 a pag. 138.](#)

[Per ulteriori informazioni, vedere l'argomento Nodo Tipo in il capitolo 4 a pag. 125.](#)

Figura 2-34
Opzioni della scheda Tipi



Istanziamento nel nodo Input

Per ottenere informazioni sull'archiviazione dei dati e sui valori dei campi sono disponibili due metodi di istanziazione. L'**istanziamento** può essere eseguita a livello del nodo di input quando i dati vengono importati in IBM® SPSS® Modeler o tramite l'aggiunta di un nodo Tipo nello stream di dati.

L'istanziamento a livello del nodo di input risulta utile nei casi seguenti:

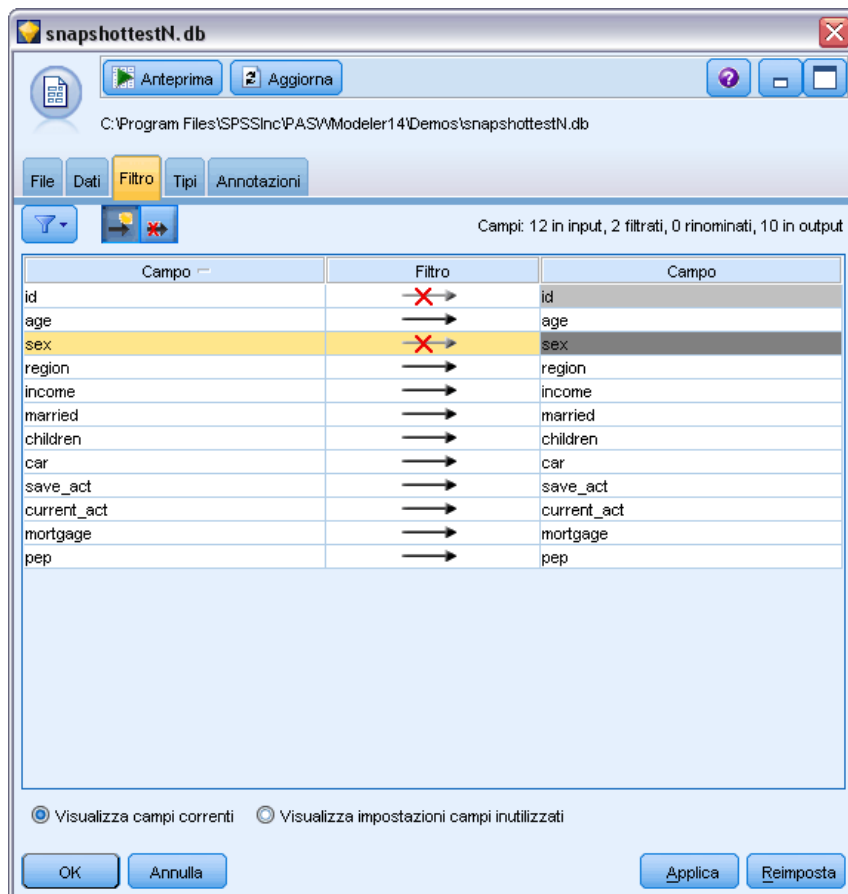
- Quando l'insieme di dati è di dimensioni ridotte.
- Quando si stabilisce di derivare nuovi campi utilizzando il generatore di espressioni (l'istanziamento rende i valori dei campi disponibili mediante il generatore di espressioni).

In genere, se l'insieme di dati non è di grandi dimensioni e non si prevede di aggiungere campi nello stream in seguito, l'istanziamento a livello del nodo di input risulta il metodo più vantaggioso.

Filtro di campi dal nodo Input

La scheda Filtro nella finestra di dialogo di un nodo di input consente di escludere campi dalle operazioni a valle in base a un'analisi iniziale dei dati. Questo risulta utile, per esempio, se nei dati sono presenti campi duplicati o se si conoscono i dati in modo tanto approfondito da poter escludere i campi irrilevanti. In alternativa, è possibile aggiungere un nodo Filtro separato più avanti nello stream. Questa funzione è analoga in entrambi i casi. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di filtro in il capitolo 4 a pag. 144.](#)

Figura 2-35
Filtro di campi dal nodo di input



Nodi Operazioni su record

Panoramica sui nodi Operazioni su record

I nodi Operazioni su record vengono utilizzati per apportare modifiche ai dati a livello di record. Queste operazioni risultano importanti durante le fasi **Data Understanding** e **Data Preparation** del processo di data mining, poiché consentono di adattare i dati alle specifiche esigenze aziendali.

In base ai risultati dell'esplorazione dati eseguita utilizzando il nodo Esplora (palette Output), per esempio, è possibile decidere se si desidera unire i record relativi agli acquisti effettuati da un cliente negli ultimi tre mesi. Utilizzando un nodo Unione sarà possibile unire i record in base ai valori di un campo chiave, quale *ID cliente*. Oppure, nel caso in cui non fosse più possibile gestire le informazioni di un database contenente oltre un milione di record relativi ai risultati di un sito Web, utilizzando un nodo Campione è possibile selezionare un sottoinsieme di dati da utilizzare per la modellazione.

La palette Operazioni su record contiene i seguenti nodi:



Il nodo Seleziona consente di selezionare o scartare un sottoinsieme di record dallo stream dei dati basato su una condizione specifica. Per esempio, è possibile selezionare i record relativi a una determinata area vendite. [Per ulteriori informazioni, vedere l'argomento Nodo Seleziona a pag. 61.](#)



Il nodo Campione seleziona un sottoinsieme di record. Sono supportati vari tipi di campioni, inclusi campioni stratificati, raggruppati e non casuali (strutturati). Il campionamento può essere utile per migliorare le prestazioni e per selezionare gruppi di record correlati o transazioni per un'analisi. [Per ulteriori informazioni, vedere l'argomento Nodo Campione a pag. 63.](#)



Il nodo Bilanciamento corregge sbilanciamenti in un insieme di dati in modo che soddisfi una determinata condizione. La direttiva di bilanciamento regola la proporzione di record in cui una condizione è vera in base al fattore specificato. [Per ulteriori informazioni, vedere l'argomento Nodo Bilanciamento a pag. 71.](#)



Il nodo Aggregazione sostituisce una sequenza di record di input con record di output aggregati di riepilogo. [Per ulteriori informazioni, vedere l'argomento Nodo Aggregazione a pag. 73.](#)



Il nodo Aggregazione RFM (Recency, Frequency, Monetary, Passato recente, Frequenza, Monetario) consente di prendere in considerazione i dati storici delle transazioni dei clienti, eliminare i dati non utilizzati e combinare tutti i dati delle transazioni rimanenti in un'unica riga che indica quanto tempo è trascorso dall'ultima transazione, il numero di transazioni effettuate e il valore monetario totale delle transazioni. [Per ulteriori informazioni, vedere l'argomento Nodo Aggregazione RFM a pag. 75.](#)



Il nodo Ordina ordina record in ordine crescente o decrescente in base ai valori di uno o più campi. [Per ulteriori informazioni, vedere l'argomento Nodo Ordina a pag. 78.](#)



Il nodo Unione prende più record di input e crea un singolo record di output contenente tutti o alcuni campi di input. È utile per unire dati da sorgenti diverse, per esempio dati interni sui clienti e dati demografici acquistati. [Per ulteriori informazioni, vedere l'argomento Nodo Unione a pag. 80.](#)



Il nodo Accodamento concatena insieme di record. Può essere utilizzato per combinare insieme di dati con strutture simili ma dati diversi. [Per ulteriori informazioni, vedere l'argomento Nodo Accodamento a pag. 90.](#)



Il nodo Elimina duplicati rimuove record duplicati passando il primo record distinto nello stream di dati oppure scartando il primo record e passando nello stream tutti i duplicati. [Per ulteriori informazioni, vedere l'argomento Nodo Elimina duplicati a pag. 92.](#)

Molti di questi nodi richiedono l'utilizzo di espressioni CLEM. Se si conosce il linguaggio CLEM, è possibile immettere un'espressione nel campo. Tuttavia, tutti i campi per le espressioni presentano un pulsante che apre il Generatore di espressioni CLEM, che aiuta a creare tali espressioni in modo automatico. [Per ulteriori informazioni, vedere l'argomento Generatore di espressioni in il capitolo 7 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

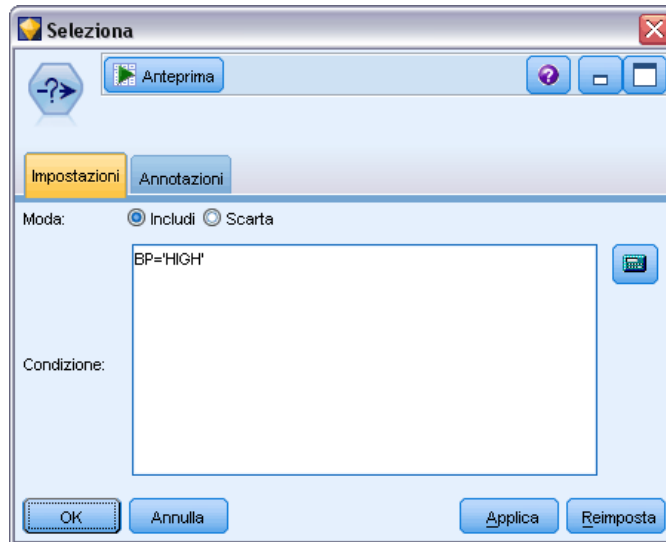
Figura 3-1
Pulsante del Generatore di espressioni



Nodo Seleziona

È possibile utilizzare i nodi Seleziona per selezionare o scartare un sottoinsieme di record dallo stream di dati in base a una condizione specificata, per esempio Pressione (pressione sanguigna) = "ALTO".

Figura 3-2
Finestra di dialogo del nodo Seleziona



Modalità. Specifica se i record che soddisfano la condizione verranno inclusi nello stream di dati oppure esclusi.

- **Includi.** Selezionare questa opzione per includere i record che soddisfano la condizione di selezione.
- **Scarta.** Selezionare questa opzione per escludere i record che soddisfano la condizione di selezione.

Condizione. Visualizza la condizione di selezione che verrà utilizzata per la verifica di ogni record. Per specificare questa condizione è necessario utilizzare un'espressione CLEM. Immettere un'espressione nella finestra oppure utilizzare il Generatore di espressioni facendo clic sul pulsante con la calcolatrice (Generatore espressioni) a destra della finestra.

Se si decide di eliminare i record in base a una condizione, per esempio la seguente:

```
(var1='value1' and var2='value2')
```

il nodo Seleziona elimina per default anche i record con valori nulli per tutti i campi di selezione. Per evitare ciò, accodare la seguente condizione a quella originale:

```
and not(@NULL(var1) and @NULL(var2))
```

I nodi Seleziona vengono inoltre utilizzati per selezionare una proporzione di record. In genere, per questa operazione si utilizza un nodo diverso, ovvero il nodo Campione. Se tuttavia la condizione che si desidera specificare risulta più complessa dei parametri disponibili, è possibile creare una condizione personalizzata utilizzando il nodo Seleziona. Per esempio, è possibile creare la condizione seguente:

```
Pressione = "ALTO" and random(10) <= 4
```

In questo modo, circa il 40% dei record contenenti valori di pressione alta verrà selezionato e passato a valle per ulteriori analisi.

Nodo Campione

È possibile utilizzare i nodi Campione per selezionare un sottoinsieme di record per l'analisi, oppure per specificare una percentuale di record da scartare. Sono supportati vari tipi di campioni, inclusi campioni stratificati, raggruppati e non casuali (strutturati). Il campionamento può essere utilizzato per vari motivi:

- Per migliorare le prestazioni stimando i modelli in un sottoinsieme di dati. I modelli stimati a partire da un campione offrono spesso lo stesso livello di precisione di quelli derivati dall'insieme di dati completo e molti di essi sono anche più precisi; le migliori prestazioni consentono così di sperimentare metodi diversi che altrimenti non verrebbero tentati.
- Per selezionare gruppi di record o di transazioni correlati per l'analisi, per esempio per selezionare tutti gli elementi di un carrello (o panierino) per gli acquisti online, oppure tutte le proprietà in una specifica vicinanza.
- Per identificare unità o casi per l'ispezione casuale a fini di controllo qualità, prevenzione delle frodi o sicurezza.

Nota: se si desidera semplicemente ripartire i dati in campioni di addestramento e di test ai fini della validazione, è possibile utilizzare invece un nodo Partizione. [Per ulteriori informazioni, vedere l'argomento Nodo Partizione in il capitolo 4 a pag. 193.](#)

Tipi di campioni

Campioni raggruppati. Gruppi o cluster di campioni anziché unità singole. Per esempio, si supponga di avere un file di dati con un record per studente. Se si effettua il raggruppamento per scuola e la dimensione del campione è 50%, verrà scelto il 50% delle scuole e saranno presi in considerazione tutti gli studenti di ciascuna scuola selezionata. Gli studenti delle scuole non selezionate saranno rifiutati. In media, ci si aspetta che venga selezionato il 50% degli studenti, ma poiché le scuole hanno dimensioni diverse, la percentuale potrebbe non essere esatta. Analogamente, è possibile raggruppare gli articoli del carrello in base all'ID transazione per essere certi che siano mantenuti tutti gli articoli delle transazioni selezionate. Per un esempio di raggruppamento delle proprietà in base alla provincia, consultare il modello di stream `complexsample_property.str`.

Campioni stratificati. Campioni selezionati in modo indipendente all'interno di sottogruppi di popolazione che non si sovrappongono, o strati. Per esempio, è possibile fare in modo che uomini e donne siano campionati in proporzioni uguali, oppure che sia rappresentata ogni regione o gruppo socio-economico all'interno di una popolazione urbana. È anche possibile specificare una dimensione diversa del campione per ogni strato (per esempio, se si pensa che un gruppo sia stato sottorappresentato nei dati originali). Per un esempio di stratificazione delle proprietà in base alla regione, consultare il modello di stream `complexsample_property.str`.

Campionamento sistematico o 1-ogni-n. Se è difficile ottenere la selezione casuale, è possibile campionare le unità in modo sistematico (a un intervallo fisso) o sequenziale.

Pesi di campionamento. I pesi di campionamento vengono calcolati automaticamente durante la formulazione di un campione complesso e corrispondono approssimativamente alla "frequenza" con cui ogni unità campionata è rappresentata nei dati originali. Pertanto, la somma dei pesi nel campione stima le dimensioni dei dati originali.

Lista di campionamento

Una lista di campionamento definisce la fonte potenziale di casi da includere in un campione o studio. In alcuni casi può essere possibile identificare ogni singolo membro di una popolazione e includerlo in un campione, per esempio durante il campionamento degli articoli risultanti da una linea di produzione. Più spesso, non sarà possibile accedere a ogni caso possibile. Per esempio, non è possibile essere certi di chi parteciperà a una consultazione elettorale fino a dopo la votazione. In tal caso, è possibile utilizzare il registro elettorale come lista di campionamento, anche se alcuni cittadini iscritti non voteranno e alcuni individui non iscritti al momento della verifica del registro si recheranno effettivamente alle urne. Le persone non incluse nella lista di campionamento non potranno essere campionate. Se la lista di campionamento rispecchi fedelmente la natura della popolazione che si sta tentando di valutare è una questione che andrà gestita nei singoli casi della vita reale.

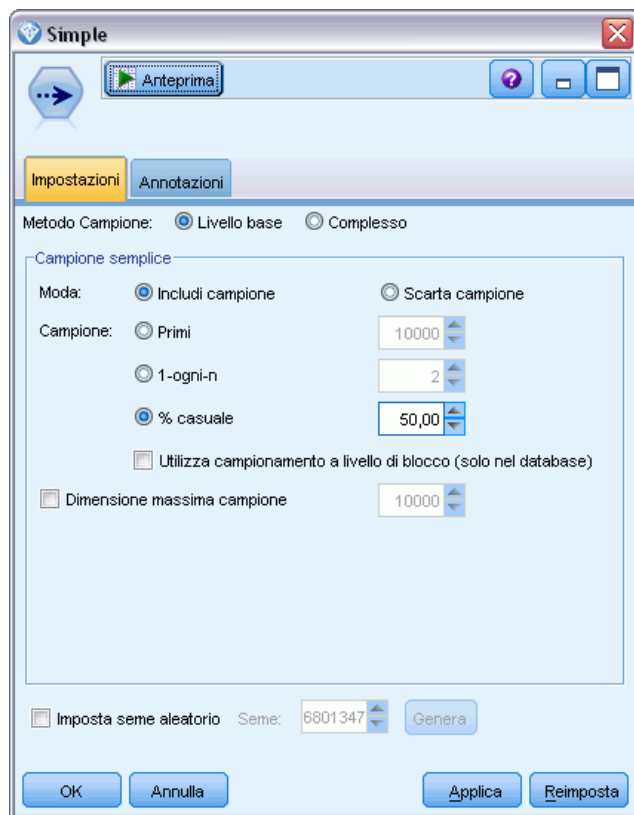
Opzioni del nodo Campione

È possibile scegliere il metodo Semplice o Complesso in base alle esigenze.

Opzioni di campionamento semplice

Il metodo di campionamento Semplice consente di selezionare una percentuale casuale di record, di selezionare record contigui oppure di selezionare ogni N record.

Figura 3-3
Opzioni di campionamento semplice



Modalità. Selezionare l'opzione desiderata per passare (includere) o scartare (escludere) record:

- **Includi campione.** Include i record selezionati nello stream di dati e scarta tutti gli altri. Per esempio, se si seleziona Includi campione come Modalità e si imposta l'opzione 1-ogni-n su 5, verrà incluso un record ogni cinque, generando un insieme di dati che corrisponde approssimativamente a un quinto della dimensione originale. Questa è la modalità di default durante il campionamento dei dati e l'unica modalità durante l'utilizzo del metodo complesso.
- **Scarta campione.** Esclude i record selezionati e include tutti gli altri. Per esempio, se si seleziona Scarta campione e si imposta l'opzione 1-ogni-n su 5, verrà scartato un record ogni cinque. Questa modalità è disponibile solo con il metodo semplice.

Esempio. Specificare il metodo di campionamento selezionando una delle opzioni seguenti:

- **Primi.** Selezionare questa opzione per utilizzare un campionamento di dati contigui. Per esempio, se la dimensione massima del campione è impostata su 10000, verranno selezionati i primi 10000 record.
- **1-ogni-n.** Selezionare questa opzione per campionare dati passando o scartando un record ogni n record. Per esempio, se N è impostato su 5, verrà selezionato un record su cinque.
- **% casuale.** Selezionare questa opzione per campionare una percentuale casuale dei dati. Per esempio, se si imposta la percentuale su 20, il 20% dei dati verrà passato nello stream di dati o scartato dallo stream, a seconda della modalità selezionata. Utilizzare il campo disponibile

per specificare una percentuale di campionamento. È inoltre possibile specificare un valore seme mediante il controllo Imposta seme casuale.

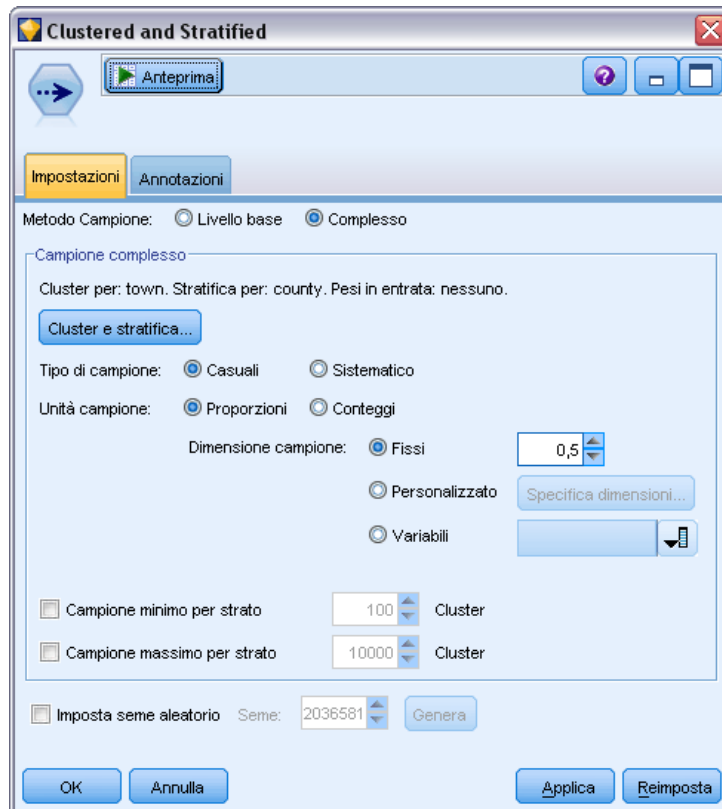
Utilizza campionamento block level (solo in-database). Questa opzione è attivata solo se si sceglie il campionamento con percentuale casuale quando si esegue il mining in-database su un database Oracle o IBM DB2. In questi casi, il campionamento block level può essere più efficiente.

Dimensione massima campione. Specifica il numero massimo di record da includere nel campione. Questa opzione è ridondante e risulta pertanto disattivata se vengono selezionati Primi e Includi. Si noti inoltre che, se utilizzata in combinazione con l'opzione % casuale, questa impostazione può impedire la selezione di determinati record. Per esempio, se un insieme di dati contiene 10 milioni di record e si seleziona il 50% dei record dopo avere impostato una dimensione massima campione di 3 milioni di record, verrà selezionato solo il 50% dei primi 6 milioni di record e i restanti 4 milioni non avranno alcuna possibilità di essere selezionati. Per evitare questa limitazione, selezionare il metodo di campionamento Complesso e richiedere un campione casuale di tre milioni di record senza specificare alcuna variabile cluster o di stratificazione.

Opzioni di campionamento complesso

Le opzioni di campionamento complesso consentono un controllo più preciso del campione che comprende campioni raggruppati, stratificati e ponderati unitamente ad altre opzioni.

Figura 3-4
Opzioni di campionamento complesso



Cluster e stratifica. Consente di specificare campi di raggruppamento, stratificazione e peso di input, se necessario. [Per ulteriori informazioni, vedere l'argomento Impostazioni di raggruppamento e stratificazione a pag. 68.](#)

Tipo di campionamento.

- **Casuale.** Seleziona casualmente i cluster o i record all'interno di ogni strato.
- **Sistematico.** Seleziona i record a intervalli fissi. Questa opzione funziona come il metodo *l-ogni-n*, tranne che per la posizione del primo record, che cambia in base a un seme aleatorio. Il valore di *N* viene determinato automaticamente in base alla dimensione o alla proporzione del campione.

Unità campione. È possibile selezionare proporzioni o conteggi come unità campione di base.

Dimensione campione. La dimensione campione può essere specificata in vari modi:

- **Fissa.** Consente di specificare la dimensione globale del campione, sotto forma di conteggio o proporzione.
- **Personalizzato.** Consente di specificare la dimensione del campione per ogni sottogruppo o strato. Questa opzione è disponibile solo se è stato specificato un campo di stratificazione nella sottofinestra di dialogo Cluster e stratifica.
- **Variabile.** Consente all'utente di scegliere un campo che definisce la dimensione del campione per ogni sottogruppo o strato. Questo campo deve presentare lo stesso valore per ogni record all'interno di un particolare strato; per esempio, se il campione è stratificato per regione, tutti i record con *regione = Lazio* devono presentare lo stesso valore. Il campo deve essere numerico e i suoi valori devono corrispondere alle unità campione selezionate. Per le proporzioni, i valori devono essere superiori a 0 e inferiori a 1; per i conteggi, il valore minimo è 1.

Campione minimo per strato. Specifica un numero minimo di record (oppure un numero minimo di cluster, se è specificato un campo cluster).

Campione massimo per strato. Specifica un numero massimo di record o cluster. Se si seleziona questa opzione senza specificare un campo cluster o di stratificazione, verrà selezionato un campione casuale o sistematico delle dimensioni specificate.

Imposta seme aleatorio. Quando si esegue il campionamento o il partizionamento di record in base a una percentuale casuale, questa opzione consente di duplicare gli stessi risultati in un'altra sessione. Specificando il valore iniziale utilizzato dal generatore di numeri casuali, è possibile garantire che vengano assegnati gli stessi record a ogni esecuzione del nodo. Immettere il valore desiderato, oppure fare clic sul pulsante Genera per generare automaticamente un valore casuale. Se questa opzione non è selezionata, verrà generato un campione diverso ogni volta che si esegue il nodo.

Nota: quando si utilizza l'opzione Imposta seme aleatorio con record letti da un database, potrebbe essere necessario un nodo Ordina prima di eseguire il campionamento, per garantire lo stesso risultato ogni volta che viene eseguito il nodo. Questo si verifica perché il seme aleatorio dipende dall'ordine dei record, per il quale non si ha la garanzia che rimanga invariato in un database relazionale. [Per ulteriori informazioni, vedere l'argomento Nodo Ordina a pag. 78.](#)

Impostazioni di raggruppamento e stratificazione

La finestra di dialogo Cluster e stratifica consente di selezionare i campi cluster, stratificazione e peso durante la formulazione di un campionamento complesso.

Figura 3-5

Impostazioni dei capi cluster e stratificazione



Raggruppamenti. Specifica un campo categoriale utilizzato per raggruppare i record. I record vengono campionati in base all'appartenenza ai cluster, con alcuni cluster inclusi e altri no. Tuttavia, se un record qualsiasi di un dato cluster è incluso, tutti i record lo sono. Per esempio, durante l'analisi delle associazioni di prodotti nei carrelli, è possibile raggruppare gli articoli in base all'ID transazione per assicurarsi che siano mantenuti tutti gli articoli delle transazioni selezionate. Anziché campionare i record, operazione che comporterebbe la distruzione delle informazioni su quali articoli sono venduti insieme, è possibile campionare le transazioni per fare in modo che vengano mantenuti tutti i record delle transazioni selezionate.

Stratifica per. Specifica un campo categoriale utilizzato per stratificare i record in modo che i campioni vengano selezionati in modo indipendente all'interno di sottogruppi della popolazione che non si sovrappongono, o strati. Per esempio, se si seleziona un campione del 50% stratificato in base al sesso, verranno considerati due campioni del 50%, uno per gli uomini e uno per le donne. Gli strati possono essere, per esempio, gruppi socioeconomici, categorie professionali, fasce di età o gruppi etnici, consentendo adeguate dimensioni di campionamento per i sottogruppi di interesse. Se nell'insieme di dati vi è un numero di donne tre volte superiore al numero di uomini, questo rapporto verrà mantenuto dal campionamento separato di ogni gruppo. È anche possibile specificare più campi di stratificazione (per esempio, campionamento delle linee di prodotti all'interno delle regioni o viceversa).

Nota: se si effettua la stratificazione in base a un campo con valori mancanti (valori vuoti o mancanti di sistema, stringhe vuote, spazi bianchi e valori vuoti o mancanti definiti dall'utente), non è possibile specificare dimensioni di campionamento personalizzate per gli strati. Se si desidera utilizzare dimensioni di campionamento personalizzate durante la stratificazione in base a un campo con valori mancanti o vuoti, è necessario compilarli a monte. [Per ulteriori informazioni, vedere l'argomento Gestione dei campi con valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Utilizza peso di input. Specifica un campo utilizzato per pesare i record prima del campionamento. Per esempio, se il campo peso presenta valori da 1 a 5, i record con peso 5 hanno cinque volte più probabilità di essere selezionati. I valori di questo campo verranno sovrascritti dai pesi di output finali generati dal nodo (vedere il paragrafo successivo).

Nuovo peso di output. Specifica il nome del campo nel quale sono scritti i pesi finali, se non è stato specificato alcun campo peso di input (se è specificato un campo peso di input, i suoi valori vengono sostituiti dai pesi finali come indicato sopra e non viene creato alcun campo peso di output separato). I valori del peso di output indicano il numero di record rappresentato da ogni record campionato nei dati originali. La somma dei valori dei pesi fornisce una stima delle dimensioni del campione. Per esempio, se viene considerato un campione casuale del 10%, il peso di output sarà 10 per tutti i record, a indicare che ogni record campionato rappresenta approssimativamente dieci record nei dati originali. In un campione stratificato o ponderato, i valori del peso di output possono variare in funzione della proporzione del campione di ogni strato.

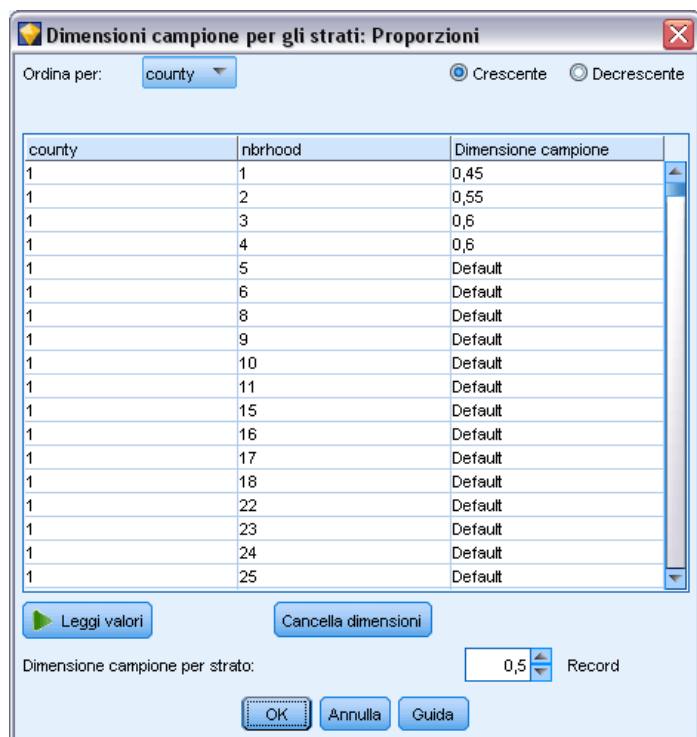
Commenti

- Il campionamento raggruppato è utile se non è possibile ottenere un elenco completo della popolazione che si desidera campionare ma si possono ottenere elenchi completi di determinati gruppi o cluster. Viene usato anche nei casi in cui un campione casuale produrrebbe un elenco di soggetti di test che risulterebbe impossibile contattare. Per esempio, risulterebbe più semplice visitare tutti gli imprenditori agricoli di una regione anziché un elenco selezionato di imprenditori agricoli sparsi in ogni regione del paese.
- È possibile specificare sia campi cluster che di stratificazione in modo da campionare i cluster in modo indipendente all'interno di ogni strato. Per esempio, è possibile campionare i valori delle proprietà stratificati per regione e raggruppati in base alle province di ogni regione. Ciò garantisce la formulazione di un campione indipendente di province da ogni regione. Alcune province verranno incluse e altre no, ma per ogni provincia inclusa verranno incluse tutte le proprietà all'interno della provincia stessa.
- Per selezionare un campione casuale di unità da ogni cluster, è possibile unire due nodi Campione. Per esempio, è possibile prima campionare le province stratificate per regione come descritto sopra, e quindi collegare un secondo nodo Campione e selezionare *provincia* come campo di stratificazione, per campionare una proporzione di record da ogni provincia.
- Nei casi in cui sia necessaria una combinazione di campi per identificare i cluster in modo univoco, è possibile generare un nuovo campo utilizzando un nodo Nuovo campo. Per esempio, se più negozi utilizzano lo stesso sistema di numerazione delle transazioni, è possibile creare un nuovo campo che concatena gli ID di negozi e transazioni.

Dimensioni dei campioni per gli strati

Durante la formulazione di un campione stratificato, l'opzione di default consente di campionare la stessa proporzione di record o cluster di ogni strato. Se un gruppo supera un altro di un fattore di 3, per esempio, in genere si desidera mantenere lo stesso rapporto nel campione. In caso contrario, tuttavia, è possibile specificare la dimensione del campione in modo separato per ogni strato.

Figura 3-6
Definizione delle dimensioni dei campioni per gli strati



La finestra di dialogo Dimensioni campione per gli strati elenca i singoli valori del campo di stratificazione, consentendo di escludere i valori di default per quel dato strato. Se sono selezionati più campi di stratificazione, viene elencata ogni possibile combinazione di valori, consentendo di specificare, per esempio, la dimensione di ogni gruppo etnico all'interno di ogni provincia, oppure di ogni provincia di ogni regione. Le dimensioni vengono specificate come proporzioni o conteggi, in base all'impostazione corrente nel nodo Campione.

Per specificare le dimensioni dei campioni per gli strati

- ▶ Nel nodo Campione, selezionare Complesso e selezionare uno o più campi di stratificazione. [Per ulteriori informazioni, vedere l'argomento Impostazioni di raggruppamento e stratificazione a pag. 68.](#)
- ▶ Selezionare Personalizzato e quindi Specifica dimensioni.
- ▶ Nella finestra di dialogo Dimensioni campione per gli strati, selezionare il pulsante Leggi valori in basso a sinistra per popolare la visualizzazione. Se necessario, è possibile istanziare i valori in un nodo Tipo o di input a monte. [Per ulteriori informazioni, vedere l'argomento Che cos'è l'istanziamento? in il capitolo 4 a pag. 130.](#)
- ▶ Fare clic su una riga qualsiasi per ignorare la dimensione di default dello strato.

Note sulle dimensioni dei campioni

Le dimensioni personalizzate dei campioni possono essere utili se strati diversi hanno varianze diverse, per esempio per rendere le dimensioni dei campioni proporzionali alla deviazione standard (se i casi all'interno dello strato sono più vari, è necessario campionarne un numero maggiore per ottenere un campione significativo). Se invece uno strato è di piccole dimensioni, può essere utile utilizzare una proporzione di campionamento superiore per assicurarsi che venga incluso un numero minimo di osservazioni.

Nota: se si effettua la stratificazione in base a un campo con valori mancanti (valori vuoti o mancanti di sistema, stringhe vuote, spazi bianchi e valori vuoti o mancanti definiti dall'utente), non è possibile specificare dimensioni di campionamento personalizzate per gli strati. Se si desidera utilizzare dimensioni di campionamento personalizzate durante la stratificazione in base a un campo con valori mancanti o vuoti, è necessario compilarli a monte. [Per ulteriori informazioni, vedere l'argomento Gestione dei campi con valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Nodo Bilanciamento

È possibile utilizzare i nodi Bilanciamento per correggere sbilanciamenti negli insiemi di dati affinché soddisfino i criteri di test specificati. Per esempio, si supponga che un insieme di dati contenga solo due valori, *basso* o *alto*, e che per il 90% dei casi il valore sia *basso* e solo per il 10% sia *alto*. Molte tecniche di modellazione utilizzeranno dati distorti di questo tipo in modo non corretto, infatti tenderanno ad apprendere solo i risultati con valore *basso* e ignorare i risultati con valore *alto*, dal momento che risulta più raro. Se i dati sono bilanciati correttamente con un numero quasi uguale di risultati con valore *basso* e *alto*, sarà più probabile che i modelli riescano a individuare schemi distintivi per i due gruppi. In questo caso, è possibile utilizzare un nodo Bilanciamento per creare una direttiva di bilanciamento in grado di ridurre i casi contenenti un risultato con valore *basso*.

Il bilanciamento viene eseguito tramite la duplicazione e il successivo scarto di record in base alle condizioni specificate. I record che non soddisfano alcuna condizione vengono passati nello stream. Dal momento che questo processo funziona tramite duplicazione e/o scarto di record, la sequenza originale dei dati viene persa nelle operazioni a valle. Pertanto, prima di aggiungere un nodo Bilanciamento allo stream di dati è necessario assicurarsi che ogni valore correlato alla sequenza venga derivato.

Nota: i nodi Bilanciamento possono essere generati automaticamente da istogrammi e grafici di distribuzione. Per esempio, è possibile bilanciare i dati per mostrare proporzioni uguali in tutte le categorie di un campo categoriale, come evidenziato in un plot di distribuzione.

Esempio. Durante la creazione di uno stream RFM per individuare i clienti recenti che hanno risposto in modo favorevole alle precedenti campagne di marketing, la divisione marketing di un'azienda utilizza un nodo Bilanciamento per bilanciare le differenze tra le risposte vere e false nei dati.

Impostazione delle opzioni per il nodo Bilanciamento

Figura 3-7
Impostazioni del nodo Bilanciamento



Direttive di bilanciamento record. Elenca le direttive di bilanciamento correnti. Ogni direttiva include un fattore e una condizione con cui è possibile indicare al software di “aumentare la proporzione di record in base a un fattore specificato se la condizione è vera”. Un fattore inferiore a 1,0 indica che la proporzione di record specificati verrà diminuita. Per esempio, se si desidera diminuire il numero di record in cui come trattamento viene utilizzata la cura denominata curaY, è possibile creare una direttiva di bilanciamento con un fattore di 0,7 e una condizione Cura = "curaY". Questa direttiva indica che il numero di record in cui il trattamento utilizzato è curaY verrà ridotto del 70% per tutte le operazioni a valle.

Nota: per la riduzione è possibile specificare fattori di bilanciamento contenenti fino a quattro decimali. I fattori impostati su valori inferiori a 0,0001 determineranno un errore, poiché i risultati non vengono calcolati correttamente.

- **Creare le condizioni** facendo clic sul pulsante a destra del campo di testo. In questo modo verrà aggiunta una riga vuota in cui è possibile specificare le nuove condizioni. Per creare un'espressione CLEM per la condizione, fare clic sul pulsante del Generatore di espressioni.
- **Eliminare le direttive** utilizzando il pulsante con la X rossa.
- **Ordinare le direttive** utilizzando i pulsanti con le frecce.

Bilancia solo dati di addestramento. Se nello stream è presente un campo di partizione, questa opzione bilancia solo i dati nella partizione di addestramento. Questo può essere utile, in particolare, se si generano punteggi di propensione regolata che richiedono una partizione di test o di validazione non bilanciata. [Per ulteriori informazioni, vedere l'argomento Punteggi di propensione in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#) Se nello stream non è presente alcun campo di partizione (oppure se sono specificati più campi di partizione), questa opzione viene ignorata e tutti i dati sono bilanciati.

Nodo Aggregazione

L'aggregazione costituisce un'attività di preparazione dei dati utilizzata spesso per ridurre le dimensioni di un insieme di dati. Prima di procedere all'aggregazione è opportuno eseguire la pulitura dei dati, prestando particolare attenzione ai valori mancanti. Con l'aggregazione potrebbero andare perse informazioni potenzialmente utili sui valori mancanti. [Per ulteriori informazioni, vedere l'argomento Panoramica sui valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

È possibile utilizzare un nodo Aggregazione per sostituire una sequenza di record di input con record di output aggregati di riepilogo. Per esempio, è possibile utilizzare un insieme di record di input come:

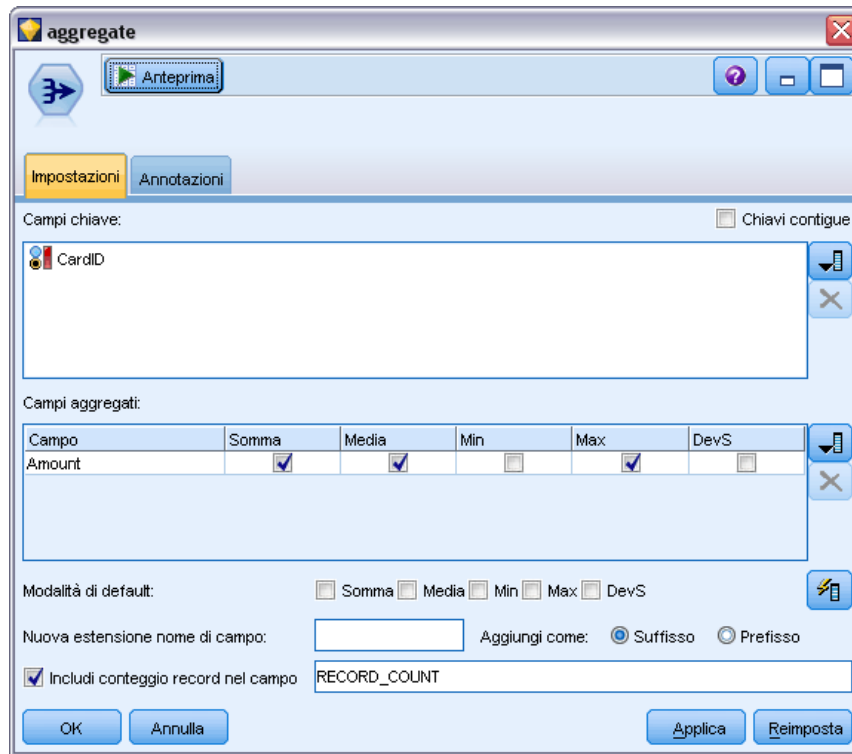
Età	Sesso	Regione	Branch	Vendite
23	M	C	8	4
45	M	C	16	4
37	M	C	8	5
30	M	C	5	7
44	M	N	4	9
25	M	N	2	11
29	F	C	16	6
41	F	N	4	8
23	F	N	6	2
45	F	N	4	5
33	F	N	6	10

È possibile aggregare questi record utilizzando *Sesso* e *Area* come campi chiave e quindi scegliendo di aggregare *Età* con la modalità Media e *Vendite* con la modalità Somma. Selezionare *Includi conteggio record* nel campo nella finestra di dialogo del nodo Aggregazione. L'output aggregato sarà il seguente:

Età	Sesso	Regione	Vendite	RECORD_COUNT
35.5	F	N	25	4
34.5	M	N	20	2
29	F	C	6	1
33.75	M	C	20	4

Nota: i campi per i quali non viene specificata la modalità di aggregazione, per esempio *Ramo*, verranno automaticamente scartati.

Figura 3-8
Finestra di dialogo del nodo Aggregazione



Impostazione delle opzioni per il nodo Aggregazione

Campi chiave. Elenca i campi che possono essere utilizzati come chiavi per l'aggregazione. Come campi chiave possono essere utilizzati sia campi numerici che campi simbolici. Se si selezionano più campi chiave, i valori verranno combinati in modo da ottenere un valore chiave per l'aggregazione dei record. Per ogni campo chiave univoco verrà generato un solo record aggregato. Per esempio, se *Sesso* e *Area* sono i campi chiave disponibili, verrà generato un record aggregato per ogni combinazione univoca di *M* e *F* con le aree *N* e *O* (ovvero quattro combinazioni univoche). Per aggiungere un campo chiave, utilizzare il pulsante di selezione dei campi a destra della finestra.

Chiavi contigue. Selezionare questa opzione se si sa che tutti i record con gli stessi valori chiave sono raggruppati insieme nell'input (per esempio, se l'input è ordinato in base ai campi chiave). In questo modo si migliorano le prestazioni.

Campi aggregati. Elenca sia i campi numerici i cui valori verranno aggregati, sia le modalità di aggregazione selezionate. Per aggiungere campi a questo elenco, utilizzare il pulsante di selezione dei campi a destra.

Modalità di default. Specificare la modalità di aggregazione di default da utilizzare per i campi aggiunti. Se si utilizza spesso la stessa aggregazione, selezionare una o più modalità e utilizzare il pulsante a destra di questo gruppo di opzioni per applicare le modalità selezionate a tutti i campi sopra elencati. Sono disponibili le seguenti modalità di aggregazione:

- **Somma.** Selezionare questa opzione per restituire i valori sommati per ogni combinazione di campi chiave.
- **Media.** Selezionare questa opzione per restituire i valori medi per ogni combinazione di campi chiave.
- **Min.** Selezionare questa opzione per restituire i valori minimi per ogni combinazione di campi chiave.
- **Max.** Selezionare questa opzione per restituire i valori massimi per ogni combinazione di campi chiave.
- **DevS.** Selezionare questa opzione per restituire la deviazione standard per ogni combinazione di campi chiave.

Nuova estensione nome di campo. Selezionare questa opzione per aggiungere un suffisso o un prefisso, per esempio “1” o “nuovo”, per duplicare i campi aggregati. Per esempio, un’aggregazione di valori minimi nel campo *Età* per cui è stata selezionata l’opzione di suffisso ed è stato specificato “1” come estensione restituirà il nome di campo *Età_Min_1*. *Nota:* le estensioni di aggregazione quali *_Min* o *_Max* vengono automaticamente aggiunte al nuovo campo, come indicazione del tipo di aggregazione eseguita. Selezionare Suffisso o Prefisso per indicare il tipo di estensione desiderato.

Includi conteggio record nel campo. Selezionare questa opzione per includere un ulteriore campo in ogni record di risultato. A questo campo verrà assegnato il nome di default *Conteggio_record*. Tale campo indica il numero dei record di input che sono stati aggregati per formare ogni record aggregato. Per assegnare un nome personalizzato a questo campo, digitare il nome desiderato nel campo di modifica.

Nota: i valori di sistema nulli verranno esclusi nel calcolo degli aggregati, ma inclusi nel conteggio dei record. I valori vuoti, invece, verranno inclusi sia nel calcolo di aggregazione sia nel conteggio dei record. Per escludere i valori vuoti è possibile utilizzare un nodo Riempimento in modo da sostituire i valori vuoti con valori nulli. È inoltre possibile rimuovere i valori vuoti utilizzando un nodo Seleziona.

Performance

Le operazioni di aggregazione possono beneficiare dell’attivazione dell’elaborazione parallela. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di ottimizzazione in il capitolo 12 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Nodo Aggregazione RFM

Il nodo Aggregazione RFM (Recency, Frequency, Monetary, Passato recente, Frequenza, Monetario) consente di prendere in considerazione i dati storici delle transazioni dei clienti, di eliminare i dati non utilizzati e di combinare tutti i dati delle transazioni rimanenti in un’unica riga (utilizzando come chiave l’ID cliente univoco) che elenca la data dell’ultima trattativa

(passato recente), il numero di transazioni effettuate (frequenza) e il valore totale delle transazioni (monetario).

Prima di procedere a qualsiasi aggregazione è opportuno eseguire la pulitura dei dati, prestando particolare attenzione ai valori mancanti. [Per ulteriori informazioni, vedere l'argomento Panoramica sui valori mancanti in il capitolo 6 in *Manuale dell'utente di IBM SPSS Modeler 14.2*.](#)

Una volta identificati e trasformati i dati utilizzando il nodo Aggregazione RFM, è possibile utilizzare un nodo Analisi RFM per eseguire ulteriori analisi. [Per ulteriori informazioni, vedere l'argomento Nodo Analisi RFM in il capitolo 4 a pag. 189.](#)

Si noti che, dopo essere stato elaborato con il nodo Aggregazione RFM, il file di dati non disporrà di alcun valore obiettivo; pertanto, prima di poterlo utilizzare come input per ulteriori analisi predittive con altri nodi Modelli quali C5.0 o CHAID, sarà necessario unirlo ad altri dati di clienti (per esempio, tramite la corrispondenza degli ID cliente). [Per ulteriori informazioni, vedere l'argomento Nodo Unione a pag. 80.](#)

I nodi Aggregazione RFM e Analisi RFM di IBM® SPSS® Modeler sono impostati per utilizzare la discretizzazione indipendente, ovvero classificano e discretizzano i dati rispetto a ogni misura di passato recente, frequenza e valore monetario, indipendentemente dai loro valori o dalle altre due misure.

Impostazione delle opzioni per il nodo Aggregazione RFM

Figura 3-9
Impostazioni di aggregazione RFM

2007-06-06

Anteprima

RFM

Impostazioni Annotazioni

Calcola passato recente relativo a: Data fissa 2007-06-06 Data odierna

ID contigui

ID: CardID

Data: Date

Valore: Amount

Nuova estensione nome di campo: Aggiungi come: Suffisso Prefisso

Scarta record con valore inferiore a: 1,0

Includi solo transazioni recenti:

Data transazione successiva a: 2007-06-06

Transazione entro gli ultimi: 6 Mesi

Salva data della seconda transazione più recente

Salva data della terza transazione più recente

OK Annulla Applica Reimposta

Calcola passato recente relativo a. Specifica la data a partire dalla quale verrà calcolato il passato recente delle transazioni. Può essere una Data fissa inserita dall'utente oppure la Data odierna impostata dal sistema. La Data odierna è inserita per default e viene aggiornata automaticamente quando viene eseguito il nodo.

ID contigui. Se i dati sono preordinati in modo che tutti i record con lo stesso ID appaiano insieme nello stream di dati, selezionare questa opzione per accelerare l'elaborazione. Se i dati non sono preordinati (o non si è certi che lo siano), lasciare questa opzione deselezionata e il nodo ordinerà i dati automaticamente.

ID. Selezionare il campo da utilizzare per identificare i clienti e le loro transazioni. Per visualizzare i campi tra i quali è possibile effettuare la selezione, utilizzare il pulsante di selezione dei campi a destra.

Data. Selezionare il campo data da utilizzare per calcolare il passato recente. Per visualizzare i campi tra i quali è possibile effettuare la selezione, utilizzare il pulsante di selezione dei campi a destra.

Si noti che è richiesto un campo contenente un'archiviazione della data o del timestamp nel formato appropriato da utilizzare come input. Per esempio, se si dispone di un campo stringa con valori come *Gen 2007*, *Feb 2007* e così via, è possibile convertirlo in campo data mediante un nodo Riempimento e la funzione `to_date()`. [Per ulteriori informazioni, vedere l'argomento Conversione della classe di archiviazione mediante il nodo Riempimento in il capitolo 4 a pag. 168.](#)

Valore. Selezionare il campo da utilizzare per calcolare il valore monetario totale delle transazioni del cliente. Per visualizzare i campi tra i quali è possibile effettuare la selezione, utilizzare il pulsante di selezione dei campi a destra. *Nota:* questo valore deve essere numerico.

Nuova estensione nome di campo. Selezionare questa opzione per aggiungere un suffisso o un prefisso, per esempio "12_mesi", ai campi passato recente, frequenza e monetario appena generati. Selezionare Suffisso o Prefisso per indicare il tipo di estensione desiderato. Questo può essere utile, per esempio, se si esaminano più periodi di tempo.

Scarta record con valore inferiore a. Se necessario, è possibile specificare un valore minimo al di sotto del quale i dettagli della transazione non vengono utilizzati per il calcolo dei totali RFM. Le unità di valore si correlano al campo Valore selezionato.

Includi solo transazioni recenti. Se si sta analizzando un database di grandi dimensioni, è possibile specificare che vengano utilizzati solo i record più recenti. È possibile scegliere di utilizzare i dati registrati dopo una certa data o in un periodo di tempo recente:

- **Data transazione successiva a.** Specificare la data della transazione dopo la quale i record verranno inclusi nell'analisi.
- **Transazione entro gli ultimi.** Specificare il numero e il tipo di periodi (giorni, settimane, mesi o anni) prima della data Calcola passato recente relativo a dopo la quale i record verranno inclusi nell'analisi.

Salva data della seconda transazione più recente. Selezionare questa casella se si desidera conoscere la data della seconda transazione più recente per ogni cliente. Inoltre, è possibile selezionare anche la casella Salva data della terza transazione più recente. Per esempio, questo può essere utile per identificare i clienti che potrebbero aver effettuato più transazioni parecchio tempo fa ma solo una in tempi più recenti.

Nodo Ordina

È possibile utilizzare i nodi Ordina per ordinare i record in ordine crescente o decrescente in base ai valori di uno o più campi. Per esempio, i nodi Ordina vengono spesso utilizzati per visualizzare e selezionare i record con i valori di dati più comuni. In genere, si aggregano innanzitutto i dati mediante il nodo Aggregazione e quindi si utilizza il nodo Ordina per ordinare i dati aggregati secondo un ordine decrescente dei conteggi di record. La visualizzazione di questi risultati in una tabella consentirà di analizzare i dati e di prendere decisioni, per esempio selezionando i record relativi ai 10 migliori clienti.

Figura 3-10
Finestra di dialogo del nodo Ordina



Ordina per. Vengono visualizzati in una tabella tutti i campi selezionati per essere utilizzati come chiavi di ordinamento. I campi numerici assicurano le migliori prestazioni di ordinamento.

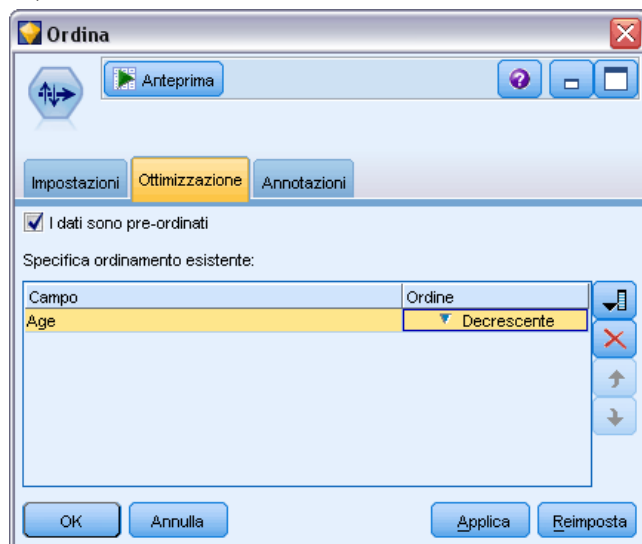
- **Aggiungere campi** all’elenco utilizzando il pulsante di selezione dei campi a destra.
- **Selezionare un ordine** facendo clic sulla freccia Crescente o Decrescente nella colonna *Ordine* della tabella.
- **Eliminare i campi** utilizzando il pulsante con la X rossa.
- **Ordinare le direttive** utilizzando i pulsanti con le frecce.

Ordinamento di default. Selezionare Crescente o Decrescente per utilizzare il relativo ordinamento come impostazione di default per i nuovi campi aggiunti nella tabella sopra.

Impostazioni di ottimizzazione dell’ordinamento

Se si sta lavorando con dei dati già ordinati in base ad alcuni campi chiave, è possibile specificare quali campi sono già ordinati, consentendo al sistema di ordinare il resto dei dati in modo più efficiente. Questo può accadere, per esempio, se si desidera effettuare l’ordinamento per *Età* (decrescente) e *Cura* (crescente) ma si sa che i dati sono già ordinati per *Età* (decrescente).

Figura 3-11
Impostazioni di ottimizzazione



I dati sono preordinati. Specifica se i dati sono già ordinati in base a uno o più campi.

Specifica ordinamento esistente. Specificare i campi già ordinati. Utilizzando la finestra di dialogo Seleziona campi, aggiungere i campi all'elenco. Nella colonna *Ordine*, specificare se ciascun campo è ordinato in ordine crescente o decrescente. Se si specificano più campi, assicurarsi di elencarli secondo l'ordinamento corretto. Utilizzare le frecce a destra dell'elenco per disporre i campi nell'ordine corretto. In caso di errore nell'indicazione dell'ordinamento esistente corretto, al momento dell'esecuzione dello stream verrà visualizzato un errore che indica il numero dei record nei quali l'ordinamento è incoerente rispetto all'impostazione specificata.

Nota: la velocità di ordinamento può beneficiare dell'attivazione dell'elaborazione parallela. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di ottimizzazione in il capitolo 12 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Nodo Unione

Un nodo Unione consente di utilizzare più record di input per creare un singolo record di output contenente alcuni o tutti i campi di input. Questa operazione risulta utile se si desidera unire dati da sorgenti diverse, per esempio dati interni sui clienti e dati demografici acquisiti dall'esterno. Esistono due modi per unire i dati:

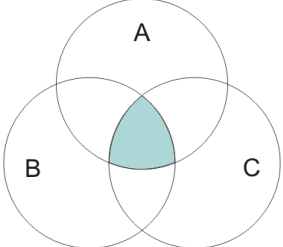
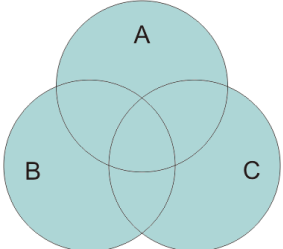
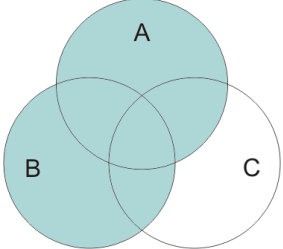
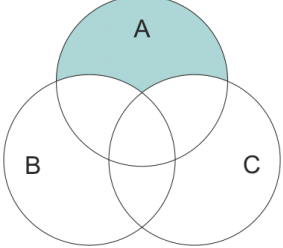
- **Unione in base all'ordine.** Questo metodo consente di concatenare record corrispondenti di tutte le sorgenti in base all'ordine di input, fino a esaurire la più piccola sorgente dati. Per utilizzare questa opzione è importante che i dati siano stati ordinati tramite un nodo Ordina.
- **Unione in base a un campo chiave.** Questo metodo consente di utilizzare un campo chiave, per esempio *ID cliente*, per specificare le modalità di corrispondenza tra record di sorgenti diverse. Sono supportati diversi tipi di join, quali join interno, join esterno completo, join esterno parziale e anti-join. [Per ulteriori informazioni, vedere l'argomento Tipi di join a pag. 81.](#)

Tipi di join

Quando si utilizza un campo chiave per l'unione di dati, è utile considerare attentamente i record da escludere e quelli da includere. È disponibile una vasta gamma di join, illustrati di seguito in modo dettagliato.

I due tipi di join di base vengono definiti join interni e join esterni. Questi metodi vengono spesso utilizzati per unire le tabelle di insiemi di dati correlati in base ai valori comuni di un campo chiave, quale *ID cliente*. I join interni consentono di ottenere un'unione corretta, con un insieme di dati di output comprendente solo record completi. Oltre a comprendere i record completi dei dati uniti, i join esterni consentono di includere dati univoci da una o più tabelle di input.

Di seguito viene fornita una descrizione più dettagliata dei tipi di join supportati.

	<p>Un join interno comprende solo i record in cui è presente un valore del campo chiave comune a tutte le tabelle di input. I record non corrispondenti non verranno quindi inclusi nell'insieme di dati di output.</p>
	<p>Un join esterno completo comprende tutti i record, sia corrispondenti che non corrispondenti, delle tabelle di input. Il join esterno sinistro e il join esterno destro, descritti di seguito, vengono definiti join esterni parziali.</p>
	<p>Un join esterno parziale comprende tutti i record corrispondenti in base al campo chiave e i record non corrispondenti delle tabelle specificate. In altre parole, tutti i record di alcune tabelle e solo i record corrispondenti di altre. È possibile selezionare le tabelle da includere nel join esterno (per esempio, A e B della figura) utilizzando il pulsante Seleziona nella scheda Unione. Quando vengono unite soltanto due tabelle, i join parziali vengono denominati anche join esterno sinistro o join esterno destro. Poiché IBM® SPSS® Modeler supporta l'unione di più di due tabelle, in questo contesto viene utilizzata la definizione di join parziale esterno.</p>
	<p>Un anti-join comprende soltanto i record non corrispondenti della prima tabella di input (tabella A della figura). Questo tipo di join rappresenta l'opposto di un join interno e non determina l'inclusione di record completi nell'insieme di dati di output.</p>

Per esempio, se si hanno informazioni sulle aziende agricole in un insieme di dati e le domande di assicurazione relative alle aziende agricole in un altro, è possibile abbinare i record della prima origine alla seconda origine utilizzando le opzioni di unione.

Per determinare se un cliente dell'azienda agricola campione ha presentato una domanda di assicurazione, utilizzare l'opzione join interno per restituire un elenco che mostra i punti di corrispondenza di tutti gli ID dei due campioni.

Figura 3-12
Esempio di output per un'unione join interno

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0...
2	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0...
3	id620	name620	north	880.000	74.000	6.000	426988.000	rapeseed	arable_de	118885.0...

L'utilizzo dell'opzione join esterno completo restituisce tutti i record corrispondenti e non corrispondenti delle tabelle di input. Per i valori incompleti verrà utilizzato il valore mancante di sistema (\$null\$).

Figura 3-13
Esempio di output per un'unione join esterno completo

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimtype	claimvalue
1	id601	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	decomm...	74703.1C
2	id602	name602	north	1780.000	42.000	9.000	734118.000	maize	\$null\$	\$nul
3	id604	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes	decomm...	281082.0
4	id605	name605	north	1700.000	46.000	8.000	621148.000	wheat	decomm...	122006.0
5	id606	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	\$null\$	arable_de	122135.0

Un join esterno parziale comprende tutti i record corrispondenti in base al campo chiave e i record non corrispondenti delle tabelle specificate. La tabella visualizza tutti i record corrispondenti del campo ID e i record corrispondenti del primo insieme di dati.

Figura 3-14
Esempio di output per un'unione join esterno parziale

	id	claimtype	claimvalue	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	\$null\$	\$null\$	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id604	decomm...	281082.0...	name604	southwest	1860.000	103.0...	3.000	625251.000	potatoes
3	id605	decomm...	122006.0...	name605	north	1700.000	46.000	8.000	621148.000	wheat
4	id607	\$null\$	\$null\$	name607	southeast	1820.000	29.000	6.000	211605.000	maize
5	id608	\$null\$	\$null\$	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
6	id609	\$null\$	\$null\$	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
7	id615	\$null\$	\$null\$	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
8	id618	\$null\$	\$null\$	name618	southeast	1180.000	98.000	3.000	368646.000	maize

Se viene utilizzata l'opzione anti-join, la tabella restituisce soltanto i record non corrispondenti della prima tabella di input.

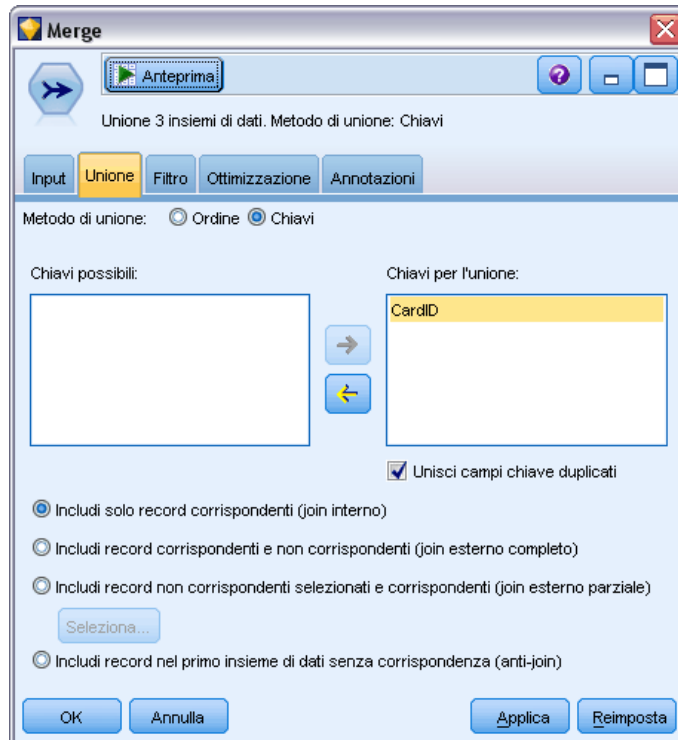
Figura 3-15
Esempio di output per un'unione anti-join

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop
1	id602	name602	north	1780.000	42.000	9.000	734118.000	maize
2	id607	name607	southeast	1820.000	29.000	6.000	211605.000	maize
3	id608	name608	southeast	1640.000	108.0...	7.000	1167040.0...	maize
4	id609	name609	southwest	1600.000	101.0...	5.000	756755.000	wheat
5	id615	name615	midlands	920.000	86.000	6.000	442554.000	potatoes
6	id618	name618	southeast	1180.000	98.000	3.000	368646.000	maize
7	id619	name619	north	840.000	64.000	8.000	457552.000	potatoes

Definizione di chiavi e metodo di unione

Figura 3-16

Utilizzo della scheda Unione per l'impostazione delle opzioni relative al metodo di unione



Metodo di unione. Selezionare Ordine o Chiavi per specificare il metodo di unione dei record. La selezione di Chiavi consente di attivare la parte inferiore della finestra di dialogo.

- **Ordine.** Unisce i record in base all'ordine in modo che il record n di ogni input venga unito per generare il record n di output. Quando sono stati generati tutti i record corrispondenti a un record di input, non vengono più creati record di output. Questo significa che il numero di record creati corrisponde al numero di record dell'insieme di dati più piccolo.
- **Chiavi.** Utilizza un campo chiave, per esempio *ID transazione*, per unire record con lo stesso valore del campo chiave. Questa opzione è equivalente a un "join di uguaglianza" (equi-join) di database. Se un valore chiave ricorre più volte, verranno restituite tutte le possibili combinazioni. Per esempio, se record con lo stesso valore di campo chiave A contengono valori B , C e D diversi in altri campi, i campi uniti genereranno un record separato per ogni combinazione di A con il valore B , A con il valore C e A con il valore D .

Nota: nel metodo di unione in base alla chiave, i valori nulli non vengono considerati identici e non verranno inclusi nel join.

Chiavi possibili. Elenca solo i campi con nomi aventi corrispondenza esatta in tutte le sorgenti dati di input. Selezionare un campo dall'elenco e utilizzare il pulsante con la freccia per aggiungerlo come campo chiave da utilizzare per l'unione dei record. È possibile utilizzare più campi chiave. I campi di input non corrispondenti si possono rinominare tramite un nodo Filtro o dalla scheda Filtro di un nodo di input.

Chiavi per l'unione. Elenca tutti i campi utilizzati per unire i record di tutte le sorgenti dati di input in base ai valori dei campi chiave. Per rimuovere una chiave dall'elenco, selezionarla e utilizzare il pulsante con la freccia per includerla nuovamente nell'elenco Chiavi possibili. L'opzione seguente è attivata solo quando vengono selezionati più campi.

Unisci campi chiave duplicati. Quando vengono selezionati più campi chiave, questa opzione evita che vengano creati più campi di output con lo stesso nome. Tale opzione è attivata per default, a meno che gli stream siano stati importati da versioni precedenti di IBM® SPSS® Modeler. Quando questa opzione è disattivata, i campi chiave duplicati devono essere rinominati o esclusi utilizzando la scheda Filtro nella finestra di dialogo del nodo Unione.

Includi solo record corrispondenti (join interno). Selezionare questa opzione per unire solo record completi.

Includi record corrispondenti e non corrispondenti (join esterno completo). Selezionare questa opzione per eseguire un "join esterno completo". In questo caso, se i valori del campo chiave non sono presenti in tutte le tabelle di input, i record incompleti vengono comunque mantenuti. Il valore non definito (\$null\$) viene aggiunto al campo chiave e incluso nel record di output.

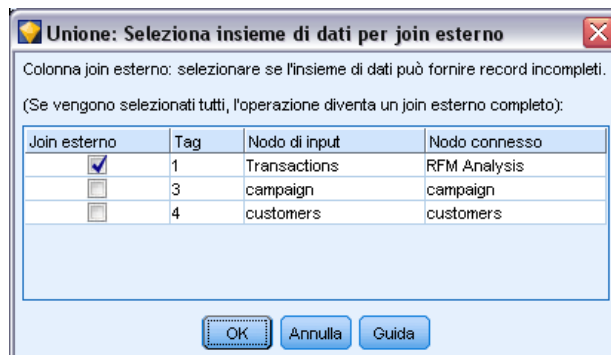
Includi record non corrispondenti selezionati e corrispondenti (join esterno parziale). Selezionare questa opzione per eseguire un "join esterno parziale" delle tabelle selezionate in una finestra di dialogo secondaria. Fare clic su Seleziona per specificare le tabelle i cui record incompleti verranno mantenuti nell'unione.

Includi record nel primo insieme di dati senza corrispondenza (anti-join). Selezionare questa opzione per eseguire un tipo di "anti-join" in cui vengono passati a valle soltanto i record non corrispondenti del primo insieme di dati. È possibile specificare l'ordine degli insiemi di dati di input utilizzando le frecce disponibili nella scheda Input. Con questo tipo di join, nell'insieme di dati di output non vengono inclusi record completi. [Per ulteriori informazioni, vedere l'argomento Tipi di join a pag. 81.](#)

Selezione di dati per join parziali

Per eseguire un join esterno parziale, è necessario selezionare una o più tabelle di cui verranno mantenuti i record incompleti. Può per esempio rivelarsi utile mantenere tutti i record di una tabella dei clienti, ma soltanto i record corrispondenti della tabella dei mutui ipotecari.

Figura 3-17
Selezione di dati per un join esterno parziale

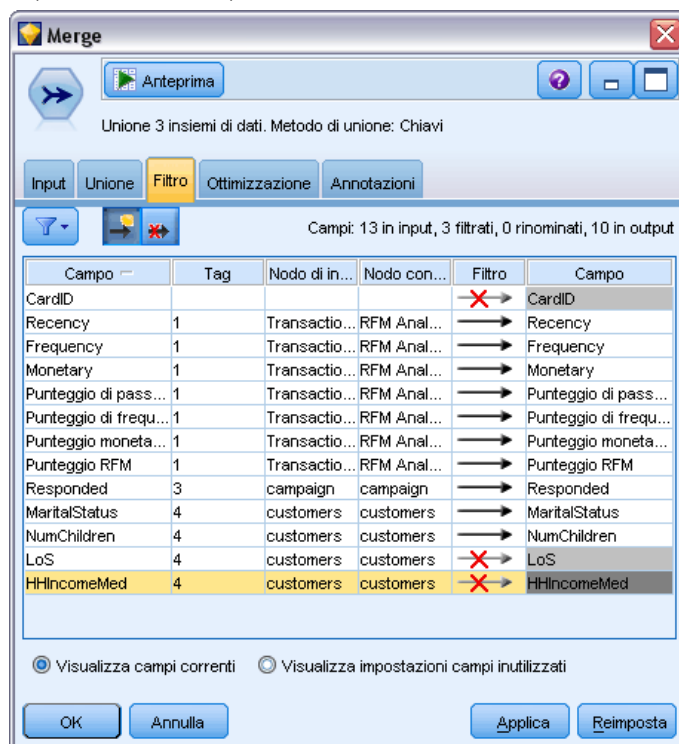


Colonna Join esterno. Nella colonna *Join esterno*, selezionare gli insiemi di dati da includere interamente. Nel caso di un join parziale, per gli insiemi di dati selezionati in questo contesto verranno mantenuti sia i record sovrapposti che i record incompleti. [Per ulteriori informazioni, vedere l'argomento Tipi di join a pag. 81.](#)

Filtro di campi dal nodo Unione

I nodi Unione includono un comodo metodo per filtrare o rinominare i campi duplicati risultanti dall'unione di più sorgenti dati. Fare clic sulla scheda Filtro nella finestra di dialogo per selezionare le opzioni di filtro.

Figura 3-18
Impostazione delle opzioni di filtro dal nodo Unione



Le opzioni illustrate di seguito sono quasi identiche a quelle disponibili per il nodo Filtro. Nel menu Filtro sono tuttavia disponibili opzioni aggiuntive che non verranno illustrate di seguito. [Per ulteriori informazioni, vedere l'argomento Applicazione di filtri o ridenominazione di campi in il capitolo 4 a pag. 143.](#)

Campo. Visualizza i campi di input dalle sorgenti dati connesse.

Tag. Elenca il nome di tag (o il numero) associato al collegamento della sorgente dati. Fare clic sulla scheda Input per modificare i collegamenti attivi per questo nodo Unione.

Nodo di input. Visualizza il nodo di input i cui dati verranno uniti.

Nodo connesso. Visualizza il nome del nodo connesso al nodo Unione. Spesso il data mining complesso richiede diverse operazioni di unione o accodamento che possono includere lo stesso nodo di input. Il nome del nodo connesso consente di operare una distinzione.

Filtro. Visualizza le connessioni correnti tra campi di input e di output. Le connessioni attive sono contrassegnate da una freccia continua. Le connessioni contrassegnate da una X rossa indicano i campi filtrati.

Campo. Elenca i campi di output dopo l'unione o l'accodamento. I campi duplicati vengono visualizzati in rosso. Fare clic sul campo Filtro per disattivare i campi duplicati.

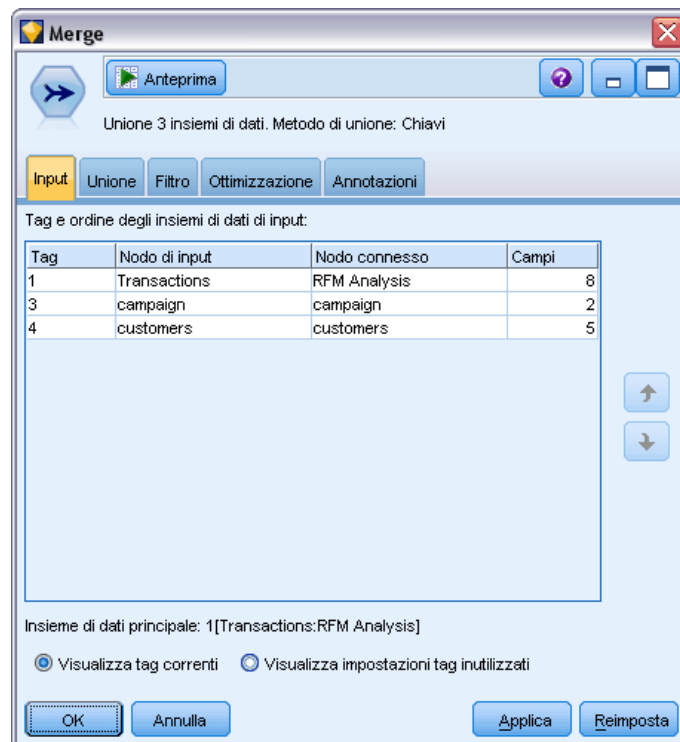
Visualizza campi correnti. Selezionare questa opzione per visualizzare informazioni sui campi selezionati da utilizzare come campi chiave.

Visualizza impostazioni campi inutilizzati. Selezionare questa opzione per visualizzare informazioni su campi attualmente non utilizzati.

Impostazione di tag e ordine dei dati di input

Mediante la scheda Input delle finestre di dialogo dei nodi Unione e Accodamento è possibile specificare l'ordine delle sorgenti dati di input e apportare qualsiasi modifica al nome di tag per ogni sorgente.

Figura 3-19
Utilizzo della scheda Input per specificare tag e ordine di input



Tag e ordine degli insiemi di dati di input. Selezionare questa opzione per unire o accodare solo record completi.

- **Tag.** Elenca i nomi di tag correnti per ogni sorgente dati di input. I nomi di tag o **tag** rappresentano un metodo per identificare in modo univoco i collegamenti dei dati per le operazioni di unione o accodamento. Per comprendere il funzionamento del flusso di dati in Clementine, si può immaginare una serie di tubi che trasportano acqua e che a un certo punto si uniscono in una sola condotta. Analogamente, in IBM® SPSS® Modeler i dati seguono percorsi separati e vengono uniti in un punto preciso, che costituisce spesso un'interazione complessa tra le diverse sorgenti dati. I tag rappresentano un mezzo per la gestione degli input (paragonabili ai tubi dell'acqua) in un nodo Unione o Accodamento, in modo che, se il nodo viene salvato o disconnesso, i collegamenti vengano mantenuti e siano facilmente identificabili.

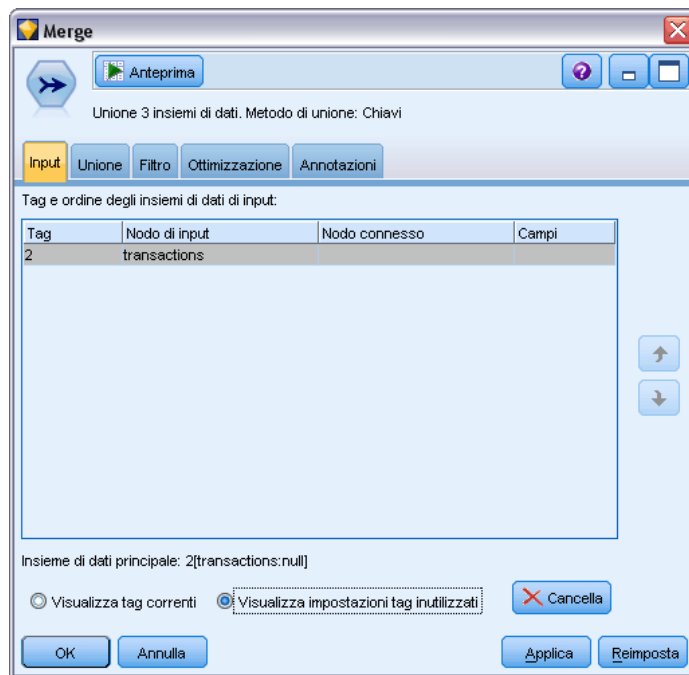
Quando si connettono sorgenti dati aggiuntive a un nodo Unione o Accodamento, vengono creati automaticamente tag di default che utilizzano numeri per rappresentare l'ordine di connessione dei nodi. Tale ordine non è correlato all'ordine dei campi negli insiemi di dati di input o output. È possibile modificare il tag di default specificando un nuovo nome nella colonna *Tag*.

- **Nodo di input.** Visualizza il nodo di input i cui dati verranno combinati.
- **Nodo connesso.** Visualizza il nome del nodo connesso al nodo Unione o Accodamento. Spesso il data mining complesso richiede diverse operazioni di unione che possono includere lo stesso nodo di input. Il nome del nodo connesso consente di operare una distinzione.
- **Campi.** Elenca il numero di campi in ogni sorgente dati.

Visualizza tag correnti. Selezionare questa opzione per visualizzare i tag attualmente utilizzati dal nodo Unione o Accodamento. In altri termini, i tag correnti identificano i collegamenti al nodo che includono il flusso di dati. Per tornare all'immagine dei tubi, i tag correnti sono analoghi ai tubi in cui l'acqua passa attualmente.

Visualizza impostazioni tag inutilizzati. Selezionare questa opzione per visualizzare i tag, o collegamenti, utilizzati in precedenza per la connessione al nodo Unione o Accodamento ma attualmente non connessi a una sorgente dati. Questi tag possono essere paragonati a tubi vuoti e inutilizzati in un sistema di tubature. È possibile scegliere se connettere questi "tubi" a una nuova sorgente oppure rimuoverli. Per rimuovere i tag inutilizzati dal nodo, fare clic su *Cancella*. In questo modo, tutti i tag inutilizzati verranno rimossi con un'unica operazione.

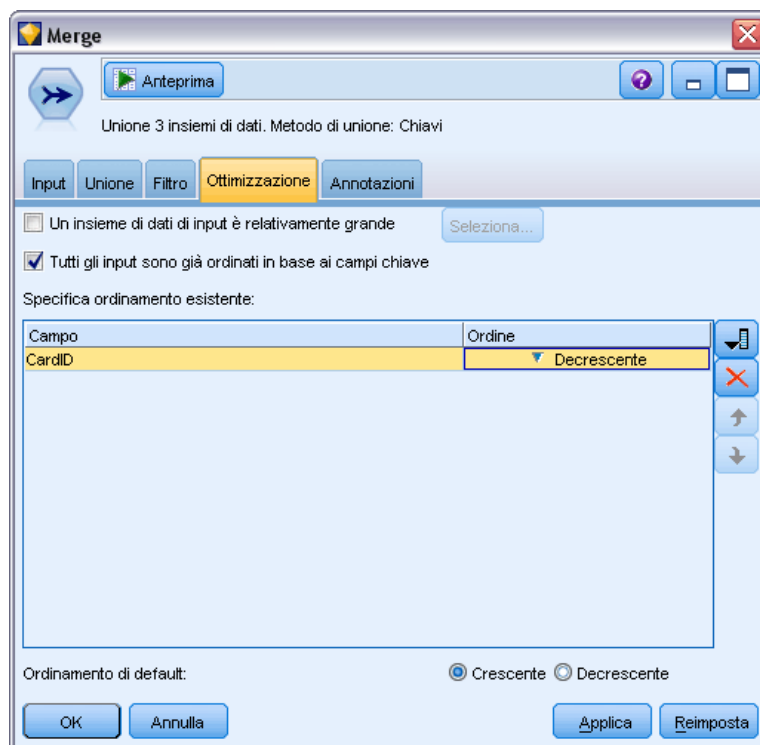
Figura 3-20
Rimozione di tag inutilizzati dal nodo Unione



Impostazioni di ottimizzazione dell'unione

Il sistema fornisce due opzioni che consentono di unire i dati con maggior efficacia in alcune situazioni. Queste opzioni consentono di ottimizzare l'unione quando un insieme di dati di input è significativamente più grande degli altri insiemi di dati oppure quando i dati sono già ordinati in base a tutti o ad alcuni campi chiave utilizzati per l'unione.

Figura 3-21
Impostazioni di ottimizzazione



Un insieme di dati di input è relativamente grande. Selezionare questa opzione per indicare che uno degli insiemi di dati di input è molto più grande degli altri. Il sistema può inserire nella cache gli insiemi di dati più piccoli e quindi eseguire l'unione elaborando l'insieme di dati grande senza inserirlo nella cache o ordinarlo. Questo tipo di join viene normalmente utilizzato con i dati presentati mediante uno schema a stella o un disegno analogo, dove vi è una grande tabella centrale di dati condivisi (è il caso, per esempio, dei dati transazionali). Se si seleziona questa opzione, fare clic su *Seleziona* per specificare l'insieme di dati grande. Si noti che è possibile selezionare solo *un* insieme di dati grande. La seguente tabella riassume i join che è possibile ottimizzare con questo metodo.

Tipo di join	Possibilità di ottimizzazione per un insieme di dati di input grande
Inner	Sì
Parziale	Sì, se l'insieme di dati grande non presenta record incompleti.
Full	No
Anti-join	Sì, se l'insieme di dati grande è il primo input.

Tutti gli input sono già ordinati in base ai campi chiave. Selezionare questa opzione per indicare che i dati di input sono già ordinati in base a uno o più dei campi chiave utilizzati per l'unione. Assicurarsi che *tutti* gli insiemi di dati di input siano ordinati.

Specifica ordinamento esistente. Specificare i campi già ordinati. Utilizzando la finestra di dialogo *Seleziona campi*, aggiungere i campi all'elenco. È possibile effettuare la selezione solo dai campi chiave utilizzati per l'unione (specificati nella scheda *Unione*). Nella colonna *Ordine*, specificare

se ciascun campo è ordinato in ordine crescente o decrescente. Se si specificano più campi, assicurarsi di elencarli secondo l'ordinamento corretto. Utilizzare le frecce a destra dell'elenco per disporre i campi nell'ordine corretto. In caso di errore nell'indicazione dell'ordinamento esistente corretto, al momento dell'esecuzione dello stream verrà visualizzato un errore che indica il numero dei record nei quali l'ordinamento è incoerente rispetto all'impostazione specificata.

Nota: la velocità di unione può beneficiare dell'attivazione dell'elaborazione parallela. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di ottimizzazione in il capitolo 12 in *Manuale dell'utente di IBM SPSS Modeler 14.2*.](#)

Nodo Accodamento

È possibile utilizzare i nodi Accodamento per concatenare insiemi di record. A differenza dei nodi Unione, che consentono di unire record da sorgenti diverse, i nodi Accodamento leggono e passano a valle tutti i record di un'unica sorgente. Vengono quindi letti i record della sorgente successiva utilizzando la stessa struttura di dati della prima o principale sorgente di input (numero di record, di campi e così via). Quando la sorgente principale include più campi rispetto a un'altra sorgente di input, per tutti i valori incompleti verrà utilizzato il valore stringa nullo di sistema (\$null\$).

I nodi Accodamento possono essere utilizzati per combinare insiemi di dati con strutture simili ma dati diversi. È possibile infatti che i dati delle transazioni siano archiviati in file diversi corrispondenti a periodi di tempo diversi, per esempio in un file di dati relativi alle vendite di marzo e in un file diverso relativo alle vendite di aprile. Presupponendo che questi file utilizzino la stessa struttura (ovvero gli stessi campi nello stesso ordine), il nodo Accodamento consente di eseguire il join dei file in un unico file di dimensioni maggiori che sarà quindi possibile analizzare.

Nota: per poter accodare i file, è necessario che i livelli di misurazione dei campi siano simili. Per esempio, a un campo *Nominale* non può essere accodato un campo con livello di misurazione *Continuo*.

Figura 3-22

Finestra di dialogo del nodo Accodamento con visualizzazione delle corrispondenze dei campi in base al nome



Impostazione delle opzioni relative al nodo Accodamento

Corrispondenze campi in base a. Selezionare un metodo da utilizzare per la corrispondenza tra i campi che si desidera accodare.

- **Posizione.** Selezionare questa opzione per accodare insiemi di dati in base alla posizione dei campi nell'origine dati principale. Quando si utilizza questo metodo, è necessario che i dati siano ordinati in modo da garantire un accodamento appropriato.
- **Nome.** Selezionare questa opzione per accodare insiemi di dati in base al nome dei campi negli insiemi di dati di input. Selezionare inoltre Maiuscole/minuscole per attivare la distinzione tra maiuscole e minuscole quando si esegue la corrispondenza tra i nomi dei campi.

Campo di output. Elenca i nodi di input connessi al nodo Accodamento. Il primo nodo nell'elenco è la sorgente di input principale. È possibile ordinare i campi nell'elenco visualizzato facendo clic sull'intestazione della colonna. Questa impostazione, tuttavia, non consente di riordinare i campi nell'insieme di dati.

Includi campi da. Selezionare Solo insieme di dati principale per creare i campi di output in base ai campi dell'insieme di dati principale. L'insieme di dati principale rappresenta il primo input, specificato nella scheda Input. Selezionare Tutti gli insiemi di dati per creare campi di output per tutti i campi di tutti gli insiemi di dati, a prescindere dalla presenza di un campo corrispondente in tutti gli insiemi di dati di input.

Contrassegna i record con insieme dei dati di origine nel campo. Selezionare questa opzione per aggiungere un campo supplementare al file di output i cui valori indicano l'insieme di dati di origine per ogni record. Specificare un nome nella casella di testo. Il nome di campo predefinito è *Input*.

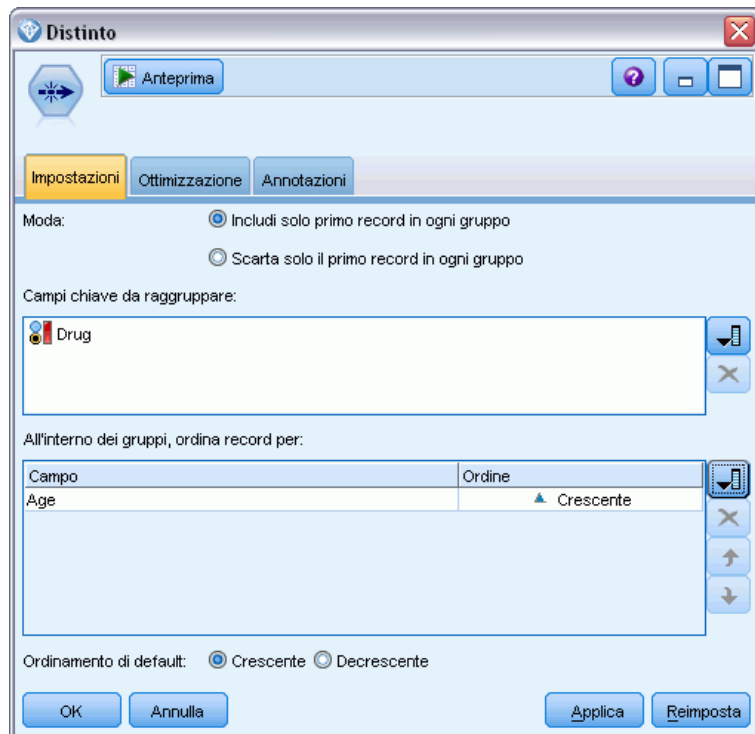
Nodo Elimina duplicati

I record duplicati di un insieme di dati devono essere rimossi prima di iniziare il data mining. Ad esempio, in un database di marketing, è possibile che gli utenti vengano visualizzati più volte con indirizzo o dati della società diversi. Il nodo Elimina duplicati consente di cercare o rimuovere i record duplicati dell'insieme di dati.

Utilizzando il nodo Elimina duplicati, è possibile rimuovere record duplicati passando il primo record distinto nello stream di dati oppure è possibile trovare i record duplicati scartando il primo record e passando nello stream tutti i duplicati.

Inoltre, è possibile definire un ordinamento all'interno dei singoli valori di chiavi distinte per i risultati restituiti. Se si desidera che venga restituita una riga specifica per ogni valore di chiave distinta, è necessario ordinare i record all'interno del nodo Elimina duplicati anziché utilizzare un nodo Ordina a monte (vedere "Ordinamento dei record nel nodo Elimina duplicati" di seguito).

Figura 3-23
Finestra di dialogo del nodo Elimina duplicati



Modalità. Specificare se si desidera includere o escludere (scartare) il primo record.

- **Includi solo primo record in ogni gruppo.** Include il primo record distinto nello stream di dati e rimuove eventuali duplicati.
- **Scarta solo il primo record in ogni gruppo.** Scarta il primo record distinto trovato e passa tutti i record duplicati nello stream di dati. Questa opzione può essere utilizzata per *trovare* duplicati nei dati in modo da analizzarli in seguito all'interno dello stream.

Campi chiave da raggruppare. Elenca il campo o i campi utilizzati per stabilire se i record sono identici. È possibile:

- Aggiungere campi all'elenco utilizzando il pulsante di selezione dei campi a destra.
- Eliminare i campi dall'elenco utilizzando il pulsante X (rimuovi) rosso.

All'interno dei gruppi, ordina record per. Elenca i campi utilizzati per stabilire in che modo i record vengono ordinati all'interno di ogni valore di chiave distinta e se questi sono in ordine crescente o decrescente. È possibile:

- Aggiungere campi all'elenco utilizzando il pulsante di selezione dei campi a destra.
- Eliminare i campi dall'elenco utilizzando il pulsante X (rimuovi) rosso.
- Spostare i campi utilizzando i pulsanti su o giù, se si esegue l'ordinamento in base a più campi.

Ordinamento di default. Specificare se, per default, i record vengono disposti in ordine Crescente o Decrescente.

Ordinamento dei record nel nodo Elimina duplicati

Utilizzando l'opzione All'interno dei gruppi, ordina record per all'interno del nodo Elimina duplicati è possibile restituire una riga specifica per ogni valore di chiave distinta, senza che sia necessario utilizzare un nodo Ordina precedente. Per esempio, si supponga di disporre dei seguenti dati relativi alle età dei consumatori di farmaci prescritti:

Età	Cura
50	Cura A
71	Cura B
44	Cura A
65	Cura X
39	Cura A
75	Cura C
72	Cura Y
57	Cura X
79	Cura Y
69	Cura C
74	Cura B
85	Cura Y
69	Cura X

Per trovare il consumatore più anziano per ogni cura, è necessario specificare l'inserimento del primo record di ogni gruppo, utilizzare Cura come campo chiave e utilizzare Età come campo di ordinamento, impostato su ordine Decrescente. L'ordine di input non ha alcun effetto sul risultato

in quanto la selezione dell'ordinamento specifica quale delle molte righe di una data Cura deve essere restituita. L'output di dati finale sarebbe:

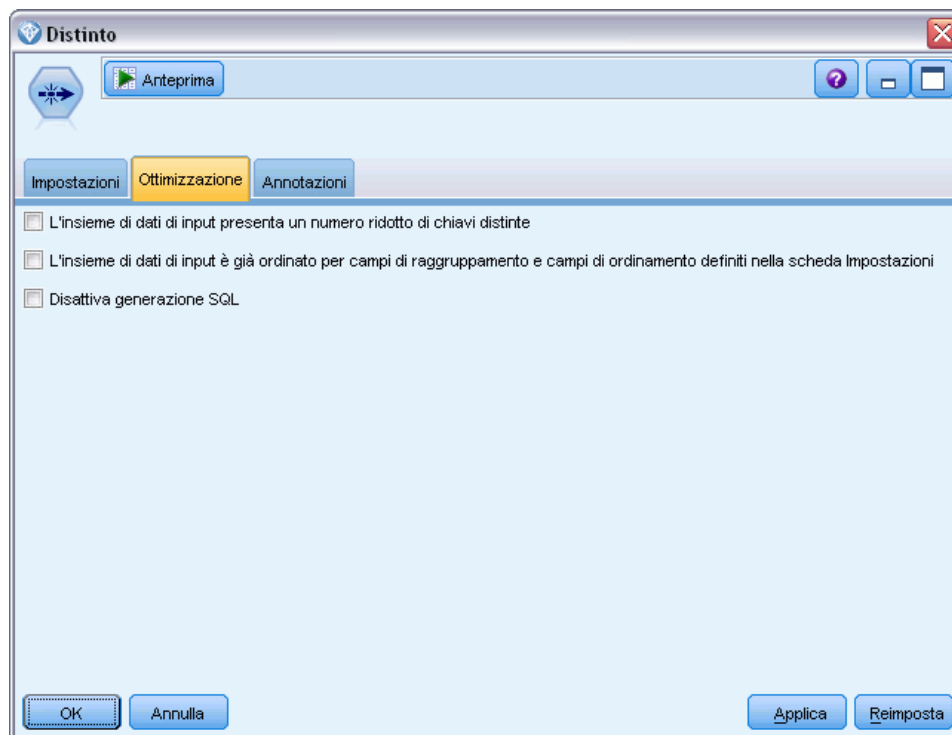
Età	Cura
50	Cura A
74	Cura B
75	Cura C
69	Cura X
85	Cura Y

Impostazioni di ottimizzazione del nodo Elimina duplicati

Se i dati su cui si lavora presentano solo un numero ridotto di record o sono già stati ordinati, è possibile ottimizzarne la gestione per consentire a IBM® SPSS® Modeler di elaborarli in modo più efficiente.

Nota: se si seleziona L'insieme di dati di input presenta un numero ridotto di chiavi distinte o si utilizza la generazione SQL per il nodo, può essere restituita qualsiasi riga nel valore della chiave distinta; per stabilire la riga che deve essere restituita in una chiave distinta, è necessario specificare l'ordinamento utilizzando i campi All'interno dei gruppi, ordina record per nella scheda Impostazioni. Le opzioni di ottimizzazione non influiscono sull'output dei risultati del nodo Elimina duplicati, a condizione che nella scheda Impostazioni sia stato specificato un ordinamento.

Figura 3-24
Impostazioni di ottimizzazione



L'insieme di dati di input presenta un numero ridotto di chiavi distinte. Selezionare questa opzione se i record e/o i valori univoci dei campi chiave sono in numero ridotto. In questo modo si migliorano le prestazioni.

L'insieme di dati di input è già ordinato per campi di raggruppamento e campi di ordinamento definiti nella scheda Impostazioni. Selezionare questa opzione solo se i dati sono già ordinati in base a tutti i campi elencati nella scheda Impostazioni in All'interno dei gruppi, ordina record per e se l'ordine dei dati è lo stesso, crescente o decrescente. In questo modo si migliorano le prestazioni.

Disattiva generazione SQL. Selezionare questa opzione per disattivare la generazione SQL relativa al nodo.

Nodi Operazioni su campi

Panoramica sui nodi Operazioni sui campi

Dopo un'analisi iniziale dei dati, probabilmente sarà necessario selezionare, ripulire o creare i dati per predisporli all'analisi. La palette Operazioni su campi contiene numerosi nodi utili per le operazioni di trasformazione e preparazione.

Per esempio, utilizzando un nodo Nuovo campo, è possibile creare un attributo che attualmente non è rappresentato nei dati. In alternativa, è possibile utilizzare un nodo Discretizza per ricodificare automaticamente i valori di campo per le analisi selezionate. Il nodo Tipo viene utilizzato spesso perché consente di assegnare un livello di misurazione, i valori e un ruolo di modellazione per ogni campo dell'insieme di dati. Queste operazioni sono utili per gestire i valori mancanti e la creazione di modelli a valle.

La palette Operazioni su campi contiene i seguenti nodi:



Il nodo Preparazione automatica dati (ADP) può analizzare i dati e individuare le correzioni, escludere i campi problematici o probabilmente inutili e derivare all'occorrenza nuovi attributi, migliorando le performance grazie allo screening intelligente e alle tecniche di campionamento. Il nodo si può utilizzare in modo completamente automatico, permettendogli di scegliere e di applicare le correzioni, oppure visualizzando in anteprima le modifiche prima dell'applicazione e accettandole, respingendole o modificandole a seconda dei casi. [Per ulteriori informazioni, vedere l'argomento Preparazione automatica dati a pag. 98.](#)



Il nodo Tipo specifica proprietà e metadati di campo. Per esempio, è possibile specificare un livello di misurazione (continuo, nominale, ordinale o flag) per ogni campo, impostare opzioni relative alla gestione dei valori mancanti e dei valori nulli di sistema, impostare il ruolo di un campo per la modellazione, specificare le etichette di campo e valore e specificare i valori per un campo. [Per ulteriori informazioni, vedere l'argomento Nodo Tipo a pag. 125.](#)



Il nodo Filtro filtra (ignora) campi, rinomina campi e mappa campi tra i nodi di input. [Per ulteriori informazioni, vedere l'argomento Applicazione di filtri o ridenominazione di campi a pag. 143.](#)



Il nodo Nuovo campo modifica valori di dati o crea nuovi campi da uno o più campi esistenti. Crea campi di tipo Formula, Flag, Nominale, Stato, Conteggio e Condizionale. [Per ulteriori informazioni, vedere l'argomento Nuovo campo a pag. 154.](#)



Il nodo Risultati classificatore binario combina due o più insiemi di campi al fine di ottenere previsioni più precise di quelle ricavabili dai singoli modelli. [Per ulteriori informazioni, vedere l'argomento Nodo Risultati classificatore binario a pag. 150.](#)



Il nodo Riempimento sostituisce valori di campo e modifica l'archiviazione. È possibile scegliere di sostituire i valori in base a una condizione CLEM, per esempio @BLANK(@FIELD). In alternativa, si può scegliere di sostituire tutti i valori nulli o vuoti con un valore specifico. Il nodo Riempimento è utilizzato spesso in combinazione con il nodo Tipo per sostituire valori mancanti. [Per ulteriori informazioni, vedere l'argomento Nodo Riempimento a pag. 166.](#)



Il nodo Anonimizza consente di mascherare i nomi o i valori dei campi, quando si utilizzano dati da includere in un modello a valle del nodo, permettendo di nascondere i dati originali. Questa funzionalità può essere utile se si desidera consentire ad altri utenti di creare modelli utilizzando dati riservati, quali nomi di clienti o altri dettagli. [Per ulteriori informazioni, vedere l'argomento Nodo Anonimizza a pag. 169.](#)



Il nodo Ricodifica trasforma un insieme di valori categoriali in un altro. L'operazione di ricodifica consente di comprimere categorie o raggruppare dati per l'analisi. [Per ulteriori informazioni, vedere l'argomento Nodo Ricodifica a pag. 174.](#)



Il nodo Discretizza crea automaticamente nuovi campi nominali (insieme) basati sui valori di uno o più campi continui (intervallo numerico) esistenti. Per esempio, è possibile trasformare un campo continuo relativo al reddito in campo categoriale contenente gruppi di reddito come deviazioni dalla media. Dopo aver creato bin per il nuovo campo, è possibile generare un nodo Nuovo campo basato sui punti di taglio. [Per ulteriori informazioni, vedere l'argomento Nodo Discretizza a pag. 178.](#)



Il nodo Analisi RFM (Recency, Frequency, Monetary, Passato recente, Frequenza, Monetario) consente di determinare in modo quantitativo i clienti potenzialmente migliori verificando quanto tempo è trascorso dal loro ultimo acquisto (passato recente), con quale frequenza hanno effettuato acquisti (frequenza) e quanto hanno speso per tutte le transazioni (monetario). [Per ulteriori informazioni, vedere l'argomento Nodo Analisi RFM a pag. 189.](#)



Il nodo Partizione genera un campo di partizione che suddivide i dati in sottoinsiemi separati per le fasi di addestramento, verifica e validazione della creazione del modello. [Per ulteriori informazioni, vedere l'argomento Nodo Partizione a pag. 193.](#)



Il nodo Crea flag crea campi flag in base ai valori categoriali di uno o più campi nominali. [Per ulteriori informazioni, vedere l'argomento Nodo Crea flag a pag. 196.](#)



Il nodo Riorganizza converte un campo nominale o flag in un gruppo di campi in cui è possibile inserire i valori di un altro campo. Per esempio, dato un campo denominato *tipo di pagamento*, con valori di *credito*, *contanti* e *debito*, verrebbero creati tre nuovi campi (*credito*, *contanti*, *debito*), ognuno dei quali può contenere il valore del pagamento effettuato. [Per ulteriori informazioni, vedere l'argomento Nodo Riorganizza a pag. 197.](#)



Il nodo Trasponi scambia i dati delle righe e delle colonne in modo da trasporre i campi in record e i record in campi. [Per ulteriori informazioni, vedere l'argomento Nodo Trasponi a pag. 200.](#)



Il nodo Intervalli di tempo specifica intervalli e, se necessario, crea etichette per la modellazione di dati di serie storica. Se i valori non sono spazati in modo uniforme, il nodo può integrare o aggregare i valori in base alle proprie esigenze per generare un intervallo uniforme tra i record. [Per ulteriori informazioni, vedere l'argomento Nodo Intervalli di tempo a pag. 204.](#)



Il nodo Cronologia crea nuovi campi contenenti dati dei campi di record precedenti. I nodi Cronologia sono utilizzati in genere per dati sequenziali, per esempio per dati di serie storica. Prima di utilizzare un nodo Cronologia, può essere utile ordinare i dati con un nodo Ordina. [Per ulteriori informazioni, vedere l'argomento Nodo Cronologia a pag. 225.](#)



Il nodo Ordina campi definisce l'ordine naturale utilizzato per visualizzare i campi a valle. Tale ordine incide sulla visualizzazione dei campi in vari contesti, quali tabelle, elenchi e Selezione campi. Questa operazione risulta utile se si desidera rendere più visibili i campi interessanti in insiemi di dati di grandi dimensioni. [Per ulteriori informazioni, vedere l'argomento Nodo Ordina campi a pag. 227.](#)

Molti di questi nodi possono essere generati direttamente dal report di esplorazione creato da un nodo Esplora. [Per ulteriori informazioni, vedere l'argomento Generazione di altri nodi per la preparazione dei dati in il capitolo 6 a pag. 383.](#)

Preparazione automatica dati

La preparazione dei dati per l'analisi è una delle fasi più importanti in qualsiasi progetto e, in genere, una delle più lunghe. La funzione Preparazione automatica dati (ADP) svolge questo compito al posto dell'utente, analizzando i dati e individuando le correzioni da apportare, escludendo i campi problematici o probabilmente inutili, derivando nuovi attributi se necessario e migliorando le prestazioni attraverso tecniche di screening intelligenti. L'algoritmo si può utilizzare in modo completamente **automatico**, permettendogli di scegliere e applicare le correzioni, oppure in modo **interattivo**, visualizzando in anteprima le modifiche prima che vengano apportate e accettandole o rifiutandole in funzione delle esigenze.

L'utilizzo di ADP consente di predisporre i dati per la creazione dei modelli in modo semplice e rapido, senza che sia necessario conoscere i concetti statistici impiegati. La creazione e il calcolo del punteggio dei modelli tenderanno a essere più rapidi; inoltre, l'utilizzo di ADP migliora la robustezza dei processi di modellazione automatica quali l'aggiornamento dei modelli e il metodo campione / sfidante.

Nota: quando ADP prepara un campo per l'analisi, crea un nuovo campo che contiene le correzioni o le trasformazioni anziché sostituire i valori e le proprietà esistenti nel vecchio campo. Il vecchio campo non viene utilizzato nelle analisi successive e il suo ruolo viene impostato su Nessuno.

Esempio. Una compagnia di assicurazioni con poche risorse per indagare sulle richieste di indennizzo dei proprietari immobiliari vuole creare un modello per evidenziare le richieste sospette e potenzialmente fraudolente. Prima di procedere, viene effettuata la preparazione automatica dei dati per la creazione del modello. Dal momento che la compagnia ha necessità di esaminare le trasformazioni proposte prima che queste vengano applicate, utilizzerà la preparazione automatica dati in modalità interattiva.

Un gruppo industriale automobilistico tiene traccia delle vendite per un'ampia gamma di autoveicoli personali. Nel tentativo di identificare modelli a basso e alto rendimento è possibile stabilire una relazione tra la vendita dei veicoli e le rispettive caratteristiche. Verrà utilizzata la preparazione automatica dei dati per l'analisi e verranno creati modelli utilizzando i dati "prima" e "dopo" la preparazione per scoprire come cambiano i risultati.

Figura 4-1
Scheda Obiettivo di Preparazione automatica dati

La preparazione automatica dei dati consiglia le fasi di preparazione dei dati che consentono di accelerare la creazione dei modelli, nonché di migliorare il potere predittivo e può comprendere la trasformazione, Creare e selezionare funzioni. È possibile trasformare l'obiettivo.

Qual è l'obiettivo?

- Bilancia velocità e precisione**
Trasforma i dati con un'enfasi sulla creazione dei modelli con un bilanciamento di velocità e precisione.
- Ottimizza per velocità**
Trasforma i dati con un'enfasi sulla creazione dei modelli alla massima velocità possibile.
- Ottimizza per precisione**
Trasforma i dati con un'enfasi sulla creazione dei modelli con il massimo potere predittivo.
- Analisi personalizzata**
Scegliere questa opzione per ottimizzare l'algoritmo nella scheda Impostazioni.

Qual è il proprio obiettivo? La Preparazione automatica dati consiglia una serie di fasi di preparazione dei dati che influiscono sulla velocità con cui altri algoritmi creano modelli e ne migliorano il potere predittivo. Possono comprendere funzioni di trasformazione, creazione e selezione. Anche l'obiettivo può essere trasformato. È possibile specificare le priorità di creazione dei modelli su cui deve concentrarsi il processo di preparazione dei dati.

- **Bilancia velocità e precisione.** Questa opzione prepara i dati in modo da dare la stessa priorità alla velocità di elaborazione dei dati da parte degli algoritmi di creazione del modello e alla precisione delle previsioni.
- **Ottimizza per velocità.** Questa opzione prepara i dati in modo da dare la priorità alla velocità di elaborazione dei dati da parte degli algoritmi di creazione del modello. Selezionare questa opzione quando si utilizzano insiemi di dati molto grandi o quando si desidera ottenere una risposta rapida.
- **Ottimizza per precisione.** Questa opzione prepara i dati in modo da dare la priorità alla precisione delle previsioni generate dagli algoritmi di creazione del modello.
- **Analisi personalizzata.** Selezionare questa opzione se si desidera modificare manualmente l'algoritmo nella scheda Impostazioni. Si noti che questa impostazione viene selezionata automaticamente se in seguito si apportano modifiche incompatibili con uno degli altri obiettivi alle opzioni della scheda Impostazioni.

Training del nodo

Il nodo ADP viene implementato come nodo di elaborazione e funziona in modo simile al nodo Tipo; **effettuare il training** del nodo ADP equivale a istanziare il nodo Tipo. Una volta eseguita l'analisi, le trasformazioni specificate vengono applicate ai dati senza ulteriore analisi, purché il modello di dati a monte resti invariato. Analogamente ai nodi Tipo e Filtro, alla disconnessione il nodo ADP ricorda il modello dati e le trasformazioni, in modo che alla successiva connessione non debba più essere sottoposto a training; in questo modo è possibile eseguirne il training su

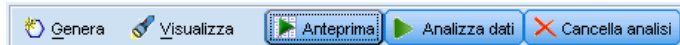
un sottoinsieme di dati tipici e quindi copiarlo o distribuirlo per l'uso su dati in tempo reale per il numero di volte necessario.

Utilizzo della barra degli strumenti

La barra degli strumenti consente di eseguire e aggiornare la visualizzazione dell'analisi dei dati e generare nodi utilizzabili congiuntamente ai dati originali.

Figura 4-2

Preparazione automatica dati - Barra degli strumenti



- **Genera** Da questo menu è possibile generare un nodo Filtro o Nuovo campo. Si noti che questo menu è disponibile solo quando nella scheda Analisi è visualizzata un'analisi.

Il nodo Filtro rimuove i campi di input trasformati. Se si configura il nodo ADP in modo da lasciare i campi di input originali nell'insieme di dati, l'insieme di input originale verrà ripristinato consentendo l'interpretazione del campo punteggio in termini di input. Questo potrebbe essere utile, per esempio, se si desidera produrre un grafico del campo punteggio rispetto a vari input.

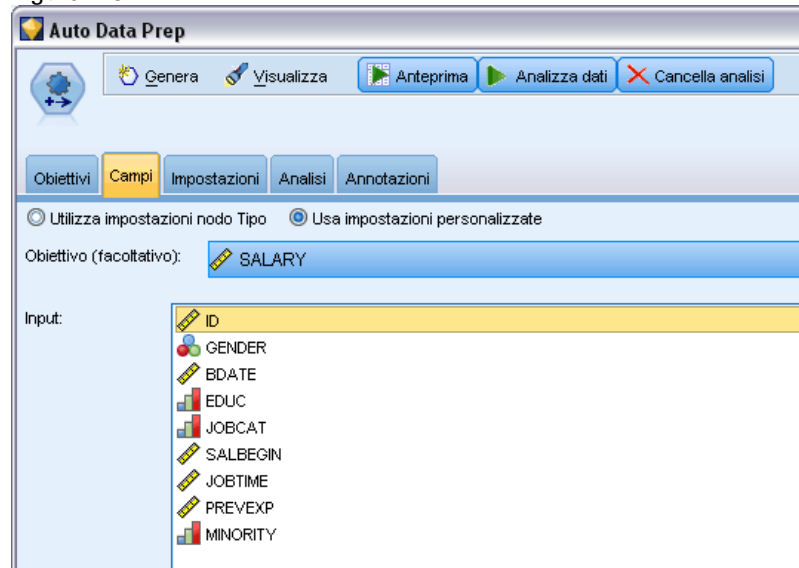
Il nodo Nuovo campo può ripristinare l'insieme di dati originale e le unità obiettivo. il nodo Nuovo campo può essere generato solo quando il nodo ADP contiene un'analisi che ridimensiona l'obiettivo intervallo (ovvero quando è selezionata la modifica della scala di Box-Cox nel riquadro Prepara input e obiettivo). Non è possibile generare un nodo Nuovo campo se l'obiettivo non è un intervallo o se non è selezionata la modifica della scala di Box-Cox. [Per ulteriori informazioni, vedere l'argomento Generazione di un nodo Nuovo campo a pag. 123.](#)
- **Visualizza** Contiene opzioni che determinano quali elementi vengono visualizzati nella scheda Analisi. Tali opzioni includono i controlli per la modifica dei grafici e le selezioni di visualizzazione sia per il riquadro principale sia per le visualizzazioni collegate.
- **Anteprima** Visualizza un esempio delle trasformazioni che saranno applicate ai dati di input. [Per ulteriori informazioni, vedere l'argomento Anteprima dei dati nei nodi in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Analizza dati** Avvia un'analisi secondo le impostazioni correnti e visualizza i risultati nella scheda Analisi.
- **Cancella analisi** Elimina l'analisi esistente (disponibile solo quando esiste un'analisi corrente).

Stato del nodo

Lo stato del nodo ADP nell'area di disegno di IBM® SPSS® Modeler è indicato da una freccia o da un segno di spunta sull'icona che mostra se l'analisi è stata effettuata oppure no.

Scheda Campi

Figura 4-3



Per poter generare un modello, è necessario prima specificare i campi da utilizzare come obiettivi e come input. Con alcune eccezioni, tutti i campi Modelli utilizzano le informazioni sui campi di un nodo Tipo a monte. Se si utilizza un nodo Tipo per selezionare i campi obiettivo e di input, non è necessario cambiare nessuna delle impostazioni presenti in questa scheda.

Utilizza impostazioni nodo Tipo. Questa opzione indica al nodo di utilizzare le informazioni sui campi da un nodo Tipo a monte. È l'impostazione di default.

Utilizza impostazioni personalizzate. Questa opzione indica al nodo di utilizzare le informazioni sui campi specificate qui al posto di quelle date in un qualsiasi nodo Tipo a monte. Dopo avere selezionato questa opzione, specificare i campi nell'area sottostante come richiesto.

Obiettivo. Per i modelli che richiedono uno o più campi obiettivo, selezionare il campo o i campi obiettivo. Questa operazione è simile all'impostazione del ruolo di un campo su *Obiettivo* in un nodo Tipo.

Input. Selezionare il campo o i campi di input. Questa operazione è simile all'impostazione del ruolo di un campo su *Input* in un nodo Tipo.

Scheda Impostazioni

La scheda Impostazioni contiene vari gruppi di impostazioni che è possibile modificare per perfezionare l'elaborazione dei dati da parte dell'algoritmo. Se si apportano modifiche alle impostazioni predefinite che risultano incompatibili con gli altri obiettivi, la scheda Obiettivo viene aggiornata automaticamente per selezionare l'opzione Personalizza analisi.

Impostazioni campo

Figura 4-4
Preparazione automatica dati: Impostazioni campo

Le impostazioni campi non vengono modificate se si cambia l'obiettivo.

Usa campo frequenza

Usa campo peso

Come gestire i campi esclusi dai modelli:

Filtra i campi inutilizzati

Imposta direzione campi inutilizzati su "Nessuna"

Se i campi in entrata non corrispondono all'analisi esistente:

Interrompi l'esecuzione e tieni l'analisi esistente

Cancella l'analisi esistente e analizza i nuovi dati

Utilizza campo frequenza. Questa opzione consente di selezionare un campo come ponderazione della frequenza. Utilizzarla se i record dei dati di training rappresentano più di un'unità ciascuno; ad esempio, se si utilizzano dei dati aggregati. I valori del campo dovrebbero essere rappresentati dal numero di unità rappresentate da ciascun record. [Per ulteriori informazioni, vedere l'argomento Utilizzo dei campi frequenza e peso. in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

Utilizza campo peso. Questa opzione consente di selezionare un campo come ponderazione del caso. I pesi dei casi si utilizzano per tenere conto delle differenze nella varianza tra i vari livelli del campo di output. [Per ulteriori informazioni, vedere l'argomento Utilizzo dei campi frequenza e peso. in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

Come gestire i campi esclusi dai modelli. Specifica cosa accade ai campi esclusi; è possibile scegliere di filtrarli per escluderli dai dati o semplicemente di impostarne il *ruolo* su Nessuno.

Se i campi in entrata non corrispondono all'analisi esistente. Specificare cosa accade se uno o più campi di input richiesti mancano dall'insieme di dati in arrivo quando si esegue un nodo ADP sottoposto a training.

- **Interrompi l'esecuzione e tieni l'analisi esistente.** Interrompe il processo di esecuzione, preserva le informazioni dell'analisi corrente e visualizza un errore.
- **Cancella l'analisi esistente e analizza i nuovi dati.** Cancella l'analisi esistente, analizza i dati in arrivo e applica le trasformazioni consigliate a tali dati.

Prepara date e ore

Figura 4-5

Preparazione automatica dati: impostazioni di Prepara date e ore



Prepara date e ore per la modellazione

Calcola durata

Calcola tempo trascorso fino alla data di riferimento

Data di riferimento

Data odierna

Data fissa

Data: 2009-05-22

Unità per la durata della data

Automatica

Unità fisse

Unità: Mesi

Calcola tempo trascorso fino all'ora di riferimento

Ora di riferimento

Ora corrente

Ora fissa

Ora: 09:40:43

Unità per la durata dell'ora

Automatica

Unità fisse

Unità: Ore

Estrai elementi di tempo ciclico

Estrai dalle date:

Anno Mese Giorno

Estrai dalle ore:

Ora Minuto Secondo

Molti algoritmi di modellazione non sono in grado di gestire direttamente i dettagli relativi a date e ore; queste impostazioni consentono di derivare nuovi dati sulle durate utilizzabili come input per i modelli dalle date e dalle ore indicate nei dati esistenti. I campi contenenti date e ore devono essere predefiniti con tipi di archiviazione data o ora. L'uso dei campi data e ora originali come input per i modelli in seguito alla preparazione automatica dei dati non è consigliato.

Prepara date e ore per la modellazione. Se si deseleziona questa opzione vengono disabilitati tutti gli altri controlli Prepara date e ore ma vengono mantenute le selezioni.

Calcola tempo trascorso fino alla data di riferimento. Produce il numero di anni/mesi/giorni a partire da una data di riferimento per ciascuna variabile che contiene delle date.

- **Data di riferimento.** Specifica la data a partire da cui sarà calcolata la durata relativamente alle informazioni sulla data presenti nei dati di input. Se si seleziona Data odierna, viene sempre utilizzata la data corrente del sistema per l'esecuzione di ADP. Per utilizzare una data specifica, selezionare Data fissa e immettere la data desiderata. La data corrente viene immessa automaticamente nel campo Data fissa quando viene creato il nodo.
- **Unità per la durata della data.** Specificare se ADP deve decidere automaticamente l'unità per la durata della data oppure selezionarne una da Unità fisse in Anni, mesi o Giorni.

Calcola tempo trascorso fino all'ora di riferimento. Produce il numero di ore/minuti/secondi a partire da un'ora di riferimento per ciascuna variabile che contiene delle ore.

- **Ora di riferimento.** Specifica l'ora a partire dalla quale sarà calcolata la durata relativamente alle informazioni sull'ora presenti nei dati di input. Se si seleziona Ora corrente, viene sempre utilizzata l'ora corrente del sistema per l'esecuzione di ADP. Per utilizzare un'ora specifica, selezionare Ora fissa e immettere l'ora desiderata. L'ora corrente viene immessa automaticamente nel campo Ora fissa quando viene creato il nodo.
- **Unità per la durata dell'ora.** Specificare se ADP deve decidere automaticamente l'unità per la durata dell'ora oppure selezionarne una da Unità fisse in Ore, minuti o Secondi.

Estrai elementi di tempo ciclico. Utilizzare queste impostazioni per suddividere un singolo campo data o ora in uno o più campi. Ad esempio, se si selezionano tutte e tre le caselle di controllo della data, il campo data di input "1954-05-23" viene suddiviso in tre campi: 1954, 5 e 23. Ciascuno utilizza il suffisso definito nel riquadro Nomi dei campi, e il campo data originale viene ignorato.

- **Estrai dalle date.** Per qualsiasi input di data, specificare se si desidera estrarre gli anni, i mesi, i giorni o una combinazione dei tre elementi.
- **Estrai dalle ore.** Per qualsiasi input di ora, specificare se si desidera estrarre le ore, i minuti, i secondi o una combinazione dei tre elementi.

Escludi campi

Figura 4-6

Preparazione automatica dati: impostazioni di Escludi campi

I campi delle costanti saranno sempre esclusi

Escludi campi di input di scarsa qualità

Escludi campi di input

Escludi campi con troppi valori mancanti

Percentuale massima di valori mancanti: %

Escludi campi nominali con troppe categorie univoche

Numero massimo di categorie:

Escludi campi categoriali con troppi valori in una categoria singola

Percentuale massima in categoria singola: %

La scarsa qualità dei dati può influire sulla precisione delle previsioni; pertanto, è possibile specificare il livello di qualità accettabile per le funzioni di input. Tutti i campi che sono costanti o hanno il 100% dei valori mancanti vengono esclusi automaticamente.

Escludi campi di input a bassa qualità. Se si deseleziona questa opzione vengono disabilitati tutti gli altri controlli Escludi campi ma vengono mantenute le selezioni.

Escludi campi con troppi valori mancanti. I campi con una percentuale di valori mancanti superiore a quella specificata vengono eliminati dalla successiva analisi. Specificare un valore superiore o uguale a 0, che equivale a deselezionare questa opzione, e inferiore o uguale a 100, benché i campi con tutti i valori mancanti vengano automaticamente esclusi. Il valore di default è 50.

Escludi campi nominali con troppe categorie univoche. I campi nominali con un numero di categorie superiore a quello specificato vengono eliminati dalla successiva analisi. Specificare un intero positivo. Il valore predefinito è 100. È utile per rimuovere automaticamente i campi

che contengono informazioni esclusive del record provenienti dalla creazione del modello, quali ID, indirizzo o nome.

Escludi campi categoriali con troppi valori in una categoria singola. I campi ordinali e nominali con una categoria contenente una percentuale di record superiore a quella specificata vengono eliminati dalla successiva analisi. Specificare un valore superiore o uguale a 0, che equivale a deselezionare questa opzione, e inferiore o uguale a 100, benché i campi costanti vengano automaticamente esclusi. Il valore di default è 95.

Preparazione degli input e degli obiettivi

Dal momento che nessun dato si trova mai in uno stato perfetto per l'elaborazione, è possibile regolare alcune delle impostazioni prima di eseguire un'analisi. Ad esempio, è possibile rimuovere i valori anomali, specificando come devono essere gestiti i valori mancanti o correggendo il tipo.

Nota: se si modificano i valori in questo riquadro, la scheda Obiettivi viene aggiornata automaticamente per selezionare l'opzione Analisi personalizzata.

Figura 4-7

Preparazione automatica dati: impostazione degli input e degli obiettivi

Prepara i campi di input e obiettivo per la modellazione

Modifica tipo e migliora qualità dati

Input	Obiettivo	
<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	Modifica il tipo di campi numerici (ordinali e continui)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Riordina campi nominali in modo da visualizzare per prima la categoria più piccola e per ultima quella più grande
<input type="checkbox"/>	<input type="checkbox"/>	Sostituisci i valori anomali nei campi continui (consigliato per campi di input se saranno messi in scala comune)
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Campi continui: sostituisci i valori mancanti con media
<input checked="" type="checkbox"/>	<input type="checkbox"/>	Campi nominali: sostituisci i valori mancanti con modalità
<input type="checkbox"/>	<input type="checkbox"/>	Campi ordinali: sostituisci i valori mancanti con mediana

Numero massimo di valori per i campi ordinali:

Numero minimo di valori per i campi continui:

Punto di interruzione dei valori anomali: (deviazioni standard)

Metodo di sostituzione dei valori anomali: Sostituisci con valore di interruzione Elimina valore

Trasforma campo continuo

Metti tutti i campi di input continui in una scala comune (fortemente consigliato se verrà creata la funzione)

Metodo di ridimensionamento: Media finale: Deviazione standard finale:

Ridimensiona obiettivo continuo con trasformazione di Box-Cox per ridurre asimmetria

Media finale: Deviazione standard finale:

Prepara i campi di input e obiettivo per la modellazione. Attiva/disattiva tutti i campi nel riquadro.

Modifica tipo e migliora qualità dati. Per gli input e l'obiettivo è possibile specificare separatamente diverse trasformazioni dei dati; ciò è utile quando non si desidera modificare i valori dell'obiettivo. Ad esempio, una previsione di reddito in dollari ha più senso di una misurata in log(dollari). Inoltre, se l'obiettivo ha dei valori mancanti, non è presente alcun guadagno previsto per riempire i valori mancanti, mentre riempire i valori mancanti negli input può rendere possibili alcuni algoritmi di elaborazione delle informazioni che altrimenti andrebbero persi.

Altre impostazioni per queste trasformazioni, quale ad esempio il valore di interruzione anomalo, sono comuni sia all'obiettivo che agli input.

È possibile selezionare le impostazioni per gli input, l'obiettivo o entrambe le categorie:

- **Modifica il tipo di campi numerici.** Selezionare questa opzione per determinare se i campi numerici con il livello di misurazione *Ordinale* possono essere convertiti in *Continuo*, o viceversa. Per controllare la conversione, si specificano i valori di soglia minimo e massimo.
- **Riordina campi nominali.** Selezionare questa opzione per ordinare i campi (insieme) nominali, dalla categoria minima a quella massima.
- **Sostituisci valori anomali nei campi continui.** Specifica se sostituire i valori anomali; utilizzare questa opzione in combinazione con le opzioni Metodo di sostituzione dei valori anomali seguenti.
- **Campi continui: sostituisci valori mancanti con media.** Selezionare questa opzione per sostituire i valori mancanti delle funzioni (intervallo) continue.
- **Campi nominali: sostituisci valori mancanti con moda.** Selezionare questa opzione per sostituire i valori mancanti delle funzioni (insieme) nominali.
- **Campi ordinali: sostituisci valori mancanti con mediana.** Selezionare questa opzione per sostituire i valori mancanti delle funzioni (insieme ordinato) ordinali.

Numero massimo di valori per i campi ordinali. Specifica la soglia per ridefinire i campi (insieme ordinato) ordinali come continui (intervallo). Il valore predefinito è 10; quindi, se un campo ordinale ha più di 10 categorie viene ridefinito come continuo (intervallo).

Numero minimo di valori per i campi continui. Specifica la soglia per ridefinire i campi scala o continui (intervallo) come ordinali (insieme ordinato). Il valore predefinito è 5; quindi, se un campo continuo ha meno di 5 valori viene ridefinito come ordinale (insieme ordinato).

Valore di interruzione anomalo. Specifica il criterio di interruzione dei valori anomali, misurato in deviazioni standard; il valore predefinito è 3.

Metodo di sostituzione dei valori anomali. Selezionare se i valori anomali devono essere sostituiti tagliandoli (coercizione) con il punto di interruzione o eliminati impostandoli come valori mancanti. Tutti i valori anomali impostati su valori mancanti rispondono alle impostazioni per la gestione dei valori mancanti selezionate in precedenza.

Metti tutti i campi di input continui in una scala comune. Per normalizzare i campi di input continui, selezionare questa casella di controllo e scegliere il metodo di normalizzazione. Il valore predefinito è trasformazione punteggio Z, dove è possibile specificare un valore per Media finale, il cui valore predefinito è 0, e per Deviazione standard finale, il cui valore predefinito è 1. In alternativa, è possibile scegliere di utilizzare Trasformazione Min/Max e specificare i valori minimo e massimo, i cui valori predefiniti sono rispettivamente 0 e 100.

Questo campo è particolarmente utile quando si seleziona Esegui creazione funzioni nel riquadro Crea e seleziona funzioni.

Ridimensiona obiettivo continuo con trasformazione di Box-Cox. Per normalizzare un campo obiettivo (scala o intervallo) continuo, selezionare questa casella di controllo. I valori predefiniti della trasformazione di Box-Cox sono 0 per Media finale e 1 per Deviazione standard finale.

Nota: se si decide di normalizzare l'obiettivo, la dimensione dell'obiettivo sarà trasformata. In questo caso può essere necessario generare un nodo Nuovo campo per applicare una trasformazione inversa al fine di riconvertire le unità trasformate in un formato riconoscibile a scopi di ulteriori elaborazioni. [Per ulteriori informazioni, vedere l'argomento Generazione di un nodo Nuovo campo a pag. 123.](#)

Creazione e selezione delle funzioni

Per migliorare il potere predittivo dei dati è possibile trasformare i campi di input o crearne di nuovi basati su quelli esistenti.

Nota: se si modificano i valori in questo riquadro, la scheda Obiettivi viene aggiornata automaticamente per selezionare l'opzione Analisi personalizzata.

Figura 4-8

Preparazione automatica dati: impostazioni di trasformazione, creazione e selezione

Trasformare, creare e selezionare campi di input per rafforzare il potere predittivo

Campi di input categoriali

Unisci categorie poco dense per aumentare al massimo l'associazione all'obiettivo Valore P:

I campi di input con una sola categoria dopo l'unione supervisionata saranno esclusi.

Quando non ci sono obiettivi, unire le categorie poco dense in base ai conteggi

Funzioni ordinali Funzioni nominali % minima di casi in una categoria:

Campi di input continui

Discretizza i campi continui e mantieni il potere predittivo (disponibile solo per l'obiettivo categoriale)

Valore P:

I campi di input con una sola categoria dopo la discretizzazione saranno esclusi.

Selezione e creazione della funzione

Esegui selezione funzione Valore P:

La selezione della funzione si applica agli input continui quando l'obiettivo è continuo e agli input categoriali

Esegui creazione funzione

La funzione di creazione viene applicata agli input continui quando l'obiettivo è continuo oppure non esiste

Trasformare, creare e selezionare campi di input per rafforzare il potere predittivo. Attiva/disattiva tutti i campi nel riquadro.

Unisci categorie poco dense per aumentare al massimo l'associazione all'obiettivo. Selezionare questa opzione per creare un modello più gestibile riducendo il numero delle variabili da elaborare in associazione all'obiettivo. Se richiesto, modificare il valore predefinito 0,05 della probabilità.

Se tutte le categorie vengono unite in una, le versioni originale e derivata del campo vengono escluse perché non hanno un valore come predittore.

Quando non ci sono obiettivi, unisci le categorie poco dense in base ai conteggi. Se si utilizzano dati privi di obiettivo, è possibile scegliere di unire categorie poco dense di funzioni ordinali (insieme ordinato) e/o nominali (insieme). Specificare la percentuale minima di casi, o record, nei dati che identificano le categorie da unire; il valore predefinito è 10.

Le categorie vengono unite mediante le seguenti regole:

- L'unione non viene effettuata sui campi binari.
- Se durante l'unione sono presenti solo due categorie, l'unione si interrompe.
- Se non è presente alcuna categoria originale né alcuna categoria creata durante l'unione, con una percentuale di casi inferiori a quella minima specificata, l'unione si interrompe.

Categorizza i campi continui e mantieni il potere predittivo. Se l'insieme di dati comprende un obiettivo categoriale, è possibile categorizzare gli input continui con associazioni forti per migliorare le prestazioni in sede di elaborazione. Se richiesto, modificare il valore predefinito 0,05 della probabilità per i sottoinsiemi omogenei.

Se con l'operazione di categorizzazione si ottiene un unico intervallo per un determinato campo, la versione originale e categorizzata del campo vengono escluse perché sono prive di valore come predittori.

Nota: la categorizzazione in ADP differisce dalla categorizzazione ottimale utilizzata in altre parti di IBM® SPSS® Modeler. La categorizzazione ottimale utilizza le informazioni relative all'entropia per convertire una variabile continua in una categoriale; per far questo è necessario ordinare i dati e archivarli tutti in memoria. ADP utilizza sottoinsiemi omogenei per categorizzare una variabile continua e, pertanto, la categorizzazione in ADP non richiede di ordinare i dati e non li archivia tutti in memoria. Quando si utilizza il metodo dei sottoinsiemi omogenei per categorizzare una variabile continua, il numero di categorie dopo la categorizzazione è sempre inferiore o uguale al numero di categorie dell'obiettivo.

Effettua selezione delle funzioni. Selezionare questa opzione per rimuovere le funzioni con un coefficiente di correlazione basso. Se richiesto, modificare il valore predefinito 0,05 della probabilità.

Questa opzione viene applicata solo alle funzioni di input continue in cui l'obiettivo è continuo e alle funzioni di input categoriali.

Esegui creazione funzioni. Selezionare questa opzione per derivare nuove funzioni da una combinazione di varie funzioni esistenti (che vengono quindi escluse dalla modellazione).

Questa opzione viene applicata solo alle funzioni di input continue quando l'obiettivo è continuo oppure non esiste.

Nomi campi

Figura 4-9

Preparazione automatica dati: impostazioni di Nomina campi

The screenshot shows a configuration window with three main sections:

- Campi trasformati e creati:**
 - Estensione nome campi obiettivo trasformati:
 - Estensione nome campi di input trasformati:
 - Nome di base funzioni create:
- Durate calcolate da date e ore:**
 - Estensioni dei nomi per le durate calcolate dalle date:
 - Anni: Mesi: Giorni:
 - Estensioni dei nomi per le durate calcolate dalle ore:
 - Ore: Minuti: Secondi:
- Elementi ciclici estratti da date e ore:**
 - Estensioni dei nomi per elementi ciclici estratti dalle date:
 - Anno: Mese: Giorno:
 - Estensioni dei nomi per elementi ciclici estratti dalle ore:
 - Ora: Minuto: Secondo:

Per individuare facilmente le funzioni nuove e trasformate, ADP crea e applica nuovi nomi, prefissi o suffissi di base. I nomi si possono modificare in modo da renderli più pertinenti rispetto alle esigenze e ai dati dell'utente. Se si desidera specificare altre etichette è necessario farlo in un nodo Tipo a valle.

Campi trasformati e creati. Specificare le estensioni dei nomi da applicare ai campi obiettivo e di input trasformati.

Si noti che nel nodo ADP l'impostazione dei campi stringa come vuoti può provocare un errore, a seconda di come si decide di gestire i campi non utilizzati. Se Come gestire i campi esclusi dai modelli è impostato su Filtra i campi inutilizzati nel riquadro Impostazioni campo della scheda Impostazioni, le estensioni dei nomi per gli input e l'obiettivo si possono impostare su "nothing". I campi originali vengono esclusi e i campi trasformati vengono salvati al loro posto; in questo caso i nuovi campi trasformati avranno lo stesso nome degli originali.

Tuttavia, se si è selezionato Imposta direzione campi inutilizzati su 'Nessuna', la presenza di estensioni di nomi vuote o nulle per l'obiettivo e per gli input provoca un errore perché in questo modo si tenta di creare nomi di campi duplicati.

Specificare inoltre il nome del prefisso da applicare a tutte le funzioni da creare con le impostazioni Seleziona e Crea. Il nuovo nome viene creato apponendo un suffisso numerico al nome radice del prefisso. Il formato del numero dipende dal numero di nuove funzioni da derivare, ad esempio:

- le funzioni create da 1 a 9 saranno denominate: da funzione1 a funzione9.
- le funzioni create da 10 a 99 saranno denominate: da funzione01 a funzione99.
- le funzioni create da 100 a 999 saranno denominate: da funzione001 a funzione999, e così via.

In questo modo, le funzioni create saranno organizzate in un ordine logico indipendentemente dal loro numero.

Durate calcolate da date e ore. Specificare le estensioni dei nomi da applicare alle durate calcolate a partire da date e ore.

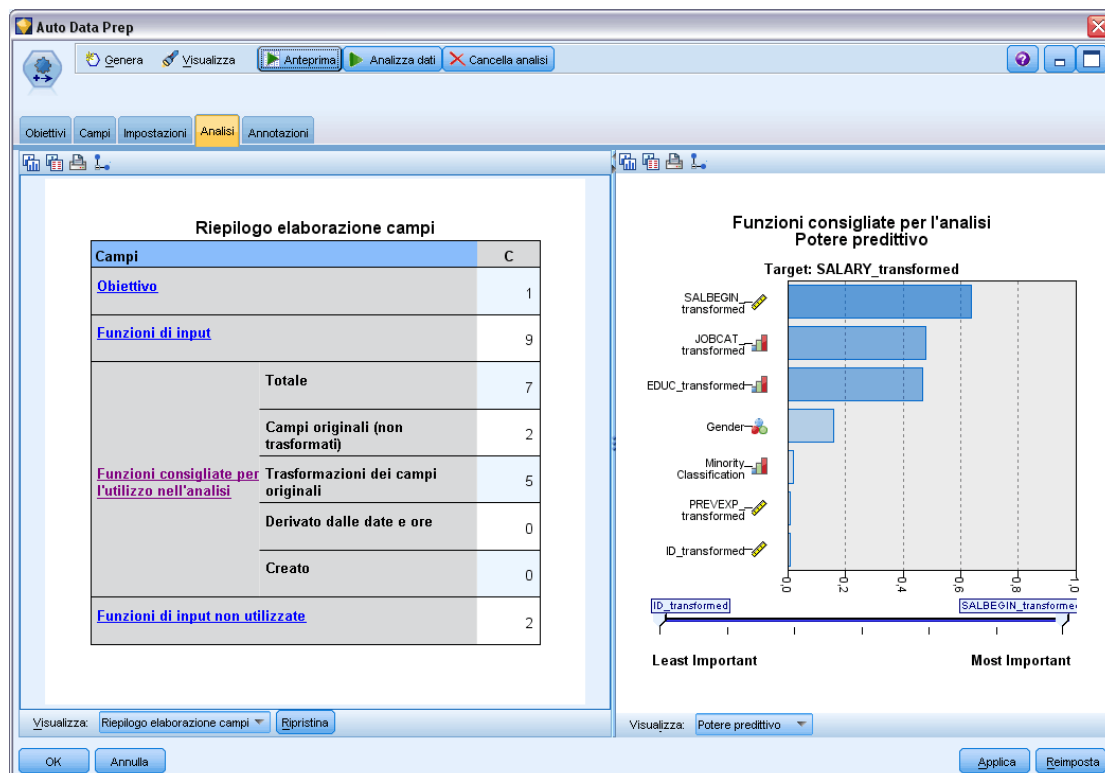
Elementi ciclici estratti da date e ore. Specificare le estensioni dei nomi da applicare agli elementi ciclici estratti da date e ore.

Scheda Analisi

- Quando le impostazioni di ADP sono soddisfacenti (comprese le eventuali modifiche apportate alla scheda Obiettivo, Campi e Impostazioni), fare clic su **Analizza dati**; l'algoritmo applica le impostazioni ai dati immessi e visualizza i risultati nella scheda Analisi.

La scheda Analisi contiene output in formato tabellare e grafico che riassume l'elaborazione dei dati e visualizza le raccomandazioni su eventuali modifiche o miglioramenti da apportare ai dati per il calcolo del punteggio. Questo consente di esaminare e accettare o rifiutare tali raccomandazioni.

Figura 4-10
Scheda Analisi di Preparazione automatica dati



La scheda Analisi è composta da due riquadri, la visualizzazione principale a sinistra e quella collegata o ausiliaria a destra. Le visualizzazioni principali sono tre:

- Riepilogo elaborazione campi (impostazione predefinita). [Per ulteriori informazioni, vedere l'argomento Riepilogo elaborazione campi a pag. 112.](#)
- Campi. [Per ulteriori informazioni, vedere l'argomento Campi a pag. 113.](#)
- Riepilogo delle azioni [Per ulteriori informazioni, vedere l'argomento Riepilogo delle azioni a pag. 115.](#)

Le visualizzazioni collegate/ausiliarie sono quattro:

- Potere predittivo (impostazione predefinita). [Per ulteriori informazioni, vedere l'argomento Potere predittivo a pag. 116.](#)
- Tabella campi. [Per ulteriori informazioni, vedere l'argomento Tabella Campi a pag. 117.](#)
- Dettagli campo. [Per ulteriori informazioni, vedere l'argomento Dettagli campo a pag. 118.](#)
- Dettagli dell'azione. [Per ulteriori informazioni, vedere l'argomento Dettagli dell'azione a pag. 120.](#)

Collegamenti tra le visualizzazioni

All'interno della visualizzazione principale, il testo sottolineato nelle tabelle controlla la visualizzazione nella visualizzazione collegata. Se si fa clic sul testo è possibile visualizzare i dettagli di un determinato campo, insieme di campi o fase di elaborazione. Il collegamento selezionato per ultimo è visualizzato con un colore più scuro per facilitare l'individuazione del collegamento tra il contenuto dei due riquadri.

Reimpostazione delle visualizzazioni

Per visualizzare nuovamente le raccomandazioni originali della scheda Analisi e annullare le eventuali modifiche apportate alle visualizzazioni Analisi, fare clic su Reimposta nella parte inferiore del riquadro di visualizzazione principale.

Riepilogo elaborazione campi

Figura 4-11
Riepilogo elaborazione campi

Campi	C
<u>Obiettivo</u>	1
<u>Funzioni di input</u>	9
Totale	7
Campi originali (non trasformati)	2
<u>Funzioni consigliate per l'utilizzo nell'analisi</u>	5
Trasformazioni dei campi originali	5
Derivato dalle date e ore	0
Creato	0
<u>Funzioni di input non utilizzate</u>	2

La tabella Riepilogo elaborazione campi offre un'istantanea dell'impatto complessivo stimato dell'elaborazione, comprese le modifiche dello stato delle funzioni e il numero di funzioni create.

Si noti che non viene effettivamente creato alcun modello e, pertanto, non vi è alcuna misura o grafico della variazione del potere predittivo generale prima e dopo la preparazione dei dati; è possibile invece visualizzare i grafici del potere predittivo dei singoli predittori consigliati.

La tabella riporta le seguenti informazioni:

- Il numero di campi obiettivo.
- Il numero di predittori (input) originali.
- I predittori consigliati per l'uso nell'analisi e nella modellazione. Sono inclusi il numero totale di campi consigliati; il numero di campi originali, non trasformati consigliati; il numero di campi trasformati consigliati (escluse le versioni intermedie di qualsiasi campo, i campi derivati dai predittori di data/ora e i predittori creati); il numero di campi consigliati derivati dai campi data/ora e il numero di predittori creati consigliati.
- Il numero di predittori di input non consigliato per l'uso in nessuna forma, che si tratti della forma originale, come campo derivato o come input per un predittore creato.

Se le informazioni dei Campi sono sottolineate, farvi clic sopra per visualizzare ulteriori dettagli in una visualizzazione collegata. I dettagli relativi a Obiettivo, Funzioni di input e Funzioni di input non utilizzate sono riportati nella visualizzazione collegata Tabella campi. [Per ulteriori informazioni, vedere l'argomento Tabella Campi a pag. 117.](#) Le funzioni consigliate per l'analisi sono visualizzate nella visualizzazione collegata Potere predittivo. [Per ulteriori informazioni, vedere l'argomento Potere predittivo a pag. 116.](#)

Campi

Figura 4-12
Fields

Campi

Obiettivo

Nome	Tipo
SALARY	

Funzioni Includi campi non raccomandati nella tabella

Versione da utilizzare	Nome	Tipo	Potere predittivo
Trasformata	SALBEGIN		0,64
Trasformata	JOB CAT		0,48
Trasformata	EDUC		0,47
Originale	GENDER		0,16
Originale	MINORITY		0,02
Trasformata	PREVEXP		0,01

La visualizzazione principale Campi mostra i campi elaborati e indica se ADP ne consiglia o meno l'utilizzo nei modelli a valle. È possibile ignorare le raccomandazioni di tutti i campi, ad esempio per escludere funzioni create o includere funzioni di cui ADP consiglia l'esclusione. Se un campo è stato trasformato, è possibile decidere se accettare la trasformazione suggerita o utilizzare la versione originale.

La visualizzazione Campi è composta da due tabelle, una per l'obiettivo e una per i predittori elaborati o creati.

Tabella Obiettivo

La tabella Obiettivo è visualizzata solo se nei dati è stato definito un obiettivo.

La tabella contiene due colonne:

- **Nome.** Si tratta del nome o dell'etichetta del campo obiettivo. Viene sempre utilizzato il nome originale, anche se il campo è stato trasformato.
- **Livello di misurazione.** In questa colonna è visualizzata l'icona che rappresenta il livello di misurazione; passare il puntatore del mouse sopra l'icona per visualizzare un'etichetta (continuo, ordinale, nominale e così via) che descrive i dati.

Se l'obiettivo è stato trasformato, la colonna Livello di misurazione riflette la versione trasformata finale. *Nota:* non è possibile disattivare le trasformazioni per l'obiettivo.

Tabella Predittori

La tabella Predittori è sempre visualizzata. Ogni riga della tabella rappresenta un campo. Per impostazione predefinita, le righe sono ordinate in modo decrescente in base al potere predittivo.

Per le funzioni ordinarie, il nome originale viene sempre utilizzato come nome della riga. Nella tabella sono riportate sia le versioni originali che quelle derivate dei campi data/ora (in righe separate); la tabella comprende anche i predittori creati.

Si noti che le versioni trasformate dei campi visualizzate nella tabella rappresentano sempre le versioni finali.

Per impostazione predefinita, nella tabella Predittori sono visualizzati solo i campi consigliati. Per visualizzare gli altri campi, selezionare la casella *Includi campi non raccomandati* nella tabella sopra la tabella; in questo modo, tali campi saranno visualizzati in fondo alla tabella.

La tabella contiene le seguenti colonne:

- **Versione da usare.** Questa colonna visualizza un elenco a discesa che controlla se un campo verrà utilizzato a valle e se usare le trasformazioni suggerite. Per impostazione predefinita, l'elenco rispecchia le raccomandazioni.

Per i predittori ordinari trasformati, l'elenco a discesa riporta tre opzioni: *Trasformata*, *Originale* e *Non utilizzare*.

Per i predittori ordinari non trasformati, le opzioni sono: *Originale* e *Non utilizzare*.

Per i campi derivati data/ora e i predittori creati, le opzioni sono: *Trasformata* e *Non utilizzare*.

Per i campi data originali l'elenco a discesa è disattivato e impostato su *Non utilizzare*.

Nota: per i predittori che hanno una versione originale e una trasformata, il passaggio dalla versione *Originale* a quella *Trasformata* aggiorna automaticamente le impostazioni *Livello di misurazione* e *Potere predittivo*.

- **Nome.** Ogni nome di campo è un collegamento. Fare clic su un nome per visualizzare ulteriori informazioni sul campo nella visualizzazione collegata. [Per ulteriori informazioni, vedere l'argomento *Dettagli campo* a pag. 118.](#)
- **Livello di misurazione.** In questa colonna è visualizzata l'icona che rappresenta il tipo di dati; passare il puntatore del mouse sopra l'icona per visualizzare un'etichetta (continuo, ordinale, nominale e così via) che descrive i dati.
- **Potere predittivo.** Il potere predittivo è visualizzato solo per i campi consigliati da ADP. Questa colonna non è visualizzata se non è stato definito un obiettivo. Il potere predittivo è compreso tra 0 e 1, e i valori più elevati indicano predittori "migliori". In generale, il potere predittivo è utile per confrontare i predittori all'interno di un'analisi ADP, ma non deve essere effettuato alcun confronto tra i valori del potere predittivo di analisi diverse.

Riepilogo delle azioni

Figura 4-13
Riepilogo delle azioni

Riepilogo delle azioni

Azione
Campi di testo
Funzioni di data e ora
Esame funzioni
Tipo di verifica
Valori anomali
Valori mancanti
Obiettivo
Funzioni categoriali
Funzioni continue

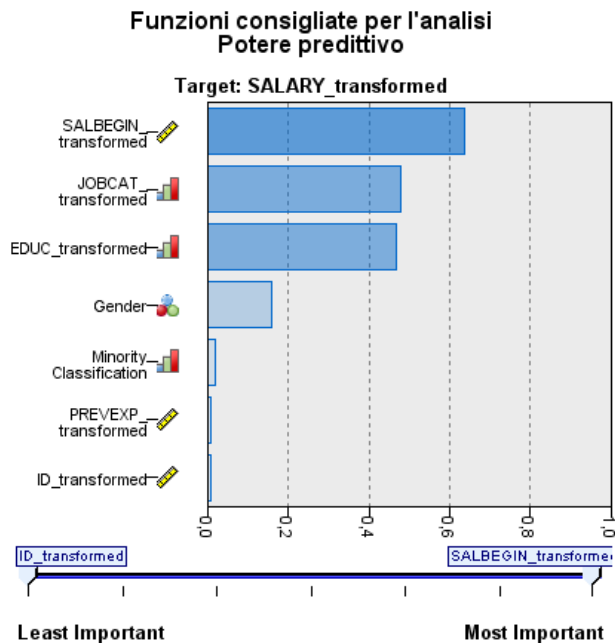
Per ogni azione svolta dalla preparazione automatica dati, i predittori di input vengono trasformati e/o eliminati tramite filtri; i campi che sopravvivono a un'azione vengono utilizzati nella successiva. I campi che superano tutti i passaggi sono quelli di cui si consiglia l'utilizzo nella modellazione, mentre gli input a predittori trasformati e creati vengono esclusi.

Il Riepilogo delle azioni è una semplice tabella che elenca le azioni di elaborazione svolte da ADP. Se vi è un'Azione sottolineata, farvi clic sopra per visualizzare ulteriori dettagli sulle azioni intraprese in una visualizzazione collegata. [Per ulteriori informazioni, vedere l'argomento Dettagli dell'azione a pag. 120.](#)

Nota: sono visualizzate solo le versioni originali e trasformate definitive di ogni campo, non quelle intermedie utilizzate durante l'analisi.

Potere predittivo

Figura 4-14
Potere predittivo



Visualizzato per impostazione predefinita la prima volta che viene eseguita l'analisi o quando si seleziona Predittori consigliati per l'uso nell'analisi nella visualizzazione principale Riepilogo elaborazione campi, il grafico mostra il potere predittivo dei predittori consigliati. I campi sono ordinati in base al potere predittivo, a partire dal campo con il valore più elevato.

Per le versioni trasformate dei predittori ordinari, il nome del campo rispecchia il suffisso scelto nel riquadro Nomi dei campi della scheda Impostazioni; ad esempio: *_transformed*.










Dopo i singoli nomi dei campi sono visualizzate le icone del livello di misurazione.

Il potere predittivo di ogni predittore consigliato è calcolato da un modello di regressione lineare o naive Bayes, a seconda che l'obiettivo sia continuo o categoriale.

Tabella Campi

Figura 4-15
Tabella campi

Funzioni di input

Nome	Tipo
ID	 Continuo
GENDER	 Insieme
BDATE	 Continuo
EDUC	 Insieme ordinato
JOBCAT	 Insieme ordinato
SALBEGIN	 Continuo
JOBTIME	 Continuo
PREVEXP	 Continuo
MINORITY	 Insieme ordinato

Visualizzata quando si fa clic su Obiettivo, Predittori o Predittori non utilizzati nella visualizzazione principale Riepilogo elaborazione campi, la visualizzazione Tabella campi mostra una semplice tabella con un elenco delle funzioni pertinenti.

La tabella contiene due colonne:

- **Nome.** Il nome del predittore.

Per gli obiettivi viene utilizzato il nome o l'etichetta originale del campo, anche se l'obiettivo è stato trasformato.

Per le versioni trasformate dei predittori ordinari, il nome rispecchia il suffisso scelto nel riquadro Nomi dei campi della scheda Impostazioni; ad esempio: *_transformed*.

Per i campi derivati da date e ore, viene utilizzato il nome della versione trasformata definitiva, ad esempio: *bdate_years*.

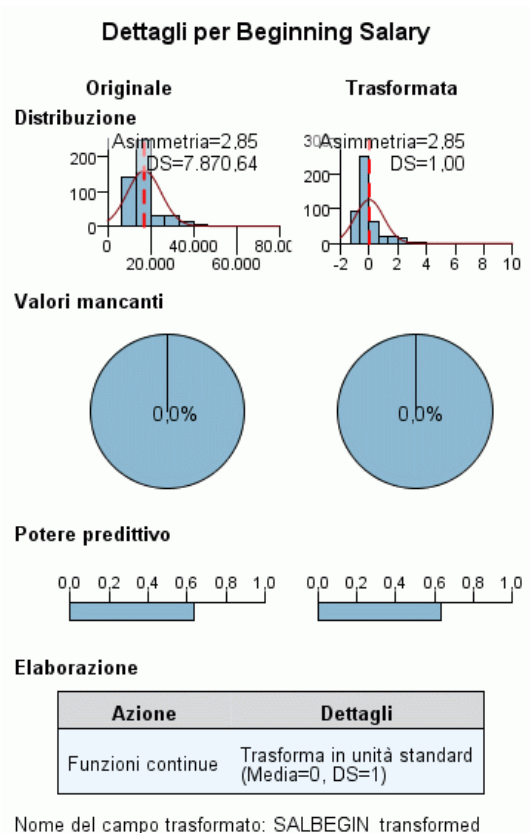
Per i predittori creati viene utilizzato il nome del predittore creato, ad esempio: *Predictor1*.

- **Livello di misurazione.** Visualizza l'icona che rappresenta il tipo di dati.

Per l'Obiettivo, il Livello di misurazione rispecchia sempre la versione trasformata se l'obiettivo è stato trasformato, ad esempio passando da ordinale (insieme ordinato) a continuo (intervallo, scala) o viceversa.

Dettagli campo

Figura 4-16
Dettagli campo



Visualizzata quando si fa clic su un Nome nella visualizzazione principale Campi, la visualizzazione Dettagli campo contiene la distribuzione, i valori mancanti e gli eventuali grafici del potere predittivo per il campo selezionato. Sono inoltre visualizzati la cronologia di elaborazione del campo e il nome del campo trasformato (se applicabile).

Per ogni grafico impostato sono visualizzate due versioni affiancate per confrontare il campo con e senza l'applicazione delle trasformazioni; se non esiste una versione trasformata del campo, il grafico viene visualizzato solo per la versione originale. Per i campi data o ora derivati e i predittori creati, i grafici sono visualizzati solo per il nuovo predittore.

Nota: se un campo viene escluso perché ha troppe categorie, viene visualizzata solo la cronologia di elaborazione.

Grafico della distribuzione

La distribuzione dei campi continui è rappresentata sotto forma di istogramma a cui è sovrapposta una curva normale e con una linea verticale di riferimento per il valore medio; i campi categoriali sono visualizzati sotto forma di grafico a barre.

Gli istogrammi sono dotati di etichette che mostrano la deviazione standard e l'asimmetria; tuttavia, l'asimmetria non è visualizzata se il numero massimo dei valori è 2 o se la varianza del campo originale è inferiore a 10-20.

Passare il puntatore del mouse sul grafico per visualizzare la media degli istogrammi o il numero e la percentuale sul totale dei record per le per le categorie dei grafici a barre.

Grafico dei valori mancanti

I grafici a torta confrontano la percentuale dei valori mancanti con e senza l'applicazione delle trasformazioni; le etichette del grafico mostrano la percentuale.

Se ADP ha utilizzato la gestione dei valori mancanti, il grafico a torta dopo la trasformazione comprende anche il valore di sostituzione (cioè il valore utilizzato al posto di quelli mancanti) sotto forma di etichetta.

Passare il puntatore del mouse sopra il grafico per visualizzare il numero dei valori mancanti e la percentuale del numero totale di record.

Grafico del potere predittivo

Per i campi consigliati, i grafici a barre mostrano il potere predittivo prima e dopo la trasformazione. Se l'obiettivo è stato trasformato, il potere predittivo calcolato è relativo all'obiettivo trasformato.

Nota: i grafici del potere predittivo non sono visualizzati se non è definito un obiettivo o se si fa clic sull'obiettivo nel riquadro della visualizzazione principale.

Passare il puntatore del mouse sul grafico per visualizzare il valore del potere predittivo.

Tabella Cronologia elaborazione

La tabella mostra come è stata derivata la versione trasformata di un campo. Le azioni intraprese da ADP sono elencate nell'ordine in cui sono state eseguite; tuttavia, per alcuni passaggi è possibile che siano state intraprese più azioni per un determinato campo.

Nota: questa tabella non è visualizzata per i campi che non sono stati trasformati.

Le informazioni della tabella sono suddivise in due o tre colonne:

- **Azione.** Il nome dell'azione. Ad esempio, Predittori continui. [Per ulteriori informazioni, vedere l'argomento Dettagli dell'azione a pag. 120.](#)
- **Dettagli.** L'elenco delle procedure eseguite. Ad esempio, Trasforma in unità standard.
- **Funzione.** Visualizzata solo per i predittori creati, mostra la combinazione lineare dei campi di input, ad esempio $0,06*age + 1,21*height$.

Dettagli dell'azione

Figura 4-17
Analisi ADP - Dettagli dell'azione

Passaggio 9: Funzioni continue

Trasformazione	Numero di funzioni	Criteri	
		Media	DS
Trasforma in unità standard	5	0	1

Creazione spazio di funzioni	C
Funzioni create	0
Funzioni escluse a causa di una scarsa associazione all'obiettivo	2
Funzioni escluse perché costanti dopo la discretizzazione	0

Visualizzata quando si seleziona un'Azione sottolineata nella visualizzazione principale Riepilogo delle azioni, la visualizzazione collegata Dettagli azione mostra informazioni generali e specifiche di un'azione per ogni fase di elaborazione effettuata; i dettagli relativi a un'azione specifica sono visualizzati per primi.

Per ogni azione, la descrizione viene utilizzata come titolo nella parte superiore della visualizzazione collegata. I dettagli specifici delle singole azioni sono visualizzati sotto al titolo e possono comprendere il numero di predittori derivati, i campi riformulati, le trasformazioni dell'obiettivo, le categorie unite o riordinate e i predittori creati o esclusi.

A ogni azione, il numero di predittori utilizzato nell'elaborazione può variare, ad esempio a causa dell'esclusione o dell'unione di predittori.

Nota: se un'azione è stata disattivata o se non è stato specificato un obiettivo, quando si fa clic sull'azione nella visualizzazione principale Riepilogo delle azioni viene visualizzato un messaggio di errore al posto dei dettagli dell'azione.

Le possibili azioni sono nove, ma non tutte sono necessariamente attive per ogni analisi.

Tabella Campi di testo

La tabella mostra il numero di:

- Valori vuoti finali tagliati.
- Predittori esclusi dall'analisi.

Tabella Predittori data e ora

La tabella mostra il numero di:

- Durate derivate da predittori di data e ora.
- Elementi di data e ora.
- Predittori di data e ora derivati in totale.

La data o l'ora di riferimento è visualizzata come nota a piè di pagina se sono state calcolate delle durate delle date.

Tabella Screening dei predittori

La tabella mostra il numero dei seguenti predittori esclusi dall'elaborazione:

- Costanti.
- Predittori con troppi valori mancanti.
- Predittori con troppi casi in un'unica categoria.
- Campi nominali (insiemi) con troppe categorie.
- Predittori esclusi in totale.

Tabella Controllo livello di misurazione

La tabella mostra il numero dei campi riformulati, suddivisi in:

- Campi ordinali (insiemi ordinati) riformulati come campi continui.
- Campi continui riformulati come campi ordinali.
- Numero totale riformulato.

Se nessun campo di input (obiettivo o predittore) è continuo o ordinale, il totale viene visualizzato come un piè di pagina.

Tabella Valori anomali

La tabella mostra il conteggio delle modalità con cui sono stati gestiti i valori anomali.

- Indica o il numero di campi continui per i quali sono stati trovati e tagliati dei valori anomali, oppure il numero di campi continui per i quali sono stati trovati e impostati come mancanti dei valori anomali, a seconda delle impostazioni nel riquadro Prepara input e obiettivo nella scheda Impostazioni.
- Il numero dei campi continui esclusi perché costanti dopo la gestione dei valori anomali.

Un piè di pagina indica il punto di interruzione dei valori anomali, mentre viene mostrato un altro piè di pagina se nessun campo di input (obiettivo o predittore) è continuo.

Tabella Valori mancanti

La tabella mostra il numero dei campi i cui valori mancanti sono stati sostituiti, suddivisi in:

- Obiettivo. Se non viene specificato alcun obiettivo, questa riga non è visualizzata.
- Predittori. A sua volta suddiviso nel numero di nominali (insieme), ordinali (insieme ordinato) e continui.
- Il numero totale di valori mancanti sostituiti.

Tabella Obiettivo

La tabella indica se l'obiettivo è stato trasformato, illustrato come:

- Trasformazione di Box-Cox alla normalità. Questo valore è ulteriormente suddiviso in colonne che mostrano i criteri specificati (media e deviazione standard) e il valore Lambda.
- Categorie obiettivo riordinate per migliorare la stabilità.

Tabella Predittori categoriali

La tabella mostra il numero di predittori categoriali:

- Le cui categorie sono state riordinate dalla più bassa alla più alta per migliorare la stabilità.
- Le cui categorie sono state unite per aumentare al massimo l'associazione all'obiettivo.
- Le cui categorie sono state unite per gestire le categorie poco dense.
- Esclusi a causa di una scarsa associazione all'obiettivo.
- Esclusi perché erano costanti dopo l'unione.

Se non erano presenti predittori categoriali viene visualizzata una nota a piè di pagina.

Tabella Predittori continui

In questo caso le tabelle sono due. La prima visualizza uno dei seguenti numeri di trasformazioni:

- Valori dei predittori trasformati in unità standard. Sono visualizzati inoltre il numero dei predittori trasformati, la media specificata e la deviazione standard.
- Valori dei predittori associati a un intervallo comune. Sono visualizzati inoltre il numero di predittori trasformati mediante una trasformazione Min/Max e i valori minimi e massimi specificati.
- Valori di predittori categorizzati e il numero di predittori categorizzati.

La seconda tabella riporta i dettagli di creazione dello spazio dei predittori, visualizzati sotto forma di numero di predittori:

- Creati.
- Esclusi a causa di una scarsa associazione all'obiettivo.
- Esclusi perché erano costanti dopo la categorizzazione.
- Esclusi perché erano costanti dopo la creazione.

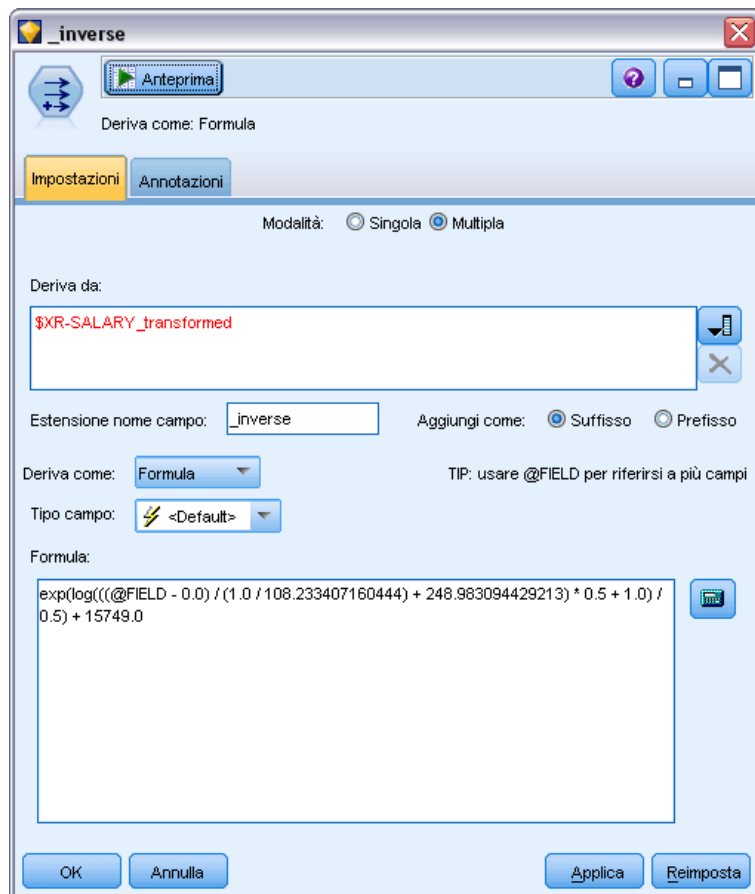
Se l'input non includeva predittori continui viene visualizzata una nota a piè di pagina.

Generazione di un nodo Nuovo campo

Quando si genera un nodo Nuovo campo, esso applica la trasformazione inversa dell'obiettivo al campo punteggio. Per impostazione predefinita, il nodo immette il nome del campo punteggio che sarebbe generato da una funzione di modellazione automatica (quale Classificatore automatico o Numerico automatico) o dal nodo Classificatore binario. Se è stato trasformato un obiettivo scala (intervallo), il campo punteggio è visualizzato nelle unità trasformate, ad esempio $\log(\$)$ anziché $\$$. Per interpretare e utilizzare questi risultati è necessario riconvertire il valore previsto nella scala originale.

Nota: il nodo Nuovo campo può essere generato solo quando il nodo ADP contiene un'analisi che ridimensiona l'obiettivo intervallo (ovvero quando è selezionata la modifica della scala di Box-Cox nel riquadro Prepara input e obiettivo). Non è possibile generare un nodo Nuovo campo se l'obiettivo non è un intervallo o se non è selezionata la modifica della scala di Box-Cox.

Figura 4-18
Nodo Nuovo campo generato con il nodo Preparazione automatica dati



Il nodo Nuovo campo viene creato in modalità multipla e utilizza @FIELD nell'espressione; è quindi possibile aggiungere l'obiettivo trasformato, se necessario. Si considerino ad esempio i seguenti dati:

- Nome campo obiettivo: risposta
- Nome campo obiettivo trasformato: response_transformed
- Nome campo punteggio: \$XR-response_transformed

Il nodo Nuovo campo creerebbe un nuovo campo: \$XR-response_transformed_inverse.

Nota: se non si utilizza una funzione di modellazione automatica o un nodo Classificatore binario, sarà necessario modificare il nodo Nuovo campo per trasformare il campo punteggio corretto per il proprio modello.

Obiettivi continui normalizzati

Per impostazione predefinita, se si seleziona la casella di controllo Ridimensiona obiettivo continuo con trasformazione di Box-Cox nel riquadro Prepara input e obiettivo, l'obiettivo viene trasformato e viene creato un nuovo campo che diventerà l'obiettivo per la creazione del proprio modello. Ad esempio, se l'obiettivo originale era *response*, il nuovo obiettivo sarà *response_transformed*; i modelli a valle del nodo ADP acquisiranno automaticamente questo nuovo obiettivo.

A seconda dell'obiettivo originale, però, questo può causare dei problemi. Ad esempio, se l'obiettivo era *Età*, i valori del nuovo obiettivo non saranno *Anni*, ma una versione trasformata di *Anni*. Ciò significa che non è possibile esaminare i punteggi e interpretarli, poiché non sono espressi in unità riconoscibili. In questo caso si può applicare una trasformazione inversa che riporterà le unità trasformate al formato che avrebbero dovuto avere. A questo scopo, eseguire le operazioni seguenti:

- ▶ Dopo avere fatto clic su Analizza dati per eseguire l'analisi ADP, selezionare *Nodo Nuovo campo* dal menu *Genera*.
- ▶ Collocare il nodo Nuovo campo dopo l'insieme di modelli nell'area di disegno del modello.

Il nodo Nuovo campo ripristinerà le dimensioni originali del campo punteggio in modo tale che la previsione sia espressa nei valori *Anni* originali.

Per impostazione predefinita, il nodo Nuovo campo trasforma il campo punteggio generato da una funzione di modellazione automatica o da un modello Classificatore binario. Se si crea un modello singolo è necessario modificare il nodo Nuovo campo per derivare il campo dal campo punteggio effettivo. Se si desidera valutare il modello, occorre aggiungere l'obiettivo trasformato al campo Deriva da del nodo Nuovo campo. In questo modo si applica la stessa trasformazione inversa all'obiettivo e tutti i nodi Valutazione o Analisi a valle useranno i dati trasformati correttamente a condizione che i nodi vengano modificati per utilizzare nomi di campi anziché metadati.

Inoltre, se si desidera ripristinare il nome originale, è possibile utilizzare un nodo Filtro per rimuovere il campo obiettivo originale, se ancora presente, e rinominare i campi obiettivo e punteggio.

Nodo Tipo

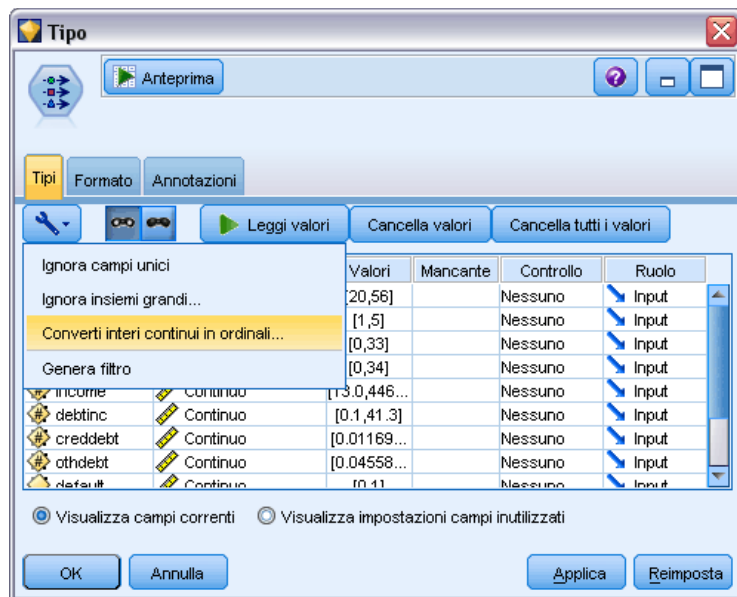
È necessario specificare le proprietà dei campi in un nodo di input o in un nodo Tipo distinto. Questa funzione è analoga in entrambi i nodi. Sono disponibili le seguenti proprietà:

- **Campo.** Fare doppio clic sul nome di un campo per specificare valore ed etichette di campo per i dati in IBM® SPSS® Modeler. Per esempio, qui è possibile visualizzare o modificare i metadati dei campi importati da IBM® SPSS® Statistics. Analogamente, è possibile creare nuove etichette per campi e loro valori. Le etichette specificate qui vengono visualizzate in SPSS Modeler in base alle selezioni effettuate nella finestra di dialogo Proprietà stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Misurazione.** Livello di misurazione utilizzato per descrivere le caratteristiche dei dati in un campo specifico. Se tutti i dettagli di un campo sono noti, il campo viene definito **completamente istanziato**. [Per ulteriori informazioni, vedere l'argomento Livelli di misurazione a pag. 127.](#)

Nota: Il livello di misurazione è diverso dalla classe di archiviazione di un campo, che indica se i dati sono archiviati come stringhe, numeri interi, numeri reali, date, ore o timestamp.

- **Valori.** Questa colonna consente di specificare le opzioni di lettura dei valori dei dati dall'insieme di dati. L'opzione Specifica permette invece di specificare i livelli di misurazione e i valori in una finestra di dialogo a parte. È anche possibile scegliere di passare i campi senza leggerne i valori. [Per ulteriori informazioni, vedere l'argomento Valori dei dati a pag. 131.](#)
- **Mancante/i.** Utilizzata per specificare come vengono gestiti i valori mancanti del campo. [Per ulteriori informazioni, vedere l'argomento Definizione dei valori mancanti a pag. 137.](#)
- **Controllo.** In questa colonna è possibile impostare opzioni per assicurarsi che i valori dei campi siano inclusi nei valori o nell'intervallo specificato. [Per ulteriori informazioni, vedere l'argomento Controllo dei valori del tipo a pag. 137.](#)
- **Ruolo.** Proprietà utilizzata per indicare ai nodi Modelli se i campi verranno trattati come campi di Input (campi predittori) o Obiettivo (campi previsti) nel processo di apprendimento automatico. Entrambi e Nessuno sono altri ruoli possibili, oltre a Partizione, che indica un campo utilizzato per eseguire la partizione dei record in campioni separati per addestramento, test e validazione. Il valore Suddivisione specifica che i modelli separati verranno costruiti per ciascun valore possibile del campo. [Per ulteriori informazioni, vedere l'argomento Impostazione del ruolo del campo a pag. 138.](#)

Figura 4-19
Opzioni del nodo Tipo



La finestra del nodo Tipo permette di specificare ulteriori opzioni:

- Quando un nodo Tipo è stato istanziato (tramite specifiche manuali, lettura di valori o esecuzione dello stream) è possibile utilizzare il pulsante del menu di strumenti per scegliere Ignora campi unici. Questa opzione consente di ignorare i campi contenenti un solo valore.
- Mediante il pulsante del menu di strumenti è inoltre possibile scegliere Ignora insiemi grandi per un nodo Tipo istanziato. In questo modo verranno automaticamente ignorati insiemi con un numero elevato di membri.
- Mediante il pulsante del menu Strumenti è possibile scegliere Converti interi continui in ordinali per un nodo Tipo istanziato. [Per ulteriori informazioni, vedere l'argomento Conversione di dati continui a pag. 129.](#)
- Mediante il pulsante del menu di strumenti è infine possibile generare un nodo Filtro per scartare i campi selezionati.
- Utilizzando i pulsanti con l'icona degli occhiali, è possibile specificare Leggi o Passa come impostazione di default per tutti i campi. La scheda Tipi nel nodo di input passa i campi per default, mentre il nodo Tipo legge i valori per default.
- Tramite il pulsante Cancella valori è possibile annullare le modifiche apportate in questo nodo ai valori dei campi (valori non ereditati) e leggere di nuovo i valori mediante operazioni a monte. Questa opzione è utile per reimpostare le modifiche apportate a monte di campi specifici.
- Tramite il pulsante Cancella tutti i valori è possibile reimpostare i valori per **tutti** i campi letti nel nodo. Questa opzione consente di impostare la colonna *Valori* su **Leggi** per tutti i campi, pertanto risulta utile per reimpostare i valori di tutti i campi e leggere nuovamente valori e tipi mediante operazioni a monte.

- Mediante il menu di scelta rapida è possibile scegliere Copia per copiare gli attributi di un campo per un altro campo. [Per ulteriori informazioni, vedere l'argomento Copia degli attributi del tipo a pag. 140.](#)
- Mediante l'opzione Visualizza impostazioni campi inutilizzati è possibile visualizzare le impostazioni del tipo relative ai campi non più inclusi nei dati o non più connessi al nodo Tipo corrente. Questa opzione è utile quando si riutilizza un nodo Tipo per insiemi di dati che sono stati modificati.

Livelli di misurazione

Il livello di misurazione (già “tipo di dati” o “tipo di utilizzo”) descrive l'utilizzo dei campi di dati in IBM® SPSS® Modeler. Il livello di misurazione si può specificare nella scheda Tipi di un nodo Tipo o di input. Per esempio, è possibile impostare il livello di misurazione per un campo intero con valori 1 e 0 su *Flag*. In genere, questo indica che 1=*Vero* e 0=*Falso*.

Classe di archiviazione e misurazione. Si noti che il livello di misurazione di un campo è diverso dal tipo di archiviazione del campo, in quanto l'archiviazione indica se i dati vengono archiviati come stringhe, numeri interi, reali, date, ore o timestamp. Mentre i tipi di dati possono essere modificati in qualsiasi punto di uno stream mediante un nodo Tipo, l'archiviazione deve essere stabilita in fase di input quando si leggono i dati in SPSS Modeler (anche se in seguito è possibile modificarla tramite una funzione di conversione). [Per ulteriori informazioni, vedere l'argomento Impostazione dell'archiviazione e della formattazione dei campi in il capitolo 2 a pag. 27.](#)

Sono disponibili i seguenti livelli di misurazione:

- **Default.** I dati di cui non si conosce il tipo di archiviazione né i valori (per esempio perché non sono ancora stati letti) sono visualizzati come <Default>.
- **Continuo.** Utilizzato per descrivere valori numerici, per esempio un intervallo 0–100 o 0.75–1.25. Un valore continuo può essere un numero intero, un numero reale o un valore di data/ora.
- **Categoriali.** Utilizzato per valori stringa nel caso in cui non sia noto un numero esatto di valori distinti. Questo è un tipo di dati **non istanziato**, ovvero, tutte le informazioni possibili sull'archiviazione e l'utilizzo dei dati non sono ancora note. Dopo la lettura dei dati, il livello di misurazione sarà definito come *Flag*, *Nominale* o *Senza tipo*, a seconda del numero massimo di membri per i campi nominali specificato nella finestra di dialogo Proprietà stream.
- **Flag.** Utilizzato per i dati con due valori distinti che indicano la presenza o l'assenza di una caratteristica, per esempio true e false, Yes e No o 0 e 1. I valori utilizzati possono variare, ma uno deve sempre essere designato come valore “true” e l'altro come valore “false”. I dati possono essere rappresentati come testo, numero intero, numero reale, data, ora o timestamp.
- **Nominale.** Utilizzato per descrivere i dati con più valori distinti, ognuno dei quali viene trattato come membro di un insieme, per esempio small/medium/large. Ai dati nominali può essere assegnata qualsiasi classe di archiviazione: numerica, stringa o data/ora. Si noti che, impostando il livello di misurazione su *Nominale*, i valori non vengono automaticamente modificati in tipo di archiviazione stringa.

- **Ordinale.** Utilizzato per descrivere i dati con più valori distinti che hanno un ordine intrinseco. Per esempio, le categorie di stipendio o le classificazioni del livello di soddisfazione possono essere immesse come dati ordinali. L'ordine è definito dall'ordine naturale degli elementi di dati. Per esempio, 1, 3, 5 è l'ordine di default per un insieme di interi, mentre HIGH, LOW, NORMAL (ordine alfabetico crescente) è l'ordine per un insieme di stringhe. Il livello di misurazione ordinale consente di definire un insieme di dati categoriali come dati ordinali per la visualizzazione, la creazione di modelli e l'esportazione ad altre applicazioni (per esempio IBM® SPSS® Statistics) che sono in grado di riconoscere i dati ordinali come un tipo distinto. È possibile utilizzare un campo ordinale in tutti i casi in cui è possibile utilizzare un campo nominale. Inoltre, è possibile definire come ordinali i campi con qualsiasi tipo di classe di archiviazione (reale, intero, stringa, data, ora e così via).
- **Senza tipo.** Utilizzato per dati che non rientrano in nessuno dei tipi sopra riportati, per i campi con un solo valore o per dati nominali in cui l'insieme ha un numero di membri superiore al massimo definito. È utile anche nei casi in cui il livello di misurazione potrebbe corrispondere a un insieme con un numero eccessivo di membri (per esempio un numero di account). Quando si seleziona Senza tipo per un campo, il ruolo viene impostato automaticamente su Nessuno, con ID record come unica alternativa. La dimensione massima di default per gli insiemi è 250 valori univoci. Questo numero può essere adeguato o disattivato nella scheda Opzioni della finestra di dialogo Proprietà stream, cui si può accedere dal menu Strumenti.

È possibile specificare i livelli di misurazione manualmente, oppure consentendo al software di leggere i dati e di stabilirne il livello di misurazione in base ai valori letti.

In alternativa, quando si dispone di diversi campi di dati continui che devono essere trattati come dati categoriali, è possibile scegliere un'opzione per convertirli. [Per ulteriori informazioni, vedere l'argomento Conversione di dati continui a pag. 129.](#)

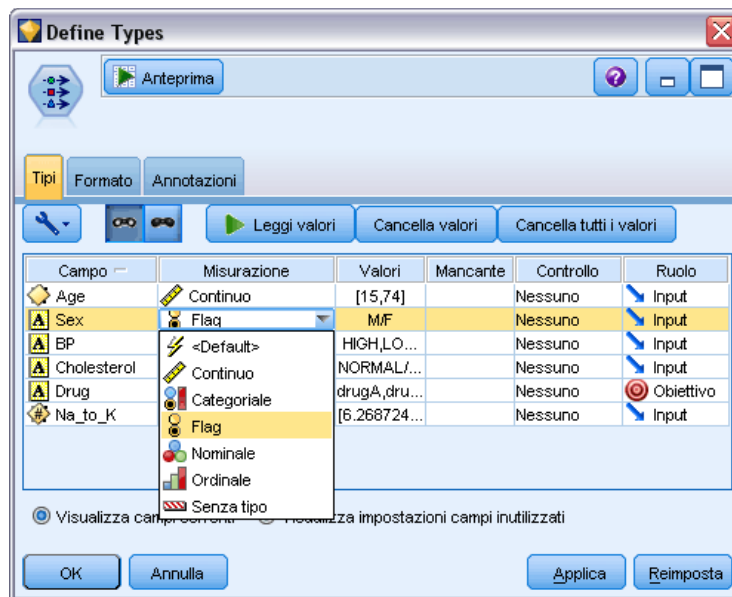
Per utilizzare l'impostazione automatica

- ▶ In un nodo Tipo o nella scheda Tipi di un nodo di input, impostare la colonna *Valori* su <Leggi> nei campi desiderati. I metadati saranno così disponibili per tutti i nodi a valle. Per impostare rapidamente tutti i campi su <Leggi> o <Passa>, utilizzare i pulsanti con l'icona degli occhiali nella finestra di dialogo.
- ▶ Per leggere subito i valori dalla sorgente dati, fare clic su Leggi valori.

Per impostare manualmente il livello di misurazione di un campo

- ▶ Selezionare un campo nella tabella.
- ▶ Dall'elenco a discesa nella colonna *Misurazione*, selezionare un livello di misurazione per il campo.
- ▶ In alternativa, utilizzare Ctrl-A o Ctrl-clic per selezionare più campi prima di selezionare un livello di misurazione dall'elenco a discesa.

Figura 4-20
Impostazione manuale dei livelli di misurazione



Conversione di dati continui

Il fatto di considerare dati categoriali come continui può avere notevoli ripercussioni sulla qualità di un modello, soprattutto se si tratta del campo obiettivo; per esempio, può determinare la creazione di un modello di regressione anziché di un modello binario. Per evitare che questo accada è possibile convertire intervalli di numeri interi in tipi categoriali come *Ordinale* o *Flag*.

- Dal pulsante del Menu di strumenti per operazioni e generazioni (con il simbolo dello strumento), selezionare Converti interi continui in ordinali. Viene visualizzata la finestra di dialogo dei valori di conversione.

Figura 4-21
Finestra di dialogo dei valori di conversione



- Specificare le dimensioni dell'intervallo che verrà automaticamente convertito; questa operazione si applica a tutti gli intervalli inferiori e uguali alle dimensioni immesse.
- Fare clic su OK. Gli intervalli interessati vengono convertiti in *Flag* o *Ordinale* e visualizzati nella scheda Tipi del nodo Tipo.

Risultati della conversione

- Quando un campo *Continuo* con tipo di archiviazione numero intero viene modificato in *Ordinale*, i valori inferiore e superiore vengono espansi in modo da includere tutti i valori interi dal valore più basso a quello più alto. Per esempio, se l'intervallo è 1, 5, l'insieme di valori è 1, 2, 3, 4, 5.
- Se il campo *Continuo* viene trasformato in *Flag*, i valori inferiore e superiore diventano i valori falso e vero del campo Flag.

Che cos'è l'istanziamento?

Per **istanziamento** si intende il processo di lettura o definizione di informazioni quali il tipo di archiviazione e i valori per un campo dati. Per ottimizzare le risorse di sistema, è l'utente a guidare il processo di istanziamento: i valori vengono letti dal software quando l'utente specifica le opzioni nella scheda Tipi di un nodo di input o esegue i dati mediante un nodo Tipo.

- I dati con tipi sconosciuti vengono anche detti **non istanziati**. I dati il cui tipo di archiviazione e i cui valori sono sconosciuti vengono visualizzati nella colonna *Misurazione* della scheda Tipi come <Default>.
- Quando si dispone di informazioni sul tipo di archiviazione di un campo, per esempio stringa o numerico, i dati vengono detti **parzialmente istanziati**. Catoriale o Continuo sono livelli di misurazione parzialmente istanziati. Catoriale specifica, per esempio, che il campo è simbolico, ma non se sia nominale, ordinale o flag.
- Quando tutti i dettagli di un tipo sono noti, compresi i valori, in questa colonna viene visualizzato un livello di misurazione **completamente istanziato**: nominale, ordinale, flag o continuo. *Nota*: il tipo *continuo* è utilizzato sia per i campi di dati parzialmente istanziati sia per quelli completamente istanziati. I dati continui possono essere numeri interi o numeri reali.

Durante l'esecuzione di uno stream di dati con un nodo Tipo, i tipi non istanziati vengono parzialmente istanziati in base ai valori dei dati iniziali. Dopo essere passati per il nodo, tutti i dati vengono completamente istanziati, a meno che i valori non siano stati impostati su <Passa>. Se l'esecuzione viene interrotta, i dati rimarranno parzialmente istanziati. Una volta istanziata la scheda Tipi, i valori di un campo sono statici in quel punto dello stream. Ciò significa che le modifiche a monte non influenzeranno i valori di un particolare campo, nemmeno se si riesegue lo stream. Per cambiare o aggiornare i valori in base ai nuovi dati o a ulteriori manipolazioni, è necessario modificarli nella scheda Tipi o impostare il valore del campo desiderato su <Leggi> o <Leggi +>.

Informazioni sull'istanziamento

In genere, se l'insieme di dati non è di grandi dimensioni e non si prevede di aggiungere campi nello stream in seguito, l'istanziamento a livello del nodo di input risulta il metodo più vantaggioso. L'istanziamento in un nodo Tipo distinto risulta tuttavia utile nei seguenti casi:

- L'insieme di dati è di grandi dimensioni e lo stream filtra un sottoinsieme precedente al nodo Tipo.
- I dati sono stati filtrati nello stream.

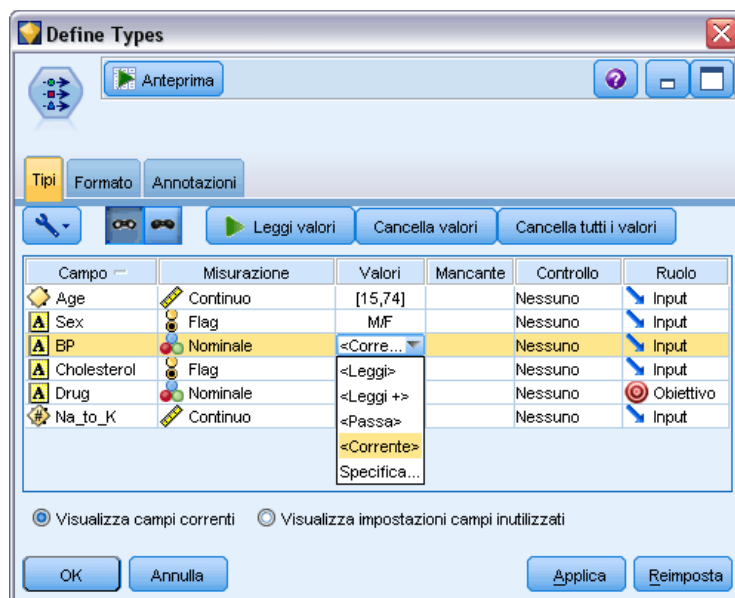
- I dati sono stati uniti o accodati nello stream.
- Durante l'elaborazione vengono derivati nuovi campi di dati.

Valori dei dati

Tramite la colonna *Valori* della scheda *Tipi* è possibile leggere automaticamente i valori dai dati o specificare livelli di misurazione e valori in una finestra di dialogo separata.

Figura 4-22

Selezione dei metodi di lettura, passaggio o definizione dei valori dei dati



Le opzioni disponibili da questo elenco a discesa forniscono le seguenti istruzioni per l'impostazione automatica:

Opzione	Funzione
<Read>	Selezionare questa opzione per leggere i valori durante l'esecuzione del nodo.
<Leggi+>	I dati vengono letti e accodati ai dati correnti (se presenti).
<Pass>	Non verrà letto alcun dato.
<Current>	Vengono mantenuti i valori dei dati correnti.
Specifica...	Viene aperta una finestra di dialogo separata in cui specificare le opzioni su valori e livelli di misurazione.

Se si esegue un nodo *Tipo* o si fa clic su *Leggi valori*, i valori vengono impostati automaticamente e letti dalla sorgente dati in base alla selezione. I valori possono essere specificati anche manualmente tramite l'opzione *Specifica* o facendo doppio clic su una cella nella colonna *Campo*.

Dopo aver apportato le modifiche per i campi nel nodo Tipo, è possibile utilizzare i seguenti pulsanti nella barra degli strumenti della finestra di dialogo per reimpostare le informazioni sui valori:

- Tramite il pulsante **Cancella valori** è possibile annullare le modifiche apportate in questo nodo ai valori dei campi (valori non ereditati) e leggere di nuovo i valori mediante operazioni a monte. Questa opzione è utile per reimpostare le modifiche apportate a monte di campi specifici.
- Tramite il pulsante **Cancella tutti i valori** è possibile reimpostare i valori per **tutti** i campi letti nel nodo. Questa opzione consente di impostare la colonna *Valori* su **Leggi** per tutti i campi, pertanto risulta utile per reimpostare i valori di tutti i campi e leggere nuovamente valori e livelli di misurazione mediante operazioni a monte.

Utilizzo della finestra di dialogo Valori

Se si fa clic sulla colonna *Valori* o *Mancante* nella scheda Tipi, viene visualizzato un elenco a discesa di valori predefiniti. Se si sceglie l'opzione *Specifica* nell'elenco, verrà aperta una finestra di dialogo distinta nella quale è possibile impostare le opzioni per la lettura, la definizione, l'assegnazione delle etichette e la gestione dei valori del campo selezionato.

Figura 4-23

Impostazione delle opzioni relative ai valori dei dati

Valori	Etichette
drugA	
drugB	
drugC	
drugX	
drugY	

Molti controlli sono comuni a tutti i tipi di dati. Tali controlli comuni vengono illustrati di seguito.

Misurazione. Visualizza il livello di misurazione selezionato. È possibile modificare l'impostazione per indicare il modo in cui si intende utilizzare i dati. Se un campo chiamato *giorno_della_settimana* contiene, per esempio, numeri che rappresentano singoli giorni, è possibile che si desideri modificare questa impostazione in dati nominali per creare un nodo di distribuzione che esamini ciascuna categoria individualmente.

Classe archiviazione. Visualizza il tipo di archiviazione, se noto. Il livello di misurazione prescelto non influisce sui tipi di archiviazione. Per modificare il tipo di archiviazione è possibile utilizzare la scheda Dati nei nodi di input Testo fisso e Testo variabile o una funzione di conversione in un nodo Riempimento.

Campo del modello. Per i campi generati in seguito al calcolo del punteggio di un insieme di modelli è possibile visualizzare anche i dettagli del campo del modello. Essi comprendono il nome del campo obiettivo, nonché il ruolo del campo nella modellazione (valore previsto, probabilità, propensione, ecc.).

Valori. Selezionare un metodo di lettura dei valori per il campo selezionato. Le selezioni eseguite in questa finestra sovrascrivono quelle eseguite in precedenza dalla colonna *Valori* della finestra di dialogo del nodo Tipo. Le scelte per la lettura dei valori includono:

- **Leggi dai dati.** Selezionare questa opzione per leggere i valori durante l'esecuzione del nodo. L'opzione è analoga a <Leggi>.
- **Passa.** Selezionare questa opzione per non leggere i dati per il campo corrente. L'opzione è analoga a <Passa>.
- **Specifica valori.** Le opzioni disponibili consentono di specificare i valori e le etichette per il campo selezionato. Se utilizzata insieme al controllo valori, questa opzione permette di specificare i valori in base alle informazioni note sul campo corrente, inoltre attiva controlli univoci per ciascun tipo di campo. Le singole opzioni relative ai valori e alle etichette verranno illustrate negli argomenti successivi. *Nota:* non è possibile specificare valori o etichette per un campo il cui livello di misurazione è *Senza tipo* o <Default>.
- **Estendi valori dai dati.** Selezionare questa opzione per accordare ai dati correnti i valori sopra specificati. Se, per esempio, il campo *campo_1* include un intervallo (0,10) e si immette un intervallo di valori (8,16), l'intervallo viene esteso aggiungendo 16, senza rimuovere il minimo originale. Il nuovo intervallo sarà (0,16). La scelta di questa opzione imposta automaticamente l'opzione di impostazione automatica su <Leggi+>.

Controlla i valori. Selezionare un metodo per forzare i valori in modo che siano conformi ai valori continui, flag o nominali specificati. Questa opzione corrisponde alla colonna *Controllo* nella finestra di dialogo del nodo Tipo, le cui impostazioni vengono sovrascritte da quelle specificate in questa finestra. Se utilizzato insieme all'opzione Specifica valori, il controllo dei valori consente di uniformare i valori nei dati ai valori previsti. Per esempio, se si specificano valori come 1, 0 e si utilizza l'opzione Scarta, è possibile scartare tutti i record con valori diversi da 1 o 0.

Definisci vuoti. Selezionare questa opzione per attivare i controlli sotto riportati, che consentono di dichiarare valori mancanti o vuoti nei dati.

- **Tabella Valori mancanti.** Consente di definire valori specifici (per esempio 99 o 0) come vuoti. Il valore deve essere appropriato per il tipo di archiviazione del campo.

- **Intervallo.** Consente di specificare un intervallo di valori mancanti, per esempio le età 1–17 o maggiori di 65. Se non si specifica un limite, l'intervallo sarà senza limite: per esempio, se si specifica un limite inferiore di 100 e nessun limite superiore, tutti i valori maggiori o uguali a 100 saranno definiti come mancanti. I valori limite sono inclusi nell'intervallo. Per esempio, un intervallo con limite inferiore 5 e limite superiore 10 includerà i valori 5 e 10. È possibile definire un intervallo di valori mancanti per qualsiasi tipo di classe di archiviazione, inclusi data/ora e stringa. In quest'ultimo caso, per stabilire se un valore rientra nell'intervallo verrà utilizzato l'ordinamento alfabetico.
- **Nulllo/Spazio bianco.** È inoltre possibile definire come vuoti i valori di sistema **nulli** (visualizzati nei dati come \$null\$) e **spazio bianco** (valori stringa senza caratteri visibili). Ai fini dell'analisi, anche il nodo Tipo considera le stringhe vuote come spazi bianchi, nonostante vengano archiviate in maniera diversa internamente e possano, in determinati casi, essere gestite diversamente. [Per ulteriori informazioni, vedere l'argomento Panoramica sui valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Nota: per codificare i vuoti come non definiti o \$null\$, è necessario utilizzare il nodo Riempimento.

Descrizione. Utilizzare questa casella di testo per specificare un'etichetta del campo. Le etichette vengono visualizzate in numerose posizioni, per esempio nei grafici, nelle tabelle, nell'output e nei browser di modelli, in base alle selezioni eseguite nella finestra di dialogo delle proprietà dello stream.

Definizione di valori ed etichette per dati continui

Il livello di misurazione *Continuo* si utilizza per i campi numerici. Per i dati continui sono disponibili tre tipi di archiviazione:

- Reale
- Intero
- Data/ora

Per modificare tutti i campi continui si utilizza la stessa finestra di dialogo; il tipo di archiviazione è visualizzato solo a scopo informativo.

Figura 4-24

Opzioni per la definizione di valori continui e delle relative etichette

Misurazione: Continuo Archiviazione: Intero

Valori: Leggi dai dati Passa Specifica valori

Inferiore: 15

Superiore: 74

Etichette...

Definizione dei valori

I controlli seguenti sono specifici dei campi continui e consentono di definire un intervallo di valori:

Ultimo. Specificare un limite inferiore per l'intervallo di valori.

Primo. Specificare un limite superiore per l'intervallo di valori.

Definizione delle etichette

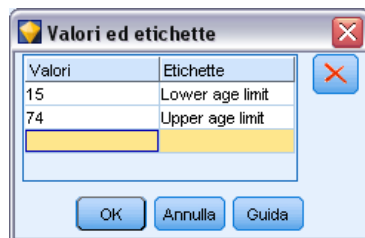
È possibile definire le etichette per qualsiasi valore di un campo intervallo. Fare clic sul pulsante Etichette per aprire una finestra di dialogo nella quale è possibile definire le etichette dei valori.

Finestra secondaria dei valori e delle etichette

Se si fa clic su Etichette nella finestra di dialogo Valori relativa a un campo intervallo, verrà aperta una nuova finestra di dialogo nella quale è possibile definire le etichette relative a qualsiasi valore compreso nell'intervallo.

Figura 4-25

Inserimento delle etichette (facoltative) relative ai valori dell'intervallo



Per definire le coppie di valori ed etichette, è possibile utilizzare le colonne *Valori* e *Etichette* della tabella. Sono visualizzate le coppie già definite. Per aggiungere nuove coppie, è possibile fare clic su una cella vuota e immettere un valore e la relativa etichetta. *Nota:* L'aggiunta di coppie valore/valore-etichetta alla tabella non comporterà l'aggiunta di nuovi valori al campo, ma verranno semplicemente creati i metadati per il valore del campo.

Le etichette specificate nel nodo Tipo vengono visualizzate in numerose posizioni (come testi descrittivi, etichette dell'output e così via), in base alle selezioni eseguite nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Definizione di valori ed etichette per dati nominali e ordinali

I livelli di misurazione nominali (insieme) e ordinali (insieme ordinato) indicano che i valori dei dati vengono utilizzati in modo discreto come un membro dell'insieme. I tipi di archiviazione per un insieme possono essere: stringa, numero intero, numero reale o data/ora.

Figura 4-26
Opzioni per la definizione dei valori nominali e delle relative etichette

Misurazione: Nominale Archiviazione: Stringa Campo del modello...

Valori: Leggi dai dati Passa
 Specifica valori

Valori	Etichette
drugA	Lisinopril
drugB	Metoprolol
drugC	Hydrochlorothiazide
drugX	Amlodipine
drugY	

I controlli seguenti sono specifici dei campi nominali e ordinali e consentono di definire i valori e le etichette:

Valori. La colonna *Valori* della tabella consente di specificare i valori in base alle informazioni note sul campo corrente. Tramite questa tabella è possibile immettere i valori previsti per il campo e quindi controllare la conformità dell'insieme di dati a questi valori utilizzando l'elenco a discesa Controlla valori. I pulsanti freccia e quello contrassegnato da una X consentono di modificare l'ordine dei valori esistenti o di eliminarli.

Etichette. La colonna *Etichette* consente di specificare le etichette per ogni valore del campo insieme. Le etichette vengono visualizzate in numerose posizioni, per esempio nei grafici, nelle tabelle, nell'output e nei browser di modelli, in base alle selezioni eseguite nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Definizione dei valori relativi a un campo flag

I campi flag vengono utilizzati per visualizzare i dati con due valori distinti. I tipi di archiviazione per i flag possono essere: stringa, numero intero, numero reale o data/ora.

Figura 4-27
Opzioni relative alla definizione dei valori di un campo flag

Misurazione: Flag Archiviazione: Stringa Campo del modello...

Valori: Leggi dai dati Passa
 Specifica valori

Vero: Etichetta:

Falso: Etichetta:

Vero. Specifica un valore flag per il campo quando la condizione viene soddisfatta.

Falso. Specifica un valore flag per il campo quando la condizione non viene soddisfatta.

Etichette. Specifica le etichette per ogni valore del campo flag. Le etichette vengono visualizzate in numerose posizioni, per esempio nei grafici, nelle tabelle, nell'output e nei browser di modelli, in base alle selezioni eseguite nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Definizione dei valori mancanti

La colonna Mancante della scheda Tipi indica se per un campo è stata definita la gestione dei valori mancanti. Le impostazioni possibili sono:

On (*) Indica che per questo campo è definita la gestione dei valori mancanti. Questo è anche possibile utilizzando il nodo Riempimento a valle oppure tramite una specifica esplicita utilizzando l'opzione Specifica (vedere qui di seguito).

Off. Il campo non ha una gestione dei valori mancanti definita.

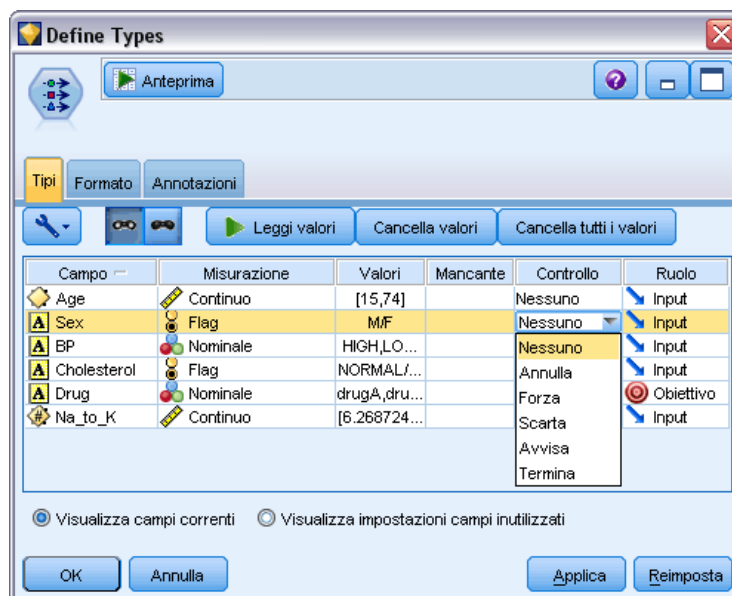
Specifica. Scegliere questa opzione per visualizzare una finestra di dialogo in cui è possibile dichiarare che i valori espliciti vengano considerati come valori mancanti per questo campo.

Per ulteriori informazioni, vedere l'argomento [Panoramica sui valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Controllo dei valori del tipo

Se si attiva l'opzione Controllo per ogni campo, verranno esaminati tutti i valori nel singolo campo in modo da stabilire se rispondono alle impostazioni correnti del tipo o ai valori specificati nella finestra di dialogo Specifica valori. Questa opzione è utile per ripulire gli insiemi di dati e ridurre le dimensioni con un'unica operazione.

Figura 4-28
Selezione delle opzioni di controllo relative al campo selezionato



L'impostazione della colonna *Controllo* nella finestra di dialogo del nodo Tipo, definisce le operazioni che verranno eseguite nel caso venga individuato un valore che non rientra nei limiti del tipo. Per cambiare le impostazioni Controllo per un campo, utilizzare l'elenco a discesa relativo a tale campo nella colonna *Controllo*. Per definire le impostazioni Controllo per tutti i

campi, fare clic nella colonna *Campo* e premere Ctrl-A. Quindi utilizzare l'elenco a discesa per ogni campo nella colonna *Controllo*.

Sono disponibili le seguenti impostazioni Controllo:

Nessuna. I valori verranno passati senza effettuare alcun controllo. (impostazione predefinita).

Annulla. Cambia i valori esterni ai limiti nel valore nullo di sistema (\$null\$).

Forza. Verranno controllati i campi i cui livelli di misurazione sono completamente istanziati, per scoprire se contengono valori che non rientrano negli intervalli specificati. I valori non specificati verranno convertiti in un valore valido per quel livello di misurazione, applicando le seguenti regole:

- Per i flag, tutti i valori diversi da quelli veri e falsi vengono convertiti nel valore falso.
- Per gli insiemi (nominali o ordinali), qualsiasi valore sconosciuto viene convertito nel primo membro dei valori dell'insieme.
- I numeri maggiori del limite superiore di un intervallo vengono sostituiti con il limite superiore.
- I numeri inferiori al limite inferiore di un intervallo vengono sostituiti con il limite inferiore.
- Ai valori nulli in un intervallo viene assegnato il valore intermedio dell'intervallo.

Scarta. Quando vengono trovati valori non validi, viene scartato l'intero record.

Avvisa. Viene conteggiato il numero di elementi non validi e il risultato del conteggio viene riportato nella finestra di dialogo delle proprietà dello stream al termine della lettura di tutti i dati.

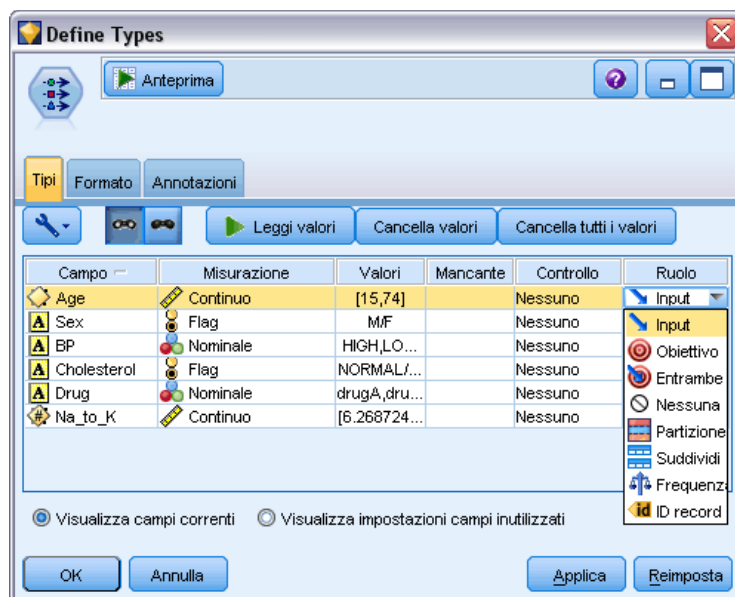
Termina. Il primo valore non valido trovato determina l'interruzione dell'esecuzione dello stream. L'errore viene riportato nella finestra di dialogo delle proprietà dello stream.

Impostazione del ruolo del campo

Il ruolo di un campo descrive il modo in cui verrà utilizzato per la creazione dei modelli, per esempio se un campo è un campo di input o un campo obiettivo (oggetto della previsione).

Nota: i ruoli Partizione, Frequenza e ID record possono essere applicati unicamente a un campo singolo.

Figura 4-29
Impostazione delle opzioni relative al ruolo dei campi per il nodo Tipo



Sono disponibili i ruoli seguenti:

Input. Il campo verrà utilizzato come input per l'apprendimento automatico (un campo predittore).

Obiettivo. Il campo verrà utilizzato come output o come obiettivo per l'apprendimento automatico (uno dei campi che il modello tenterà di prevedere).

Entrambi. Il campo verrà utilizzato sia come input che come output dal nodo Apriori. Tutti gli altri nodi Modelli ignoreranno il campo.

Nessuna. Il campo verrà ignorato dall'apprendimento automatico. I campi il cui livello di misurazione è impostato su Senza tipo vengono automaticamente impostati su Nessuno nella colonna *Ruolo*.

Partizione. Indica un campo utilizzato per partizionare i dati in campioni di addestramento, di test e di validazione (facoltativo) distinti. Questo campo deve essere un tipo di insieme istanziato con due o tre valori possibili (definiti nella finestra di dialogo Valori di campo). Il primo valore rappresenta il campione di addestramento, il secondo rappresenta il campione di test e il terzo (se presente) rappresenta il campione di validazione. Tutti i valori aggiuntivi vengono ignorati e non è possibile utilizzare i campi flag. Si noti che, per utilizzare la partizione in un'analisi, è necessario attivare il partizionamento nella scheda Opzioni modello del nodo di creazione del modello o di analisi appropriato. Se il partizionamento è attivato, i record con valori nulli del campo di partizione vengono esclusi dall'analisi. Se nello stream sono stati definiti più campi di partizione, nella scheda Campi di ogni nodo Modelli applicabile è necessario selezionare un singolo campo di partizione. Se nei dati non esiste già un campo appropriato, è possibile crearne uno utilizzando un nodo Partizione o Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Nodo Partizione a pag. 193.](#)

Suddivisione. (Solo per campi nominali, ordinali e flag) Specifica che deve essere creato un modello per ciascun valore possibile del campo. [Per ulteriori informazioni, vedere l'argomento Creazione di modelli suddivisi in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

Frequenza. (Solo per campi numerici) L'impostazione di questo ruolo consente di utilizzare un valore del campo come fattore di peso della frequenza per il record. Questa funzione è supportata soltanto dai nodi C&R Tree, CHAID, QUEST e Modelli lineari; tutti gli altri nodi ignorano questo ruolo. Il peso della frequenza si attiva mediante l'opzione Utilizza campo frequenza nella scheda Campi dei nodi Modelli che supportano la funzione. [Per ulteriori informazioni, vedere l'argomento Utilizzo dei campi frequenza e peso. in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

ID record. Il campo viene utilizzato come identificatore univoco del record. Questa funzione è ignorata dalla maggior parte dei nodi, ma è supportata dai modelli lineari.

Copia degli attributi del tipo

È facile copiare gli attributi del tipo di un campo (per esempio valori, opzioni di controllo e valori mancanti) e incollarli in un altro campo:

- ▶ Fare clic con il pulsante destro del mouse sul campo di cui si desidera copiare gli attributi.
- ▶ Dal menu di scelta rapida scegliere Copia.
- ▶ Fare clic con il pulsante destro del mouse sul campo o sui campi di cui si desidera cambiare gli attributi.
- ▶ Dal menu di scelta rapida scegliere Incolla speciale. *Nota:* è possibile selezionare più campi utilizzando il metodo Ctrl-clic o l'opzione Seleziona campi dal menu di scelta rapida.

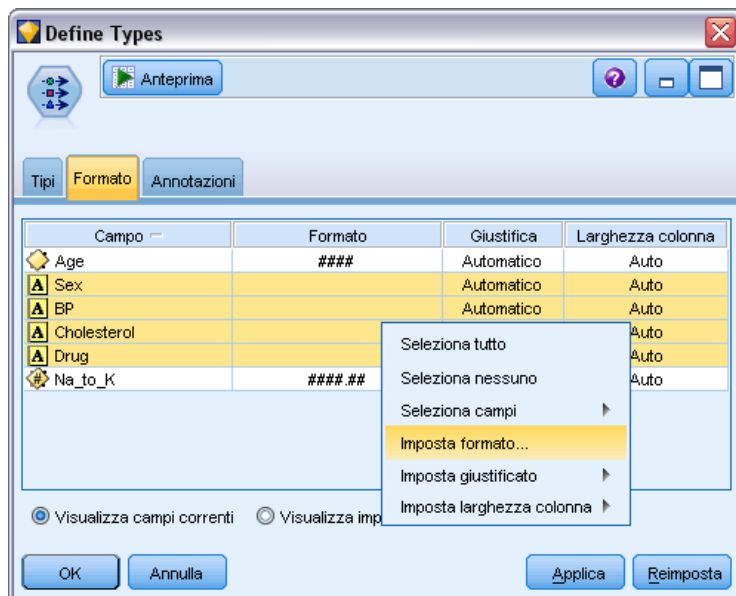
Verrà aperta una nuova finestra di dialogo dalla quale selezionare gli attributi specifici che si desidera incollare. Se gli attributi vengono incollati in più campi, le opzioni selezionate in questa finestra verranno applicate a tutti i campi obiettivo.

Incolla i seguenti attributi. Selezionare dall'elenco sotto riportato l'attributo che si desidera incollare in un altro campo.

- **Tipo.** Selezionare questa opzione per incollare il livello di misurazione.
- **Valori.** Selezionare questa opzione per incollare i valori di campo.
- **Mancante.** Selezionare questa opzione per incollare le impostazioni del valore mancante.
- **Controllo.** Selezionare questa opzione per incollare le opzioni di controllo dei valori.
- **Ruolo.** Selezionare questa opzione per incollare il ruolo di un campo.

Scheda per l'impostazione dei formati dei campi

Figura 4-30
Nodo Tipo, scheda Formato



Nella scheda Formato dei nodi Tabella e Tipo sono visualizzati l'elenco dei campi correnti o inutilizzati e le opzioni di formattazione relativi a ogni campo. Di seguito è riportata la descrizione di ogni colonna disponibile nella tabella di formattazione dei campi:

Campo. Mostra il nome del campo selezionato.

Formato. Se si fa doppio clic su una cella della colonna, verrà aperta una finestra di dialogo nella quale è possibile specificare la formattazione dei singoli campi. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni relative al formato dei campi a pag. 142.](#) Le impostazioni di formattazione definite qui hanno la precedenza su quelle definite nelle proprietà generali dello stream.

Nota: I nodi Esporta Statistics e Output Statistics esportano file *.sav* che includono nei metadati informazioni di formattazione relative ai singoli campi. Se si specifica un formato per ognuno dei campi che non è supportato dal formato di file *.sav* di IBM® SPSS® Statistics, il nodo utilizzerà il formato di default di SPSS Statistics.

Giustifica. Utilizzare questa colonna per specificare le modalità di giustificazione dei valori nella colonna della tabella. L'impostazione di default è Automatico e giustifica a sinistra i valori simbolici e a destra i valori numerici. Per ignorare l'impostazione di default, selezionare A sinistra, A destra o Al centro.

Larghezza colonna. Per default, le larghezze delle colonne vengono calcolate automaticamente in base ai valori dei campi. Per ignorare il calcolo automatico della larghezza, fare clic su una cella della tabella e selezionare una larghezza dall'elenco a discesa. Per specificare una larghezza personalizzata, aprire la finestra secondaria relativa ai formati dei campi facendo doppio clic su una cella della tabella nella colonna *Campo* o *Formato*. In alternativa, fare clic con il pulsante destro del mouse su una cella e selezionare Imposta formato.

Visualizza campi correnti. Per default, la finestra di dialogo mostra l'elenco dei campi attivi. Per visualizzare l'elenco dei campi inutilizzati, selezionare Visualizza impostazioni campi inutilizzati.

Menu di scelta rapida. Il menu di scelta rapida per questa scheda offre numerose opzioni di selezione e aggiornamento delle impostazioni.

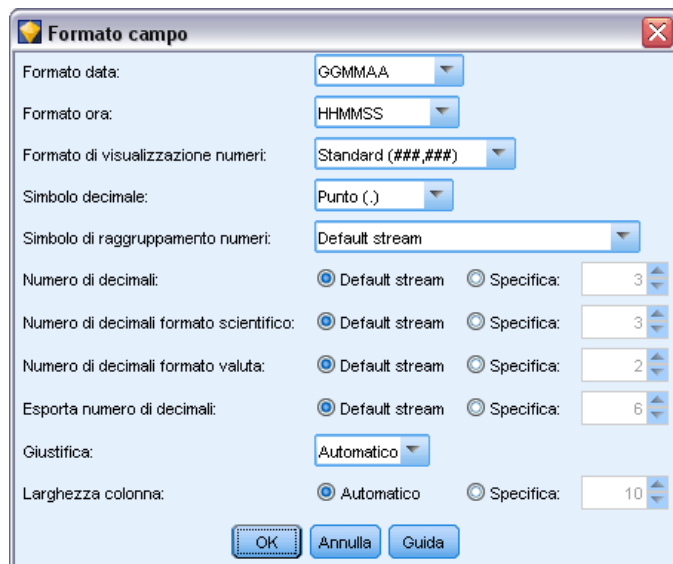
- **Seleziona tutto.** Seleziona tutti i campi.
- **Seleziona nessuno.** Cancella la selezione.
- **Seleziona campi.** Seleziona i campi in base al tipo o alle caratteristiche di archiviazione. Le opzioni sono Seleziona categoriali, Seleziona continui (numerici), Seleziona senza tipo, Seleziona stringhe, Seleziona numeri o Seleziona data/ora. [Per ulteriori informazioni, vedere l'argomento Livelli di misurazione a pag. 127.](#)
- **Imposta formato.** Apre una finestra secondaria in cui è possibile specificare la data, l'ora e le opzioni decimali per ognuno dei campi.
- **Imposta giustificato.** Imposta la giustificazione per il campo o i campi selezionati. Le opzioni sono Automatico, Al centro, A sinistra o A destra.
- **Imposta larghezza colonna.** Imposta la larghezza dei campi selezionati. Specificare Automatica per leggere il valore dai dati oppure impostare il valore su 5, 10, 20, 30, 50, 100 o 200.

Impostazione delle opzioni relative al formato dei campi

La formattazione dei campi viene definita in una finestra di dialogo secondaria della scheda Formato dei nodi Tipo e Tabella. Se prima di aprire questa finestra sono stati selezionati più campi, verranno utilizzate le impostazioni del primo campo selezionato. Se si fa clic su OK dopo avere impostato le opzioni, tali opzioni verranno applicate a tutti i campi selezionati nella scheda Formato.

Figura 4-31

Impostazione delle opzioni di formattazione per uno o più campi



Per ognuno dei campi sono disponibili le opzioni seguenti, molte delle quali possono anche essere impostate nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#) Le impostazioni stabilite a livello di campo hanno tuttavia la precedenza sulle impostazioni di default definite per lo stream.

Formato data. Selezionare un formato di data da utilizzare per i campi di archiviazione delle date o per le stringhe interpretate come date dalle funzioni di data CLEM.

Formato ora. Selezionare un formato di ora da utilizzare per i campi di archiviazione dell'ora o per le stringhe interpretate come orari dalle funzioni di ora CLEM.

Formato di visualizzazione numeri. È possibile scegliere il formato di visualizzazione standard (#####.###), scientifico (#.###E+##) o valuta (\$###.##).

Simbolo decimale. Selezionare una virgola (,) o un punto (.) come separatore decimale.

Simbolo di raggruppamento. Per i formati di visualizzazione dei numeri, selezionare il simbolo utilizzato per raggruppare i valori (per esempio la virgola per 3.000,00). Le opzioni disponibili sono Nessuno, Punto, Virgola, Spazio e Definito in base alle opzioni internazionali (in questo caso verrà utilizzato il valore di default per le opzioni internazionali correnti).

Numero di decimali (Standard, Scientifico, Valuta, Esporta). Per i formati di visualizzazione dei numeri, specifica il numero di decimali da utilizzare per la visualizzazione, la stampa o l'esportazione di numeri reali. Questa opzione viene specificata separatamente per ogni formato di visualizzazione. Il formato di esportazione è applicabile solo ai campi con classe di archiviazione reale.

Giustifica. Specifica le modalità di giustificazione dei valori nella colonna. L'impostazione di default è Automatico e giustifica a sinistra i valori simbolici e a destra i valori numerici. Per ignorare l'impostazione di default, selezionare A sinistra, A destra o Al centro.

Larghezza colonna. Per default, le larghezze delle colonne vengono calcolate automaticamente in base ai valori dei campi. È possibile specificare una larghezza personalizzata in base a intervalli di cinque, utilizzando le frecce a destra della casella di riepilogo.

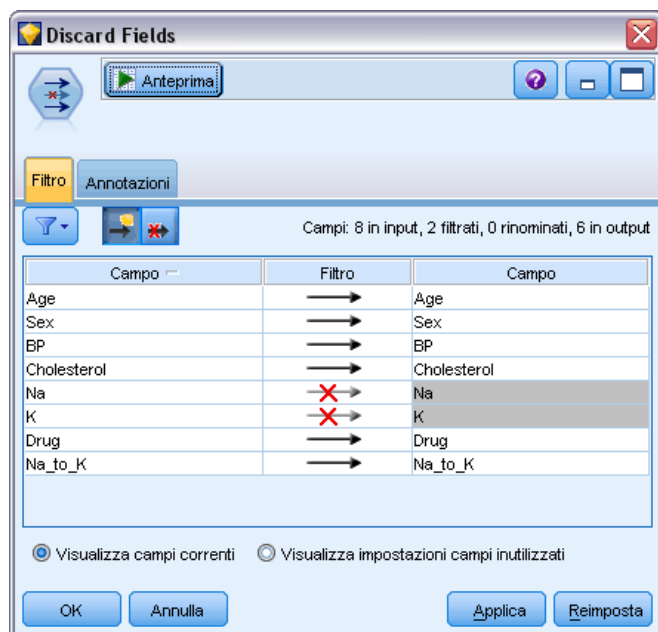
Applicazione di filtri o ridenominazione di campi

È possibile rinominare o escludere i campi in qualsiasi punto di uno stream. Per esempio, è possibile che un ricercatore medico non sia interessato al livello di potassio (dati a livello di campo) dei pazienti (dati a livello di record). Grazie a questo nodo il ricercatore può filtrare il campo *K* (potassio). Si può eseguire questa operazione mediante un nodo Filtro separato o tramite la scheda Filtro di un nodo di input o di output. La funzionalità è identica, indipendentemente dal nodo da cui vi si accede.

- Da nodi di input quali Testo variabile, Testo fisso, File Statistics e XML, è possibile rinominare o filtrare i campi mentre i dati vengono letti in IBM® SPSS® Modeler.
- Un nodo Filtro consente di rinominare o filtrare i campi in qualsiasi punto dello stream.

- Dai nodi Esporta Statistics, Trasformazioni Statistics, Modelli Statistics e Output Statistics è possibile filtrare o rinominare i campi per adeguarli agli standard di denominazione di IBM® SPSS® Statistics. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics in il capitolo 8 a pag. 451.](#)
- In tutti i nodi sopracitati è possibile utilizzare la scheda Filtro per definire o modificare gli insiemi di risposte multiple. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple a pag. 147.](#)
- Infine, il nodo Filtro può essere utilizzato per mappare i campi tra nodi di input. [Per ulteriori informazioni, vedere l'argomento Mappatura degli stream di dati in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Figura 4-32
Impostazione delle opzioni relative al nodo Filtro



Impostazione delle opzioni di filtro

Nella tabella utilizzata nella scheda Filtro, per ogni campo viene riportato il nome del campo in entrata nel nodo e in uscita dal nodo. È possibile utilizzare le opzioni in questa tabella per rinominare o filtrare i campi duplicati o non necessari per operazioni a valle.

- **Campo.** Visualizza i campi di input dalle sorgenti dati connesse.

- **Filtro.** Visualizza lo stato del filtro di tutti i campi di input. I campi filtrati sono contrassegnati da una X rossa in questa colonna e non verranno passati a valle. Fare clic nella colonna *Filtro* per attivare e disattivare il filtro di un campo selezionato. È anche possibile selezionare le opzioni per più campi contemporaneamente, tramite il metodo di selezione Maiusc-clic.
- **Campo.** Visualizza lo stato dei campi in uscita dal nodo Filtro. I nomi duplicati sono visualizzati in rosso. È possibile modificare i nomi dei campi facendo clic su questa colonna e immettendo il nuovo nome. Per rimuovere i campi, fare clic nella colonna *Filtro* per disattivare i campi duplicati.

Tutte le colonne nella tabella possono essere ordinate facendo clic sulla relativa intestazione.

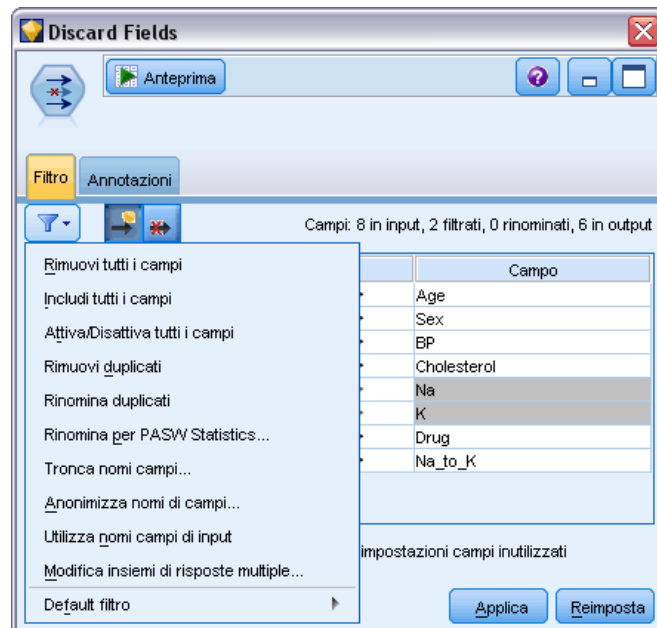
Visualizza campi correnti. Selezionare questa opzione per visualizzare campi per insiemi di dati attivi connessi al nodo Filtro. Questa opzione è selezionata per default e rappresenta il metodo di utilizzo dei nodi Filtro più comune.

Visualizza impostazioni campi inutilizzati. Selezionare questa opzione per visualizzare campi di insiemi di dati non più connessi al nodo Filtro. Questa opzione risulta utile quando si copiano nodi Filtro da uno stream in un altro stream e quando si ricaricano nodi Filtro.

Menu del pulsante Filtro

Fare clic sul pulsante Filtro nell'angolo superiore sinistro della finestra di dialogo per accedere a un menu contenente numerose scelte rapide e altre opzioni.

Figura 4-33
Opzioni del menu Filtro



È possibile scegliere di:

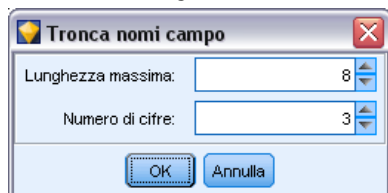
- Rimuovere tutti i campi.
- Includere tutti i campi.

- Attivare/disattivare tutti i campi.
- Rimuovere i duplicati. *Nota:* se si seleziona questa opzione, verranno rimosse tutte le occorrenze del nome duplicato, inclusa la prima.
- Rinominare i campi e gli insiemi di risposte multiple per conformarli a quelli delle altre applicazioni. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics in il capitolo 8 a pag. 451.](#)
- Troncare i nomi dei campi.
- Anonimizzare i nomi dei campi e degli insiemi di risposte multiple.
- Utilizzare i nomi dei campi di input.
- Modificare gli insiemi di risposte multiple. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple a pag. 147.](#)
- Impostare lo stato di default dei filtri.

È anche possibile utilizzare i pulsanti con le frecce nella parte superiore della finestra di dialogo per includere o scartare i campi per default. Questa opzione risulta utile per insiemi di dati di grandi dimensioni in cui per le operazioni a valle è necessario includere solo alcuni campi. Per esempio, è possibile selezionare solo i campi da conservare e specificare che tutti gli altri devono essere scartati (anziché selezionare i campi da scartare uno per uno).

Troncamento dei nomi di campo

Figura 4-34
Finestra di dialogo Tronca nomi campo



Dal menu del pulsante Filtro (nell'angolo superiore sinistro della scheda Filtro) è possibile scegliere di troncamento i nomi di campo.

Lunghezza massima. Consente di specificare un numero di caratteri per limitare la lunghezza dei nomi di campo.

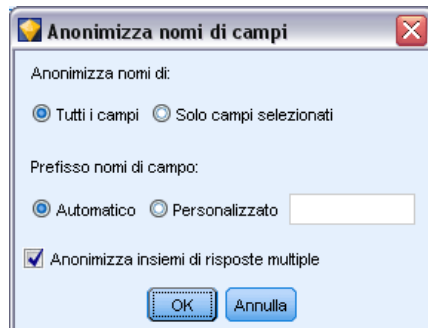
Numero di cifre. Se i nomi di campo non sono più univoci dopo il troncamento, verranno ulteriormente troncati e differenziati mediante l'aggiunta di cifre. È possibile specificare il numero di cifre utilizzate. I pulsanti con le frecce permettono di modificare il numero.

Per esempio, nella tabella seguente viene illustrato il troncamento di nomi di campo in un insieme di dati medici tramite le impostazioni di default (lunghezza massima = 8 e numero di cifre = 2).

Nomi di campi	Nomi di campo troncati
InputPaziente1	InputP01
InputPaziente2	InputP02
Battito Cardiaco	BattitoC
Pressione	Pressione

Anonimizzazione dei nomi di campi

Figura 4-35
Finestra di dialogo Anonimizza nomi di campi



È possibile anonimizzare i nomi dei campi da qualunque nodo contenente una scheda Filtro facendo clic sul menu del pulsante Filtro nell'angolo superiore sinistro e selezionando Anonimizza nomi di campi. I nomi dei campi anonimizzati consistono in un prefisso sotto forma di stringa e in un valore su base numerica univoco.

Anonimizza nomi di. Scegliere Solo campi selezionati per anonimizzare solo i nomi dei campi già selezionati nella scheda Filtro. L'impostazione di default è Tutti i campi, che anonimizza i nomi di tutti i campi.

Prefisso dei nomi dei campi. Il prefisso di default per i nomi di campi anonimizzati è anon_; se si preferisce utilizzare un prefisso diverso, scegliere Personalizzato e digitare il prefisso.

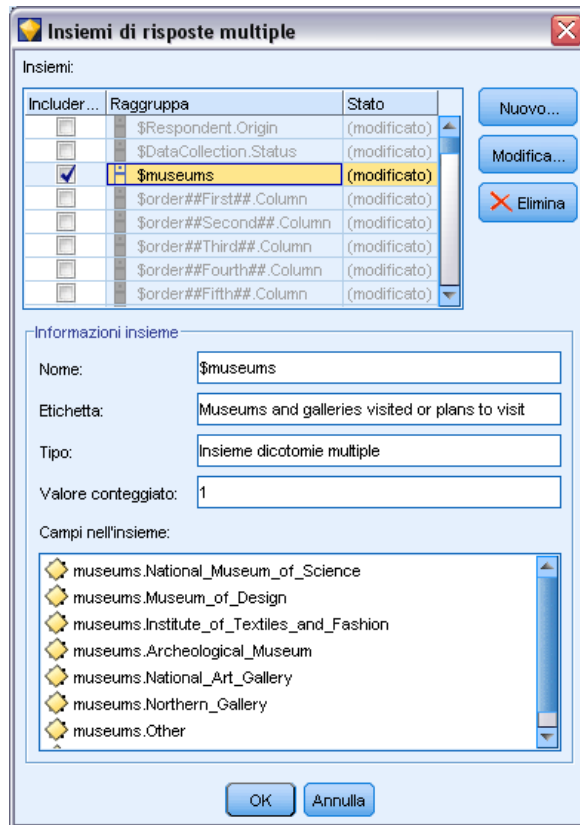
Anonimizza insiemi di risposte multiple. Anonimizza i nomi degli insiemi di risposte multiple analogamente a quanto avviene per i campi. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple a pag. 147.](#)

Per ripristinare i nomi di campi originali, scegliere Utilizza nomi campi di input dal menu di pulsanti di filtro.

Modifica di insiemi di risposte multiple

È possibile aggiungere o modificare gli insiemi di risposte multiple da qualunque nodo contenente una scheda Filtro facendo clic sul menu del pulsante Filtro nell'angolo superiore sinistro e selezionando Modifica insiemi di risposte multiple.

Figura 4-36
Finestra di dialogo *Insiemi di risposte multiple*



Gli insiemi di risposte multiple vengono utilizzati per registrare i dati che possono avere più di un valore per ogni caso, per esempio quando in un sondaggio si chiede agli intervistati quali musei hanno visitato o quali riviste hanno letto. Gli insiemi di risposte multiple possono essere importati in IBM® SPSS® Modeler mediante un nodo di input di Data Collection o un nodo di input File Statistics e possono essere definiti in SPSS Modeler mediante un nodo Filtro.

- Fare clic su Nuovo per creare un nuovo insieme di risposte multiple, oppure su Modifica per modificare un insieme esistente.

Figura 4-37
Modifica di un insieme di risposte multiple

Nome ed Etichetta. Riportano il nome e la descrizione dell'insieme.

Tipo. Le domande a risposta multipla possono essere gestite in due diversi modi:

- **Insieme dicotomie multiple.** Viene creato un campo flag separato per ogni risposta possibile: in questo modo, se le riviste sono 10, vi saranno 10 campi flag, ciascuno dei quali può avere dei valori quali 0 o 1 per indicare *vero* o *falso*. Il valore conteggiato consente di specificare quale valore viene conteggiato come vero. Questo metodo è utile quando si desidera permettere agli intervistati di scegliere tutte le opzioni pertinenti.
- **Insieme di categorie multiple.** Viene creato un campo nominale per ogni risposta fino al numero massimo di risposte date da un determinato intervistato. Ogni campo nominale contiene valori che rappresentano le risposte possibili, quali 1 per *Time*, 2 per *Newsweek* e 3 per *PC Week*. Questo metodo è utile soprattutto quando si desidera limitare il numero delle risposte, per esempio quando si chiede agli intervistati di scegliere le tre riviste che leggono con maggiore frequenza.

Campi nell'insieme. Utilizzare le icone a destra per aggiungere o rimuovere i campi.

Figura 4-38
Domanda a risposta multipla

Commenti

- Tutti i campi compresi in un insieme di risposte multiple devono avere lo stesso tipo di archiviazione.
- Gli insiemi e i campi in essi contenuti sono due entità separate. Per esempio, se si elimina un insieme non saranno eliminati anche i campi in esso contenuti, ma solo i collegamenti tra tali campi. L'insieme risulterà ancora visibile a monte del punto di eliminazione, ma non a valle.
- Se vengono rinominati dei campi tramite un nodo Filtro (direttamente dalla scheda o selezionando l'opzione Rinomina per IBM® SPSS® Statistics, Tronca o Anonimizza nel menu Filtro), saranno aggiornati anche tutti i riferimenti a tali campi utilizzati negli insiemi di risposte multiple. I campi di un insieme di risposte multiple eliminati mediante il nodo Filtro non saranno invece rimossi dall'insieme di risposte multiple. Tali campi, sebbene non più visibili nello stream, continuano ad avere dei riferimenti nell'insieme di risposte multiple; questo potrebbe essere un aspetto da tenere presente quando si esegue un'esportazione, per esempio.

Nodo Risultati classificatore binario

Il nodo Risultati classificatore binario combina due o più insiemi di modelli per ottenere previsioni più precise di quelle ricavabili dai singoli modelli. Se si combinano le previsioni di più modelli è possibile superare le limitazioni dei singoli modelli e ottenere un livello di precisione complessiva più elevato. I modelli combinati in questo modo forniscono in genere prestazioni almeno equivalenti a quelle del singolo modello migliore, quando non addirittura superiori.

Questa combinazione di nodi viene eseguita automaticamente nei nodi Modelli automatici Classificatore automatico, Numerico automatico e Cluster automatico. [Per ulteriori informazioni, vedere l'argomento Nodi Modelli automatici in il capitolo 5 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

Dopo avere utilizzato il nodo Risultati classificatore binario è possibile utilizzare un nodo Analisi o Valutazione per confrontare la precisione dei risultati combinati con quelli di ciascun modello di input. Per eseguire questa operazione, verificare che l'opzione Filtra campi generati dai modelli Classificatore binario sia deselezionata nella scheda Impostazioni del nodo Risultati classificatore binario.

Campi di output

Ogni nodo Risultati classificatore binario genera un campo contenente i punteggi combinati. Il nome del campo è composto dal campo obiettivo specificato con il prefisso $\$XF_$, $\$XS_$ o $\$XR_$, a seconda del livello di misurazione del campo - rispettivamente, flag, nominale (insieme) o continuo (intervallo). Per esempio, se il campo obiettivo è un campo flag denominato *risposta*, il campo di output sarà $\$XF_risposta$.

Campi di confidenza o propensione. Per i campi flag e nominali vengono creati dei campi di confidenza o propensione aggiuntivi in base al metodo di classificatore binario, come illustrato dalla seguente tabella:

Metodo Classificatore binario	Nome del campo
Voting Confronto ponderato con confidenza Confronto ponderato con propensione grezza Confronto ponderato con propensione grezza Confidenza più elevata vince	\$XFC_<campo>
Propensione grezza media	\$XFRP_<campo>
Propensione grezza/regolata media	\$XFAP_<campo>

Impostazioni del nodo Risultati classificatore binario

Campo obiettivo per classificatore binario. Selezionare un campo singolo che sarà utilizzato come obiettivo da due o più modelli a monte. I modelli a monte possono utilizzare obiettivi flag, nominali o continui, ma per poter combinare i punteggi è necessario che almeno due modelli condividano lo stesso obiettivo.

Filtra campi generati dai modelli Classificatore binario. Elimina dall'output tutti i campi aggiuntivi generati dai singoli modelli che alimentano il nodo Risultati classificatore binario. Selezionare questa casella di controllo se si è interessati unicamente al punteggio combinato di tutti i modelli di input. Verificare che questa opzione sia deselezionata se, per esempio, si desidera utilizzare un nodo Analisi o Valutazione per confrontare la precisione del punteggio combinato con quella dei singoli modelli di input.

Figura 4-39

Nodo Risultati classificatore binario con un campo continuo selezionato come obiettivo



Le impostazioni disponibili dipendono dal livello di misurazione del campo selezionato come obiettivo.

Obiettivi continui

Per gli obiettivi continui viene calcolata la media dei punteggi. Questo è l'unico metodo disponibile per combinare i punteggi.

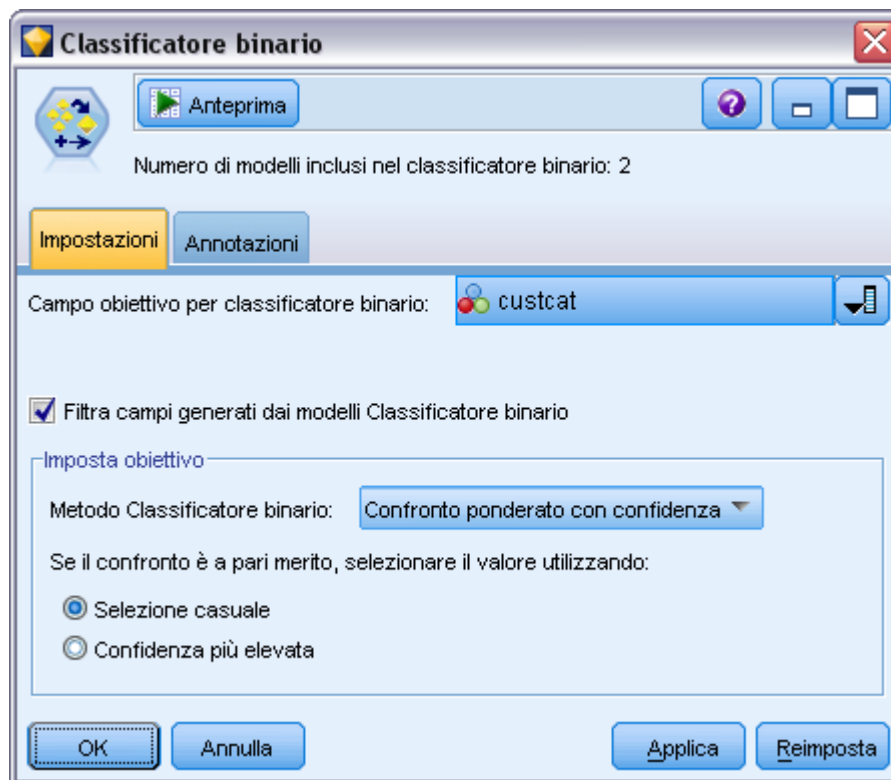
Quando si esegue la media di punteggi o stime, il nodo Risultati classificatore binario utilizza un calcolo degli errori standard per stabilire la differenza tra i valori misurati o stimati e quelli veri, e per mostrare la corrispondenza di queste stime. I calcoli degli errori standard vengono generati di default per i nuovi modelli; tuttavia, è possibile deselezionare la casella di controllo per i modelli esistenti, per esempio se devono essere rigenerati.

Obiettivi categoriali

Per gli obiettivi categoriali sono supportati numerosi metodi fra cui il **confronto**, che consiste nel conteggio del numero di volte in cui ogni possibile valore previsto viene scelto e nella selezione del valore con il totale più elevato. Per esempio, se tre modelli su cinque prevedono *sì* e gli altri due *no*, il *sì* vince per un voto di 3 a 2. In alternativa, i voti possono essere **ponderati** in base al valore di confidenza o di propensione di ogni previsione. Vengono quindi sommati i pesi e viene nuovamente selezionato il valore con il totale più alto. La confidenza della previsione finale è la somma dei pesi del valore vincente divisa per il numero dei modelli compresi nel classificatore binario.

Figura 4-40

Nodo Risultati classificatore binario con un campo nominale selezionato come obiettivo

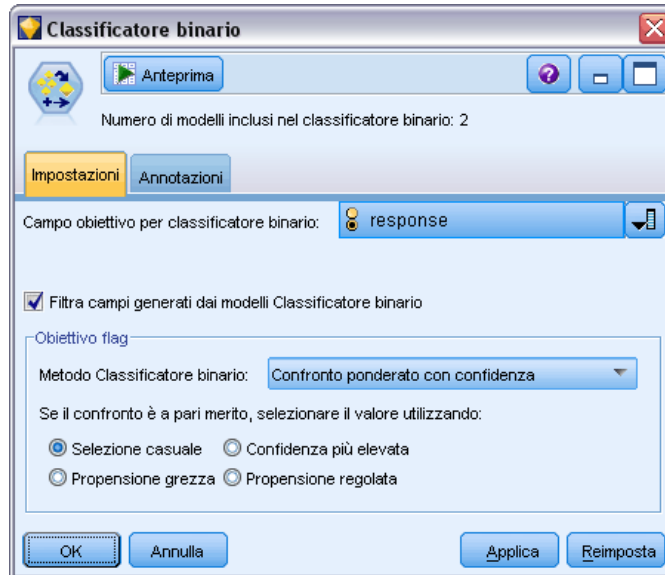


Tutti i campi categoriali. Sono supportati i seguenti metodi sia per i campi flag che per i campi nominali:

- Voting
- Confronto ponderato con confidenza
- Confidenza più elevata vince

Figura 4-41

Nodo Risultati classificatore binario con un campo flag selezionato come obiettivo



Solo i campi flag. Solo per i campi flag, sono disponibili anche numerosi metodi basati sulla propensione:

- Confronto ponderato con propensione grezza
- Confronto ponderato con propensione regolata
- Propensione grezza media
- Propensione regolata media

Confronti a pari merito. Per i metodi di confronto è possibile specificare le modalità di risoluzione delle situazioni di pari merito.

- **Selezione casuale.** Viene scelto a caso uno dei valori pari merito.
- **Confidenza più elevata.** Vince il valore pari merito previsto con la confidenza più elevata. Si noti che questa non corrisponde necessariamente alla confidenza più elevata di tutti i valori previsti.
- **Propensione grezza o regolata (solo campi flag).** Il valore pari merito previsto con la massima propensione assoluta, dove la propensione assoluta è calcolata come:

$$\frac{\text{abs}(0.5 - \text{propensione})}{2}$$

O, nel caso della propensione regolata:

$\text{abs}(0.5 - \text{propensione regolata}) * 2$

Per ulteriori informazioni, vedere l'argomento Opzioni della scheda Analizza di un nodo Modelli in il capitolo 3 in *IBM SPSS Modeler 14.2 Nodi Modelli*.

Nuovo campo

Una delle funzioni più efficaci di IBM® SPSS® Modeler è rappresentata dalla possibilità di modificare i valori dei dati e derivare nuovi campi dai dati esistenti. Nel corso di lunghi progetti di data mining è normale eseguire diverse derivazioni, per esempio estrarre gli ID clienti da una stringa di dati di registro Web oppure creare un valore del profitto futuro prodotto dal cliente in base ai dati demografici e alle transazioni. Queste trasformazioni possono essere eseguite tramite una grande varietà di nodi Operazioni su campi.

Alcuni nodi consentono di derivare nuovi campi:



Il nodo Nuovo campo modifica valori di dati o crea nuovi campi da uno o più campi esistenti. Crea campi di tipo Formula, Flag, Nominale, Stato, Conteggio e Condizionale. Per ulteriori informazioni, vedere l'argomento Nuovo campo a pag. 154.



Il nodo Ricodifica trasforma un insieme di valori categoriali in un altro. L'operazione di ricodifica consente di comprimere categorie o raggruppare dati per l'analisi. Per ulteriori informazioni, vedere l'argomento Nodo Ricodifica a pag. 174.



Il nodo Discretizza crea automaticamente nuovi campi nominali (insieme) basati sui valori di uno o più campi continui (intervallo numerico) esistenti. Per esempio, è possibile trasformare un campo continuo relativo al reddito in campo categoriale contenente gruppi di reddito come deviazioni dalla media. Dopo aver creato bin per il nuovo campo, è possibile generare un nodo Nuovo campo basato sui punti di taglio. Per ulteriori informazioni, vedere l'argomento Nodo Discretizza a pag. 178.



Il nodo Crea flag crea campi flag in base ai valori categoriali di uno o più campi nominali. Per ulteriori informazioni, vedere l'argomento Nodo Crea flag a pag. 196.



Il nodo Riorganizza converte un campo nominale o flag in un gruppo di campi in cui è possibile inserire i valori di un altro campo. Per esempio, dato un campo denominato *tipo di pagamento*, con valori di *credito*, *contanti* e *debito*, verrebbero creati tre nuovi campi (*credito*, *contanti*, *debito*), ognuno dei quali può contenere il valore del pagamento effettuato. Per ulteriori informazioni, vedere l'argomento Nodo Riorganizza a pag. 197.



Il nodo Cronologia crea nuovi campi contenenti dati dei campi di record precedenti. I nodi Cronologia sono utilizzati in genere per dati sequenziali, per esempio per dati di serie storica. Prima di utilizzare un nodo Cronologia, può essere utile ordinare i dati con un nodo Ordina. Per ulteriori informazioni, vedere l'argomento Nodo Cronologia a pag. 225.

Utilizzo del nodo Nuovo campo

Il nodo Nuovo campo consente di creare sei tipi di nuovi campi da uno o più campi esistenti:

- **Formula.** Il nuovo campo è il risultato di un'espressione CLEM arbitraria.
- **Flag.** Il nuovo campo è un flag che rappresenta una condizione specificata.
- **Nominale.** Il nuovo campo è nominale, ovvero i suoi membri sono un gruppo di valori specificati.
- **Stato.** Il nuovo campo è uno di due stati. La commutazione da uno stato all'altro è attivata da una condizione specificata.
- **Conteggio.** Il nuovo campo si basa sul numero di volte in cui una condizione è risultata vera.
- **Condizionale.** Il nuovo campo è il valore di una di due espressioni, in base al valore di una condizione.

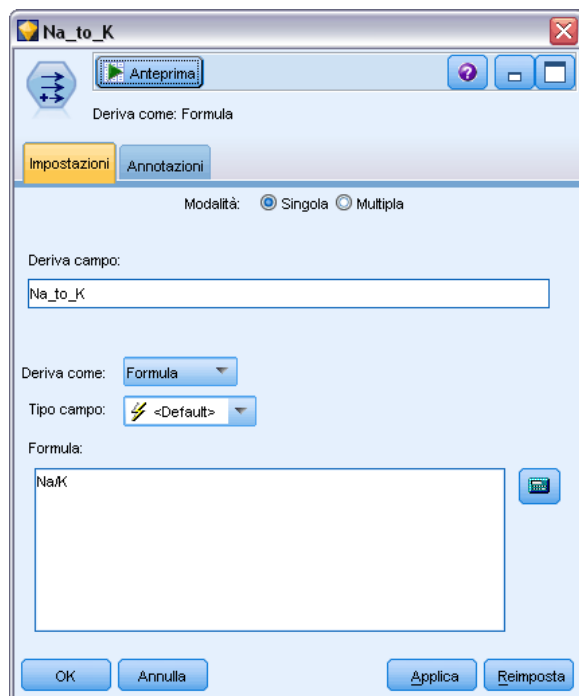
Ogni nodo contiene un insieme di opzioni speciali nella finestra di dialogo del nodo Nuovo campo. Tali opzioni verranno illustrate negli argomenti seguenti.

Impostazione delle opzioni di base relative al nodo Crea variabile

Nella parte superiore della finestra di dialogo per i nodi Crea variabile sono incluse diverse opzioni per la selezione del tipo di nodo Crea variabile desiderato.

Figura 4-42

Finestra di dialogo del nodo Nuovo campo



Modalità. Selezionare Singola o Multipla per derivare un solo campo o più campi. Se si seleziona Multipla, la finestra di dialogo cambia per includere opzioni per più campi.

Deriva campo. Per nodi Nuovo campo con modalità singola, specificare il nome del campo che si desidera derivare e aggiungere a ciascun record. Il nome di default è Nuovo campo N , dove N è il numero di nodi Nuovo campo creati fino a questo momento nella sessione corrente.

Deriva come. Selezionare un tipo di nodo Nuovo campo, per esempio Formula o Nominale, dall'elenco a discesa. Per ogni tipo viene creato un nuovo campo in base alle condizioni specificate nella finestra di dialogo relativa al tipo.

La selezione di un'opzione dall'elenco a discesa comporta l'aggiunta di un nuovo insieme di controlli alla finestra di dialogo principale, a seconda delle proprietà di ogni tipo di nodo Nuovo campo.

Tipo campo. Selezionare un livello di misurazione, per esempio continuo, categoriale o flag, per il nodo derivato. Questa opzione è comune a tutte le forme di nodi Nuovo campo.

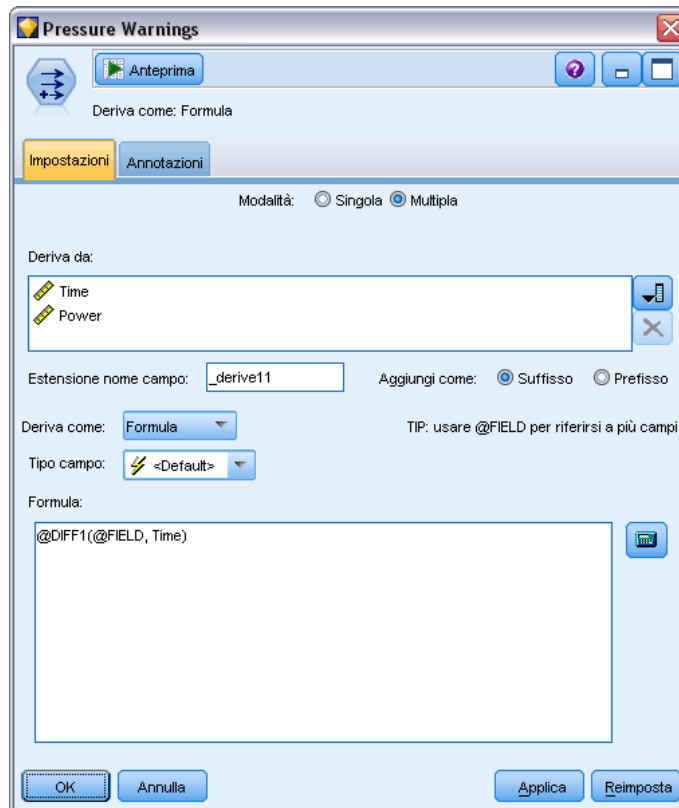
Nota: per derivare nuovi campi è spesso necessario utilizzare funzioni speciali o espressioni matematiche. Il generatore di espressioni, disponibile nella finestra di dialogo per tutti i tipi di nodi Nuovo campo, semplifica la creazione di tali espressioni e rende disponibile il controllo delle regole e un elenco completo delle espressioni CLEM. [Per ulteriori informazioni, vedere l'argomento Informazioni su CLEM in il capitolo 7 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Derivazione di campi multipli

L'impostazione della modalità Multipla in un nodo Crea variabile permette di derivare più campi basati sulla stessa condizione all'interno dello stesso nodo. Questa funzione consente di risparmiare tempo quando si desidera eseguire trasformazioni identiche per più campi dell'insieme di dati. Se si vuole, per esempio, generare un modello di regressione per la previsione del salario corrente in base al salario iniziale e all'esperienza precedente, potrebbe essere utile applicare una trasformazione logaritmica a tutte e tre le variabili asimmetriche. Invece di aggiungere un nuovo nodo Nuovo campo per ogni trasformazione, è possibile applicare contemporaneamente la stessa funzione a tutti i campi. È sufficiente selezionare tutti i campi dai quali si desidera derivare un nuovo campo, quindi digitare l'espressione di derivazione utilizzando la funzione @FIELD all'interno delle parentesi del campo.

Nota: la funzione @FIELD è uno strumento importante per la derivazione contemporanea di più campi, in quanto consente di fare riferimento al contenuto dei campi correnti senza specificare il nome esatto del campo. Per esempio, un'espressione CLEM utilizzata per applicare una trasformazione logaritmica a più campi è $\log(@FIELD)$.

Figura 4-43
Derivazione di più campi



Se si seleziona la modalità Multipla, alla finestra di dialogo vengono aggiunte le seguenti opzioni:

Deriva da. Utilizzare Selezione campi per selezionare i campi dai quali derivare nuovi campi. Per ogni campo selezionato verrà generato un campo di output. *Nota:* i campi selezionati non devono necessariamente essere dello stesso tipo di archiviazione. Tuttavia, se la condizione non è valida per *tutti* i campi l'operazione di derivazione non verrà eseguita.

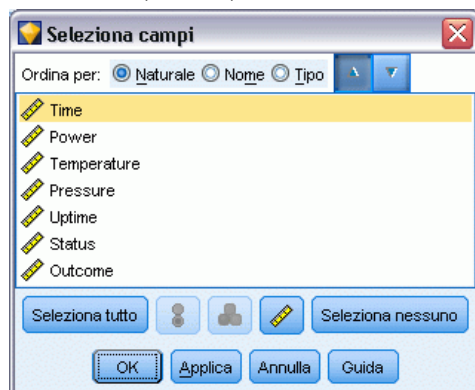
Estensione nome campo. Digitare l'estensione da aggiungere ai nuovi nomi di campo. Se un nuovo campo contiene, per esempio, il registro di *Salario Corrente*, è possibile aggiungere l'estensione *_registro* al nome del campo, ottenendo *Salario Corrente_registro*. Utilizzare i pulsanti di scelta rapida per aggiungere l'estensione come prefisso (all'inizio) o come suffisso (alla fine) del nome di campo. Il nome di default è Nuovo campo N , dove N è il numero di nodi Nuovo campo creati fino a questo momento nella sessione corrente.

Come nel nodo Nuovo campo in modalità singola, è necessario generare l'espressione da utilizzare per la derivazione di un nuovo campo. Le opzioni disponibili per creare una condizione variano a seconda del tipo di operazione di derivazione selezionata. Tali opzioni verranno illustrate negli argomenti seguenti. Per generare un'espressione, è possibile digitare l'espressione direttamente nei campi della formula oppure utilizzare il Generatore di espressioni facendo clic sul pulsante del calcolatore. Ricordarsi di utilizzare la funzione @FIELD quando ci si riferisce a manipolazioni su più campi.

Selezione di più campi

Per tutti i nodi che eseguono operazioni su più campi di input, per esempio Nuovo campo (modalità multipla), Aggregazione, Ordina, Plot multiplo e Plot tempo è possibile selezionare facilmente più campi tramite la finestra di dialogo Seleziona campi.

Figura 4-44
Selezione di più campi



Ordina per. Per visualizzare i campi disponibili in un determinato ordine, sono disponibili le seguenti opzioni:

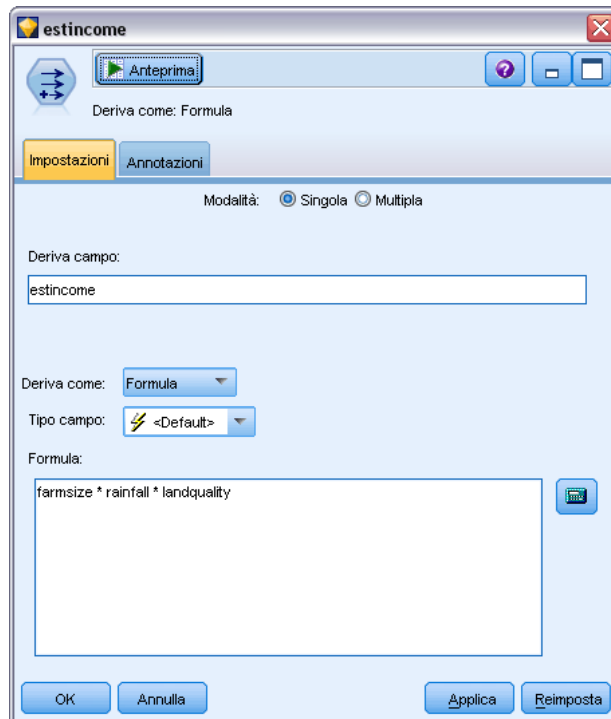
- **Naturale.** Visualizza l'ordine dei campi secondo la modalità di passaggio a valle nello stream di dati nel nodo corrente.
- **Nome.** Visualizza i campi in ordine alfabetico.
- **Tipo.** Visualizza i campi ordinati in base al livello di misurazione. Questa opzione è utile quando si selezionano campi con un determinato livello di misurazione.

Selezionare i campi dall'elenco uno per volta, oppure utilizzare i metodi Maiusc-clic e Ctrl-clic per selezionare più campi contemporaneamente. È anche possibile utilizzare i pulsanti nella parte inferiore dell'elenco per selezionare gruppi di campi in base al livello di misurazione, oppure per selezionare o deselezionare tutti i campi nella tabella.

Impostazione delle opzioni relative al nodo Crea variabile formula

I nodi Crea variabile formula creano un nuovo campo per ciascun record in un insieme di dati, in base ai risultati di un'espressione CLEM. Si noti che questa espressione non può essere condizionale. Per derivare valori in base a un'espressione condizionale, utilizzare il tipo flag o condizionale del nodo Nuovo campo.

Figura 4-45
Impostazione delle opzioni relative al nodo Nuovo campo formula



Formula. Specificare una formula utilizzando il linguaggio CLEM per derivare un valore per il nuovo campo. [Per ulteriori informazioni, vedere l'argomento Informazioni su CLEM in il capitolo 7 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Impostazione delle opzioni relative al nodo Nuovo campo flag

I nodi Nuovo campo flag vengono utilizzati per indicare una condizione specifica, per esempio pressione sanguigna elevata o inattività dell'account cliente. Per ciascun record viene creato un campo flag e, se la condizione è vera, nel campo viene aggiunto il valore flag per vero.

Figura 4-46
Derivazione di un campo flag

healthy

Deriva come: Flag

Impostazioni Annotazioni

Modalità: Singola Multipla

Deriva campo:
healthy

Deriva come: Flag

Tipo campo: Flag

Valore vero: T Valore falso: F

Valore vero quando:
'fruitveg' == 'T' and 'fish' == 'T'

OK Annulla Applica Reimposta

Valore vero. Specificare il valore da includere nel campo flag per i record che soddisfano la condizione specificata sotto. L'impostazione di default è V.

Valore falso. Specificare il valore da includere nel campo flag per i record che *non* soddisfano la condizione specificata sotto. L'impostazione di default è F.

Valore vero quando. Specificare una condizione CLEM per valutare determinati valori di ciascun record e assegnare al record un valore vero o falso (definito sopra). Si noti che il valore vero verrà assegnato ai record nel caso di valori numerici non falsi.

Nota: Per restituire una stringa vuota, digitare virgolette di apertura e di chiusura senza immettere alcun carattere tra di esse (""). Le stringhe vuote vengono utilizzate spesso, per esempio come valore falso, per dare maggior risalto ai valori veri presenti in una tabella. Le virgolette dovrebbero essere utilizzate anche se si desidera ottenere un valore stringa che altrimenti verrebbe trattato come numero

Esempio

Nelle versioni di IBM® SPSS® Modeler precedenti alla 12.0, le risposte multiple venivano importate in un unico campo, con i valori separati da virgole.

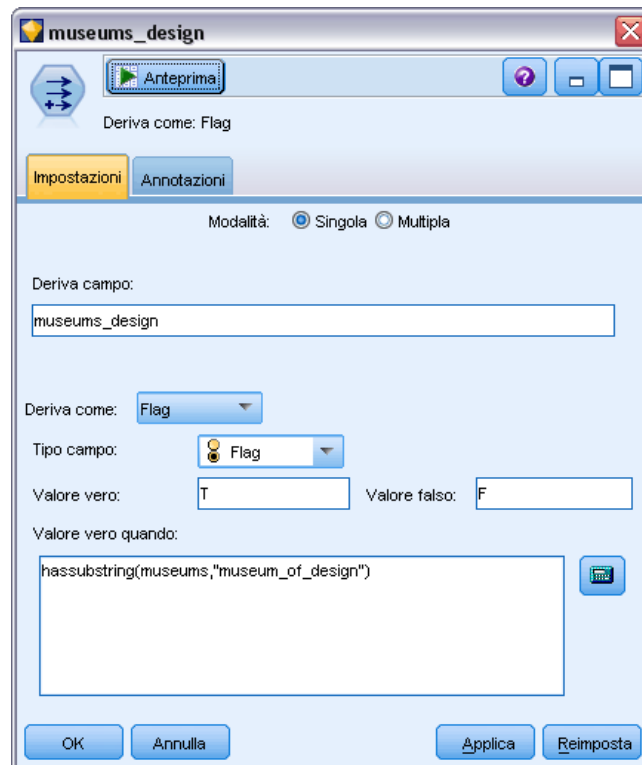
musei
museum_of_design,institute_of_textiles_and_fashion
museum_of_design
archeological_museum

musei
\$null\$
national_art_gallery,national_museum_of_science,other

Per preparare i dati per l'analisi è possibile utilizzare la funzione `hassubstring` per generare un campo flag separato per ciascuna risposta con un'espressione quale:

```
hassubstring(museums,"museum_of_design")
```

Figura 4-47
Derivazione di un campo flag con la funzione `hassubstring`



Impostazione delle opzioni relative al nodo Crea variabile insieme

I nodi Nuovo campo insieme sono utilizzati per eseguire un insieme di condizioni CLEM allo scopo di stabilire quale condizione venga soddisfatta da ciascun record. Ogni volta che viene rispettata una condizione per ciascun record, al nuovo campo derivato verrà aggiunto un valore indicante l'insieme di condizioni soddisfatte.

Figura 4-48
Utilizzo di un nodo Nuovo campo insieme



Valore di default. Specificare il valore da utilizzare nel nuovo campo se nessuna delle condizioni viene soddisfatta.

Imposta campo su. Specificare il valore da immettere nel nuovo campo quando viene soddisfatta una determinata condizione. A ogni valore dell'elenco è associata una condizione, specificata nella colonna adiacente.

Se questa condizione è vera. Specificare una condizione per ogni membro nel campo insieme da elencare. Utilizzare il Generatore di espressioni per selezionare funzioni e campi disponibili. È possibile utilizzare i pulsanti con le frecce e quello contrassegnato da una X per riordinare o rimuovere le condizioni.

Una condizione esegue la verifica dei valori di un particolare campo nell'insieme di dati. Al termine della verifica di ogni condizione, i valori specificati sopra verranno assegnati al nuovo campo in modo da indicare quale condizione è stata rispettata. Se non viene soddisfatta alcuna condizione, verrà utilizzato il valore di default.

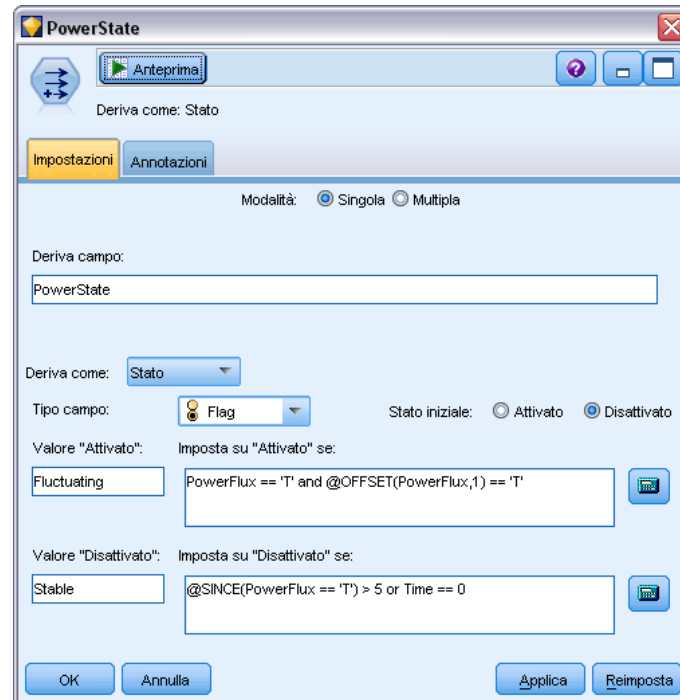
Impostazione delle opzioni relative al nodo Nuovo campo stato

I nodi Nuovo campo stato sono simili ai nodi Nuovo campo flag. Un nodo Nuovo campo flag imposta i valori in base al modo in cui viene soddisfatta una *singola* condizione del record corrente, invece un nodo Nuovo campo stato è in grado di modificare i valori di un campo in base

al modo in cui vengono soddisfatte *due condizioni indipendenti*. Ciò significa che il valore verrà modificato (ovvero attivato o disattivato) quando viene soddisfatta ciascuna condizione.

Figura 4-49

Utilizzo di un nodo Nuovo campo stato



Stato iniziale. Selezionare l'opzione desiderata per assegnare inizialmente il valore Attivato o Disattivato a ciascun record del nuovo campo. Si noti che questo valore può cambiare quando viene soddisfatta ciascuna condizione.

Valore "Attivato". Specificare il valore per il nuovo campo quando viene soddisfatta la condizione Attivato.

Imposta su "Attivato" se. Specificare una condizione CLEM per modificare lo stato in Attivato se la condizione è vera. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Valore "Disattivato". Specificare il valore per il nuovo campo quando viene soddisfatta la condizione Disattivato.

Imposta su "Disattivato" se. Specificare una condizione CLEM per modificare lo stato in Disattivato se la condizione è falsa. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Nota: Per specificare una stringa vuota, digitare virgolette di apertura e di chiusura senza immettere alcun carattere tra di esse (""). Le virgolette dovrebbero essere utilizzate anche se si desidera ottenere un valore stringa che altrimenti verrebbe trattato come numero.

Impostazione delle opzioni relative al nodo Crea variabile conteggio

Il nodo Crea variabile conteggio è utilizzato per applicare una serie di condizioni ai valori di un campo numerico nell'insieme di dati. Quando ogni condizione viene soddisfatta, il valore del campo conteggio derivato viene aumentato in base a un incremento prestabilito. Questo tipo di nodo Nuovo campo è utile per dati di serie storica.

Figura 4-50

Opzioni di conteggio nella finestra di dialogo del nodo Nuovo campo



Valore iniziale. Imposta un valore utilizzato per il nuovo campo durante l'esecuzione. Il valore iniziale deve essere una costante numerica. Utilizzare i pulsanti con le frecce per aumentare o diminuire il valore.

Incrementa quando. Specificare la condizione CLEM che, se soddisfatta, determina la modifica del valore derivato in base al numero specificato in Incrementa di. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Incrementa di. Imposta il valore utilizzato per incrementare il conteggio. È possibile utilizzare una costante numerica oppure il risultato di un'espressione CLEM.

Reimposta quando. Specificare una condizione che, se soddisfatta, determinerà la nuova impostazione del valore derivato sul valore iniziale. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Impostazione delle opzioni relative al nodo Nuovo campo condizionale

I nodi Nuovo campo condizionale utilizzano una serie di istruzioni Se, Allora, Altrimenti per derivare il valore del nuovo campo.

Figura 4-51
Utilizzo di un nodo Nuovo campo condizionale



Se. Specifica una condizione CLEM che verrà valutata per ciascun record durante l'esecuzione. Se la condizione è vera (o non falsa, nel caso di numeri), al nuovo campo viene assegnato il valore specificato sotto dall'espressione Allora. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

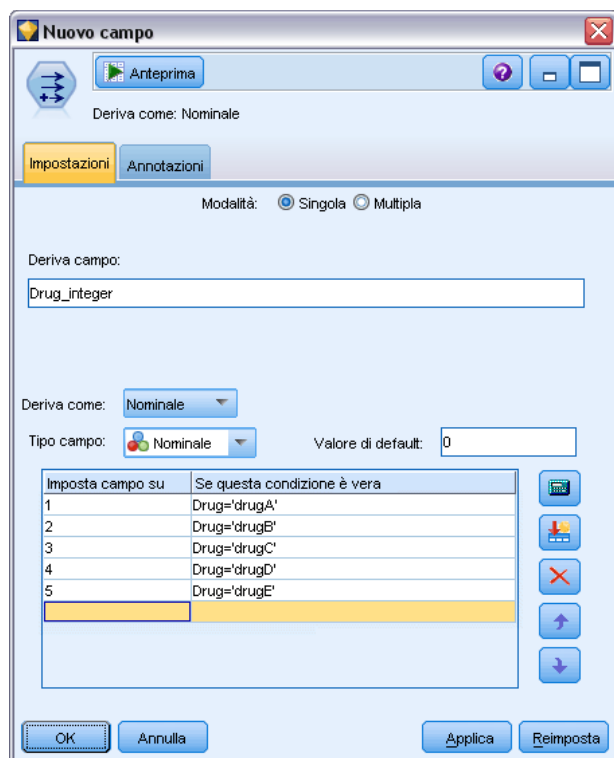
Allora. Specifica un valore o un'espressione CLEM per il nuovo campo quando l'istruzione Se sopra specificata è vera (o non falsa). Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Altrimenti. Specifica un valore o un'espressione CLEM per il nuovo campo quando l'istruzione Se sopra specificata è falsa. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Ricodifica di valori con il nodo Nuovo campo

I nodi Nuovo campo possono essere utilizzati anche per ricodificare i valori, per esempio convertendo un campo stringa con valori categoriali in un campo nominale (insieme) numerico.

Figura 4-52
Ricodifica di valori stringa

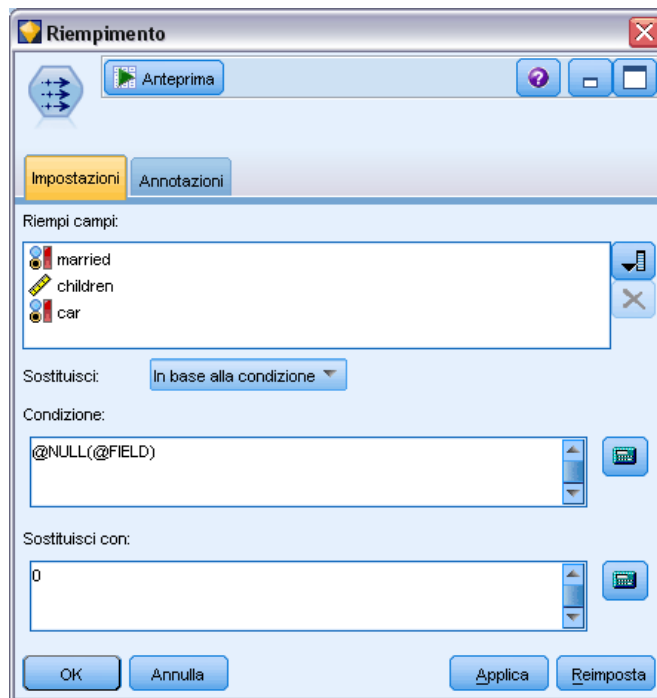


- Per Deriva come, selezionare il tipo di campo (nominale, flag e così via) desiderato.
- Specificare le condizioni per la ricodifica dei valori. Per esempio, è possibile impostare il valore su 1 se Drug='drugA', 2 se Drug='drugB' e così via.

Nodo Riempimento

I nodi Riempimento sono utilizzati per sostituire i valori di campo e modificare la classe di archiviazione. È possibile scegliere di sostituire i valori in base a una condizione CLEM specificata, per esempio @BLANK(FIELD). In alternativa, si può scegliere di sostituire tutti i valori nulli o vuoti con un valore specifico. I nodi Riempimento sono utilizzati spesso insieme al nodo Tipo per sostituire i valori mancanti. Per esempio, è possibile riempire i vuoti con il valore medio di un campo, specificando un'espressione @GLOBAL_MEAN. L'espressione riempirà tutti i vuoti con il valore medio calcolato da un nodo Globali.

Figura 4-53
Finestra di dialogo del nodo Riempimento



Riempi campi Tramite Selezione campi (pulsante a destra del campo di testo), selezionare i campi dall'insieme di dati i cui valori verranno esaminati e sostituiti. Per default, i valori vengono sostituiti in base alle espressioni Condizione e Sostituisci con specificate sotto. Le opzioni Sostituisci riportate sotto consentono di selezionare un metodo di sostituzione alternativo.

Nota: quando si selezionano più campi da sostituire con un valore definito dall'utente, è importante che i tipi di campo siano simili (tutti numerici o tutti simbolici).

Sostituisci. Selezionare uno dei seguenti metodi per sostituire i valori dei campi selezionati:

- **In base alla condizione.** Questa opzione attiva il campo Condizione e il Generatore di espressioni per consentire all'utente di creare un'espressione utilizzata come condizione per effettuare la sostituzione con il valore specificato.
- **Sempre.** Sostituisce tutti i valori del campo selezionato. Per esempio, questa opzione potrebbe essere utilizzata per convertire la classe di archiviazione del reddito in una stringa tramite la seguente espressione CLEM: (to_string(income)).
- **Valori vuoti.** Sostituisce tutti i valori vuoti specificati dall'utente nel campo selezionato. Per selezionare i valori vuoti viene utilizzata la condizione standard @BLANK(@FIELD). *Nota:* per definire i valori vuoti, utilizzare la scheda Tipi del nodo di input o un nodo Tipo.
- **Valori nulli.** Questa opzione sostituisce tutti i valori nulli di sistema nel campo selezionato. Per selezionare i valori nulli viene utilizzata la condizione standard @NULL(@FIELD).
- **Valori vuoti e nulli.** Questa opzione sostituisce sia i valori vuoti sia i nulli di sistema nel campo selezionato. È utile quando non si è sicuri se i valori nulli sono stati o meno definiti come valori mancanti.

Condizione. Questa opzione è disponibile se è stata selezionata l'opzione In base alla condizione. Utilizzare questa casella di testo per specificare un'espressione CLEM per la valutazione dei campi selezionati. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

Sostituisci con. Specificare un'espressione CLEM per assegnare un nuovo valore ai campi selezionati. Per sostituire il valore con un valore nullo, digitare undef nella casella di testo. Fare clic sul pulsante del calcolatore per aprire il Generatore di espressioni.

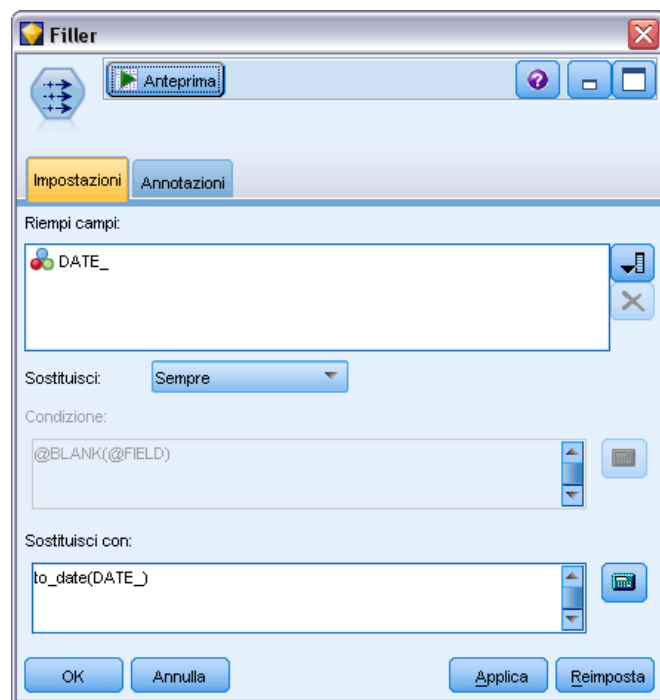
Nota: quando i campi selezionati sono stringhe, è necessario sostituirli con un valore stringa. L'utilizzo dello 0 di default o di un altro valore numerico come valore sostitutivo per i campi stringa determinerà un errore.

Conversione della classe di archiviazione mediante il nodo Riempimento

La condizione Sostituisci di un nodo Riempimento consente di convertire facilmente la classe di archiviazione per uno o più campi. Per esempio, utilizzando la funzione di conversione `to_integer`, è possibile convertire `income` da stringa in numero intero tramite la seguente espressione CLEM: `to_integer(income)`.

Figura 4-54

Utilizzo di un nodo Riempimento per la conversione della classe di archiviazione dei campi



È possibile visualizzare le funzioni di conversione disponibili e generare automaticamente un'espressione CLEM utilizzando il Generatore di espressioni. Dall'elenco a discesa delle funzioni, selezionare Conversione per visualizzare l'elenco delle funzioni di conversione della classe di archiviazione. Sono disponibili le seguenti opzioni di esecuzione:

- `to_integer(ITEM)`

- `to_real(ITEM)`
- `to_number(ITEM)`
- `to_string(ITEM)`
- `to_time(ITEM)`
- `to_timestamp(ITEM)`
- `to_date(ITEM)`
- `to_datetime(ITEM)`

Conversione dei valori di data e ora. Si noti che le funzioni di conversione, e qualsiasi altra funzione che richieda un tipo di input specifico quale un valore di data o di ora, dipendono dai formati correnti specificati nella finestra di dialogo Opzioni stream. Per esempio, se si desidera convertire un campo stringa con valori *Gen 2003*, *Feb 2003* e così via, per l'archiviazione dei dati, selezionare MES AAAA come formato di data di default per lo stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Le funzioni di conversione sono disponibili anche dal nodo Nuovo campo per la conversione temporanea durante un calcolo di derivazione. È inoltre possibile utilizzare il nodo Nuovo campo per eseguire altre manipolazioni, quali ricodifiche di campi stringa con valori categoriali. [Per ulteriori informazioni, vedere l'argomento Ricodifica di valori con il nodo Nuovo campo a pag. 165.](#)

Nodo Anonimizza

Il nodo Anonimizza consente di mascherare i nomi dei campi, i valori dei campi o entrambi, quando si utilizzano dati da includere in un modello a valle del nodo. In questo modo, il modello generato può essere distribuito liberamente, per esempio al supporto tecnico, senza rischiare che i dati riservati, quali record dei dipendenti o anamnesi dei pazienti, possano essere visualizzati da utenti non autorizzati.

A seconda del punto dello stream in cui il nodo Anonimizza viene situato, potrebbe essere necessario apportare modifiche ad altri nodi. Per esempio, se si inserisce un nodo Anonimizza a monte di un nodo Seleziona, i criteri di selezione di quest'ultimo dovranno essere modificati se agiscono su valori che sono stati anonimizzati.

Il metodo da utilizzare per l'anonimizzazione dipende da diversi fattori. Per i nomi dei campi e tutti i valori di campo ad eccezione di Livelli di misurazione continui, i dati vengono sostituiti da una stringa con il seguente formato:

prefix_Sn

dove *prefix_* è una stringa specificata dall'utente o la stringa di default *anon_* e *n* è un valore intero che parte da 0 e viene incrementato per ciascun valore univoco (per esempio, *anon_S0*, *anon_S1* e così via).

I valori di campo di tipo Continuo devono essere trasformati, poiché gli intervalli numerici si riferiscono a valori interi o reali e non a stringhe. In tal senso, possono essere anonimizzati solo mediante la trasformazione dell'intervallo in un intervallo diverso, vale a dire, mascherando i dati originali. La trasformazione di un valore x nell'intervallo viene eseguita nel modo seguente:

$$A*(x + B)$$

dove:

A è un fattore di scala che deve essere maggiore di 0.

B è un valore di traslazione da aggiungere ai valori.

Esempio

Nel caso di un campo $ETÀ$ in cui il fattore di scala A è impostato su 7 e il valore di traslazione B è impostato su 3, i valori di $ETÀ$ vengono trasformati in:

$$7*(ETÀ + 3)$$

Impostazione delle opzioni relative al nodo Anonimizza

A questo punto è possibile scegliere i campi i cui valori dovranno risultare mascherati a valle.

Si noti che i campi di dati devono essere istanziati a monte del nodo Anonimizza prima che possano essere eseguite le operazioni di anonimizzazione. Istanziare i dati facendo clic sul pulsante Leggi valori di un nodo Tipo o nella scheda Tipi di un nodo di input.

Figura 4-55
Impostazione delle opzioni di anonimizzazione



Campo. Elenca i campi nell'insieme di dati corrente. Se alcuni dei campi sono già stati anonimizzati, i relativi nomi anonimizzati vengono visualizzati qui.

Misurazione. Il livello di misurazione del campo.

Anonimizza valori. Selezionare uno o più campi, fare clic sulla colonna e scegliere Sì per anonimizzare il valore dei campi utilizzando il prefisso di default anon_; scegliere Specifica per visualizzare una finestra di dialogo in cui immettere il proprio prefisso o, nel caso dei valori di campi di tipo *Continuo*, specificare se per la trasformazione dei valori dei campi è necessario utilizzare valori casuali o specificati dall'utente. Si noti che i tipi di campo *Continuo* e non *Continuo* non possono essere specificati nell'ambito della stessa operazione. L'operazione va eseguita separatamente per ciascun tipo di campo.

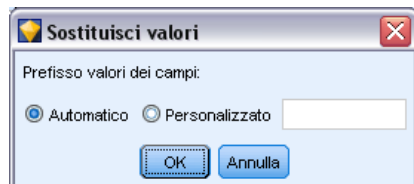
Visualizza campi correnti. Selezionare questa opzione per visualizzare campi per insiemi di dati attivi connessi al nodo Anonimizza. Questa opzione è selezionata per default.

Visualizza impostazioni campi inutilizzati. Selezionare questa opzione per visualizzare campi di insiemi di dati non più connessi al nodo. Questa opzione è utile quando si copiano nodi da uno stream all'altro o quando si salvano e si ricaricano nodi.

Specifica del metodo di anonimizzazione dei valori di campo

La finestra di dialogo Sostituisci valori consente di scegliere se utilizzare il prefisso di default o un prefisso personalizzato per i valori di campo anonimizzati. Se si fa clic su OK in questa finestra di dialogo, l'impostazione di Anonimizza valori nella scheda Impostazioni viene modificata in Sì per il campo o i campi selezionati.

Figura 4-56
Finestra di dialogo Sostituisci valori



Prefisso dei valori di campo. Il prefisso di default per i valori di campo anonimizzati è anon_; se si preferisce utilizzare un prefisso diverso, scegliere Personalizzato e digitare il prefisso.

La finestra di dialogo Trasforma valori viene visualizzata solo per i campi di tipo Continuo e consente di specificare se per la trasformazione dei valori dei campi devono essere utilizzati valori casuali o valori specificati dall'utente.

Figura 4-57
Finestra di dialogo *Trasforma valori*



Casuale. Scegliere questa opzione se si desidera utilizzare valori casuali per la trasformazione. L'opzione *Imposta seme aleatorio* è selezionata per default; specificare un valore nel campo *Seme* oppure utilizzare il valore di default.

Fissa. Scegliere questa opzione se si desidera specificare i propri valori per la trasformazione.

- **Scala per.** Il numero per il quale i valori dei campi verranno moltiplicati durante la trasformazione. Il valore minimo è 1 e il valore massimo solitamente è 10, ma può essere ridotto per evitare l'overflow.
- **Transla per.** Il numero che verrà aggiunto ai valori dei campi durante la trasformazione. Il valore minimo è 0 e il valore massimo solitamente è 1000, ma può essere ridotto per evitare l'overflow.

Anonimizzazione dei valori dei campi

I campi che sono stati selezionati per l'anonimizzazione nella scheda *Impostazioni* subiscono l'anonimizzazione dei valori:

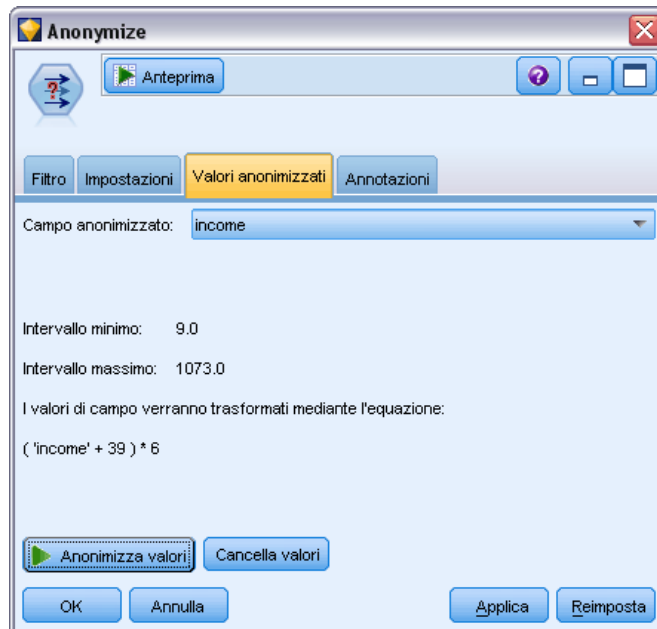
- Quando si esegue lo stream contenente il nodo *Anonimizza*
- Quando si visualizzano i valori in anteprima

Per visualizzare i valori in anteprima, fare clic sul pulsante *Anonimizza valori* nella scheda *Valori anonimizzati*. Selezionare quindi un nome di campo dall'elenco a discesa.

Se il livello di misurazione è *Continuo*, vengono visualizzati:

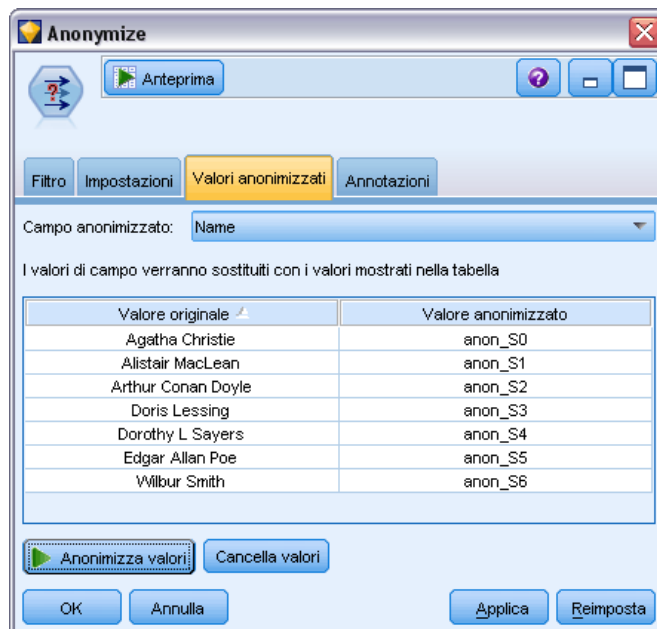
- Valori minimo e massimo dell'intervallo originale
- Equazione utilizzata per trasformare i valori

Figura 4-58
Anonimizzazione dei valori dei campi



Se il livello di misurazione è diverso da Continuo, vengono visualizzati i valori originale e anonimizzato per tale campo.

Figura 4-59
Anonimizzazione dei valori dei campi



Quando la visualizzazione presenta uno sfondo giallo, significa che l'impostazione del campo selezionato è stata modificata dopo l'ultima anonimizzazione dei valori oppure che sono state apportate delle modifiche ai dati a monte del nodo Anonimizza e i valori anonimizzati potrebbero non essere più corretti. Viene visualizzato l'insieme di valori corrente; fare clic nuovamente sul pulsante Anonimizza valori per generare un nuovo insieme di valori in base all'impostazione corrente.

Anonimizza valori. Consente di creare valori anonimizzati per il campo selezionato e di visualizzarli in una tabella. Se si utilizza un seme aleatorio per un campo di tipo Continuo, ogni volta che si fa clic su questo pulsante viene creato un insieme di valori diverso.

Cancella valori. Cancella i valori originali e anonimizzati dalla tabella.

Nodo Ricodifica

Il nodo Ricodifica consente la trasformazione da un insieme di valori categoriali a un altro. L'operazione di ricodifica consente di comprimere categorie o raggruppare dati per l'analisi. Per esempio, è possibile ricodificare i valori per *Prodotto* in tre gruppi, quali *Stoviglie*, *Biancheria* ed *Elettrodomestici*. Questa operazione viene spesso eseguita direttamente da un nodo Distribuzione raggruppando valori e generando un nodo Ricodifica. [Per ulteriori informazioni, vedere l'argomento Utilizzo di un nodo Distribuzione in il capitolo 5 a pag. 264.](#)

La ricodifica può essere eseguita per uno o più campi simbolici. È anche possibile scegliere di sostituire i nuovi valori per il campo esistente o generare un nuovo campo.

Prima di utilizzare un nodo Ricodifica, valutare se un altro nodo Operazioni su campi può risultare più appropriato all'attività in corso:

- Per trasformare intervalli numerici in insiemi utilizzando un metodo automatico, come i ranghi o i percentili, è necessario utilizzare un nodo Discretizza. [Per ulteriori informazioni, vedere l'argomento Nodo Discretizza a pag. 178.](#)
- Per codificare manualmente intervalli numerici in insiemi, è necessario utilizzare un nodo Nuovo campo. Se, per esempio, si desidera comprimere i valori di salario in specifiche categorie di intervalli di salari, è necessario utilizzare un nodo Nuovo campo per definire manualmente ogni categoria.
- Per creare uno o più campi flag basati sui valori di un campo categoriale, per esempio *Tipo_mutuo*, è necessario utilizzare un nodo Crea flag.
- Per convertire un campo categoriale in campo di archiviazione numerica è possibile utilizzare un nodo Nuovo campo. Per esempio, è possibile convertire i valori *No* e *Sì* rispettivamente in 0 e 1. [Per ulteriori informazioni, vedere l'argomento Ricodifica di valori con il nodo Nuovo campo a pag. 165.](#)

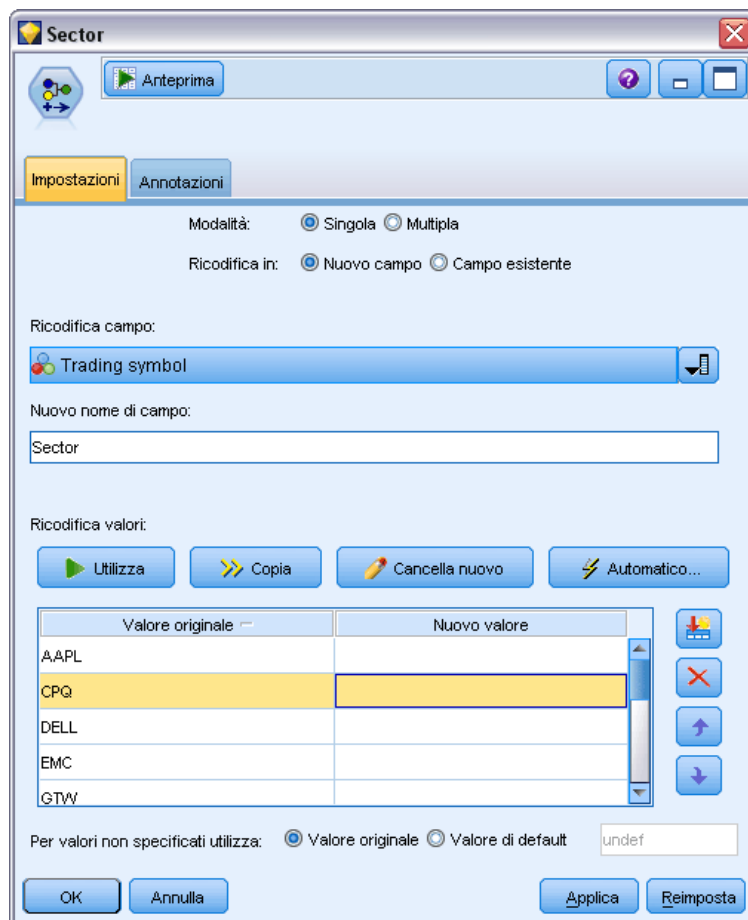
Impostazione delle opzioni per il nodo Ricodifica

Per utilizzare il nodo Ricodifica è necessario seguire tre passaggi:

- ▶ Innanzitutto, selezionare se si desidera ricodificare uno o più campi.
- ▶ Quindi, scegliere se eseguire la ricodifica nel campo esistente o creare un nuovo campo.

- Infine, utilizzare le opzioni dinamiche nella finestra di dialogo del nodo Ricodifica per mappare gli insiemi in base alle esigenze.

Figura 4-60
Finestra di dialogo Nodo Ricodifica



Modalità. Selezionare Singola per ricodificare le categorie per un campo. Selezionare Multipla per attivare le opzioni che consentono la trasformazione di più campi contemporaneamente.

Ricodifica in. Selezionare Nuovo campo per utilizzare il campo nominale originale e derivare un campo aggiuntivo contenente i valori ricodificati. Selezionare Campo esistente per sovrascrivere i valori nel campo originale con le nuove classificazioni. In questo caso si tratta essenzialmente di un'operazione di riempimento.

Dopo aver specificato le opzioni di sostituzione e relative alla modalità, è necessario selezionare il campo di trasformazione e specificare i nuovi valori di classificazione utilizzando le opzioni dinamiche nella parte inferiore della finestra di dialogo. Tali opzioni variano in relazione alla modalità selezionata.

Ricodifica campo (campi). Utilizzare il pulsante Selezione campi a destra per selezionare uno (modalità Singola) o più (modalità Multipla) campi categoriali.

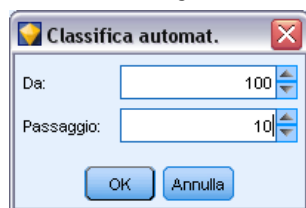
Nome nuovo campo. Specificare un nome per il nuovo campo nominale contenente i valori ricodificati. Questa opzione è disponibile solo in modalità Singola se è stato selezionato Nuovo campo. Se si seleziona Campo esistente, il nome originale del campo viene mantenuto. Se si opera in modalità Multipla, questa opzione viene sostituita con controlli per specificare l'aggiunta di un'estensione a ogni nuovo campo. [Per ulteriori informazioni, vedere l'argomento Ricodifica di più campi a pag. 177.](#)

Ricodifica valori. Questa tabella consente di specificare una nuova mappatura tra i valori di insieme preesistenti e quelli specificati in questo contesto.

- **Valore originale.** In questa colonna sono elencati i valori esistenti per i campi selezionati.
 - **Nuovo valore.** Utilizzare questa colonna per digitare nuovi valori di categoria o selezionarne uno dall'elenco a discesa. Se si genera automaticamente un nodo Ricodifica utilizzando i valori di un grafico di distribuzione, tali valori verranno inseriti nell'elenco a discesa. Ciò consente di mappare rapidamente i valori esistenti su un insieme di valori noto. Per esempio, le organizzazioni sanitarie a volte raggruppano le diagnosi in modi diversi, in base alla rete o alle opzioni internazionali. Dopo una fusione o un'acquisizione, a tutte le parti verrà richiesto di ricodificare in modo uniforme i nuovi dati o i dati esistenti. Anziché reinserire manualmente tutti i valori di destinazione di un lungo elenco, è possibile leggere l'elenco principale di valori in IBM® SPSS® Modeler, eseguire un grafico di distribuzione per il campo *Diagnosi* e generare un nodo Ricodifica (valori) per il campo direttamente dal grafico. In questo modo tutti i valori di destinazione Diagnosi saranno disponibili nell'elenco a discesa Nuovi valori.
- ▶ Fare clic su Utilizza per leggere i valori originali di uno o più campi selezionati in precedenza.
 - ▶ Fare clic su Copia per incollare i valori originali nella colonna *Nuovo valore* per i campi che non sono ancora stati mappati. I valori originali non mappati verranno aggiunti all'elenco a discesa.
 - ▶ Fare clic su Cancella nuovo per cancellare tutte le specifiche presenti nella colonna *Nuovo valore*. *Nota:* questa opzione non cancella i valori dall'elenco a discesa.
 - ▶ Fare clic su Automatico per generare automaticamente numeri interi consecutivi per ogni valore originale. Possono essere generati solo valori interi (nessun valore reale, come 1.5, 2.5 e così via).

Figura 4-61

Finestra di dialogo di classificazione automatica



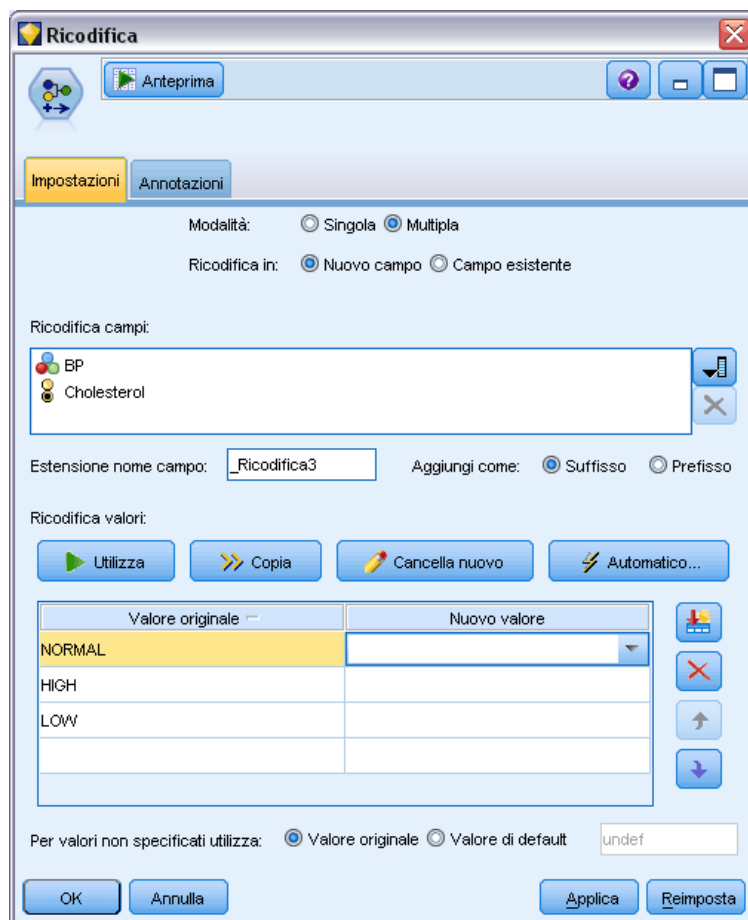
Per esempio, è possibile generare automaticamente numeri ID di prodotto consecutivi per i nomi di prodotto o numeri dei corsi per le offerte di un'università. Questa funzionalità corrisponde alla trasformazione di ricodifica automatica degli insiemi in IBM® SPSS® Statistics.

Per valori non specificati utilizza. Questa opzione consente di immettere valori non specificati nel nuovo campo. È possibile scegliere di mantenere il valore originale selezionando Valore originale oppure specificare un valore di default.

Ricodifica di più campi

Per mappare i valori di categoria per più campi contemporaneamente, impostare la modalità Multipla. In questo modo vengono attivate nuove impostazioni nella finestra di dialogo Ricodifica, descritte di seguito.

Figura 4-62
Opzioni della finestra di dialogo Dinamico per la ricodifica di più campi



Ricodifica campi. Utilizzare il pulsante Selezione campi a destra per selezionare i campi che si desidera trasformare. In questo modo è possibile selezionare tutti i campi contemporaneamente oppure campi di un determinato tipo, per esempio nominale o flag.

Estensione nome campo. Quando si ricodificano più campi contemporaneamente, è consigliabile specificare un'estensione comune da aggiungere a tutti i nuovi campi anziché singoli nomi di campo. Specificare un'estensione quale `_recode` e selezionare se inserirla come prefisso o suffisso ai nomi originali dei campi.

Classe di archiviazione e livello di misurazione per i campi ricodificati

Il nodo Ricodifica crea sempre un campo nominale dall'operazione di ricodifica. In alcuni casi, il livello di misurazione può cambiare quando si utilizza la modalità di ricodifica Campo esistente.

L'archiviazione del nuovo campo (il modo in cui vengono *archiviati* i dati anziché come vengono *utilizzati*) viene definita in base alle seguenti opzioni della scheda Impostazioni:

- Se nel caso di valori non specificati si sceglie di utilizzare un valore di default, il tipo di archiviazione appropriato viene determinato esaminando i nuovi valori e il valore di default. Se, per esempio, tutti i valori possono essere analizzati come numeri interi, per il campo verrà utilizzato il tipo di archiviazione come numero intero.
- Se nel caso di valori non specificati si sceglie di utilizzare i valori originali, il tipo di archiviazione viene basato sulla classe di archiviazione del campo originale. Se tutti i valori possono essere analizzati con la classe di archiviazione del campo originale, tale classe viene mantenuta; in caso contrario, la classe di archiviazione viene determinata individuando il tipo di archiviazione più appropriato che includa i valori nuovi e quelli preesistenti. Per esempio, se si ricodifica un insieme di numeri interi { 1, 2, 3, 4, 5 } con la ricodifica $4 \Rightarrow 0, 5 \Rightarrow 0$, viene generato un nuovo insieme di numeri interi { 1, 2, 3, 0 }, mentre la ricodifica $4 \Rightarrow$ "Maggiore di 3", $5 \Rightarrow$ "Maggiore di 3" genererà l'insieme stringa {"1", "2", "3", "Maggiore di 3"}.

Nota: se il tipo originale non era stato istanziato, verrà applicato lo stesso formato per il nuovo tipo.

Nodo Discretizza

Il nodo Discretizza consente di creare automaticamente nuovi campi nominali basati sui valori di uno o più campi continui (intervallo numerico). Per esempio, è possibile trasformare un campo continuo relativo al reddito in campo categoriale contenente gruppi di reddito di larghezza equivalente o come deviazioni dalla media. In alternativa, è possibile selezionare un campo "supervisore" categoriale per mantenere la forza dell'associazione originale tra i due campi.

La discretizzazione può essere utile per diversi motivi, fra cui:

- **Requisiti degli algoritmi.** Alcuni algoritmi, come Bayes naive e Regressione logistica, richiedono input categoriali.
- **Prestazioni.** Alcuni algoritmi, come l'algoritmo logistico multinomiale, possono avere migliori performance se viene ridotto il numero dei valori distinti dei campi di input. Per esempio, utilizzare il valore mediano o medio per ogni bin invece dei valori originali.
- **Privacy dei dati.** Le informazioni personali riservate quali, per esempio, lo stipendio, possono essere riportate sotto forma di intervalli invece che di cifre effettive, in modo da proteggere la privacy.

Sono disponibili diversi metodi di discretizzazione. Dopo avere creato bin per il nuovo campo, è possibile generare un nodo Nuovo campo basato sui punti di taglio.

Prima di utilizzare un nodo Discretizza, valutare se un'altra tecnica potrebbe essere più appropriata per l'attività in corso:

- Per specificare manualmente punti di taglio per le categorie, come specifici intervalli di salario predefiniti, utilizzare un nodo Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Nuovo campo a pag. 154.](#)
- Per creare nuove categorie per insiemi esistenti, utilizzare un nodo Ricodifica. [Per ulteriori informazioni, vedere l'argomento Nodo Ricodifica a pag. 174.](#)

Gestione dei valori mancanti

Il nodo Discretizza gestisce i valori mancanti come segue:

- **Vuoti definiti dall'utente.** I valori mancanti specificati come vuoti vengono inclusi durante la trasformazione. Per esempio, se si specifica -99 per indicare un valore vuoto mediante il nodo Tipo, questo valore verrà incluso nel processo di discretizzazione. Per ignorare i vuoti durante la discretizzazione, utilizzare un nodo Riempimento per sostituire i valori vuoti con il valore nullo di sistema.
- **Valori mancanti di sistema (\$null\$).** I valori nulli vengono ignorati durante la discretizzazione e restano nulli dopo la trasformazione.

Nella scheda Impostazioni sono incluse le opzioni per le tecniche disponibili. Nella scheda Visualizza sono riportati i punti di taglio stabiliti per i dati eseguiti precedentemente attraverso il nodo.

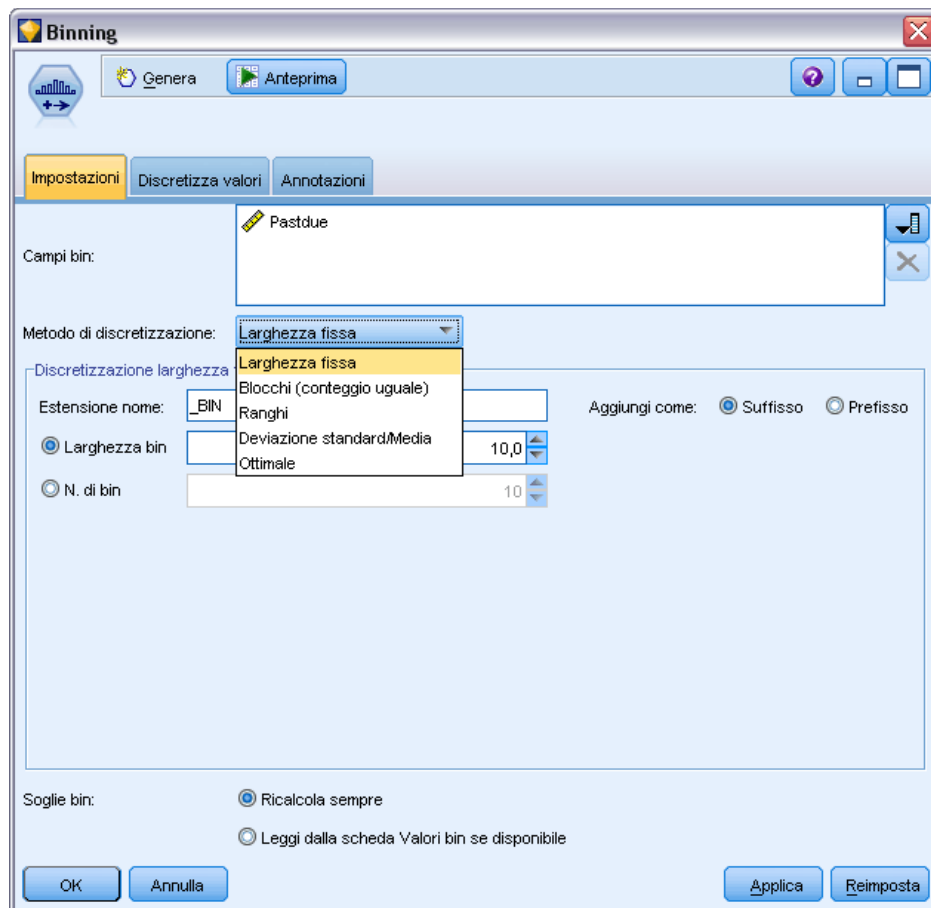
Impostazione delle opzioni relative al nodo Discretizza

Il nodo Discretizza consente di generare automaticamente bin (categorie) utilizzando le seguenti tecniche:

- Discretizzazione larghezza fissa
- Blocchi (conteggio o somma uguale)
- Media e deviazione standard
- Ranghi
- Ottimizzazione relativa a un campo "supervisore" categoriale

La parte inferiore della finestra cambia in modo dinamico in relazione al metodo di discretizzazione selezionato.

Figura 4-63
Finestra di dialogo del nodo Discretizza, scheda Impostazioni



Campi bin. Le trasformazioni in sospeso per i campi continui (intervallo numerico) sono visualizzate in questo punto. Il nodo Discretizza consente di eseguire la discretizzazione di più campi simultaneamente. Aggiungere o rimuovere campi mediante i pulsanti a destra.

Metodo di discretizzazione. Selezionare il metodo utilizzato per determinare i punti di taglio per i nuovi bin di campo (categorie).

Soglie bin. Specificare se i punti di taglio e le allocazioni dei bin vengono sempre rielaborati all'esecuzione del nodo, o se vengono elaborati solo quando è necessario (per esempio in occasione dell'aggiunta di nuovi dati). Se si seleziona Leggi dalla scheda Valori bin se disponibile è possibile modificare i punti di taglio superiore e inferiore dei vari bin nella scheda Discretizza valori.

I metodi di discretizzazione disponibili sono esaminati negli argomenti seguenti.

Bin a larghezza fissa

Quando si sceglie Larghezza fissa come metodo di discretizzazione, viene visualizzato un nuovo insieme nella finestra di dialogo.

Figura 4-64

Finestra di dialogo del nodo Discretizza (scheda Impostazioni) con le opzioni relative ai bin a larghezza fissa

Estensione nome. Specificare un'estensione da utilizzare per i campi generati. *_BIN* è l'estensione di default. È inoltre possibile specificare se l'estensione verrà aggiunta all'inizio (Prefisso) o alla fine (Suffisso) del nome del campo. Per esempio, è possibile generare un nuovo campo denominato *reddito_BIN*.

Larghezza bin. Specificare un valore (numero intero o reale) utilizzato per calcolare la "larghezza" del bin. Per esempio, è possibile utilizzare il valore di default, 10, per discretizzare il campo *Età*. Poiché al campo *Età* è assegnato un intervallo 18–65, i bin generati saranno i seguenti:

Tabella 4-1

Bin per *Età* con intervallo 18–65

Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
>=da 13 a <23	>=da 23 a <33	>=da 33 a <43	>=da 43 a <53	>=da 53 a <63	>=da 63 a <73

L'inizio dell'intervallo dei bin viene calcolato sottraendo la metà della larghezza del bin al valore più basso rilevato (come indicato). Nei bin illustrati sopra, per esempio, il valore 13 viene utilizzato per iniziare gli intervalli in base al seguente calcolo: $18 [\text{valore più basso}] - 5 [0.5 \times (\text{Larghezza del bin di } 10)] = 13$.

N. di bin. Questa opzione consente di specificare un numero intero utilizzato per determinare il numero di bin a larghezza fissa (categorie) per i nuovi campi.

Dopo aver eseguito il nodo Discretizza in uno stream, è possibile visualizzare le soglie dei bin generate facendo clic sulla scheda Anteprema nella finestra di dialogo del nodo Discretizza. [Per ulteriori informazioni, vedere l'argomento Anteprema dei bin generati a pag. 188.](#)

Blocchi (conteggio o somma uguale)

Il metodo di discretizzazione a blocchi crea dei campi nominali che possono essere utilizzati per suddividere i record analizzati in gruppi percentili (o quartili, decili e così via) in modo che ogni gruppo contenga lo stesso numero di record o che la somma dei valori di ciascun gruppo sia uguale. I record vengono classificati in ordine crescente a seconda del valore del campo bin specificato, in modo tale che ai record con i valori più bassi per la variabile bin selezionata viene assegnato il rango 1, all'insieme di record seguente il rango 2 e così via. I valori di soglia di ogni bin vengono generati automaticamente in base ai dati e al metodo di discretizzazione a blocchi utilizzato.

Figura 4-65

Finestra di dialogo del nodo Discretizza (scheda Impostazioni) con le opzioni relative ai bin con conteggio uguale

Discretizzazione conteggio uguale

Estensione nome blocco: Aggiungi come: Suffisso Prefisso

Nota: il conteggio del blocco verrà aggiunto alla fine dell'estensione

Estensione blocco personalizzata: Aggiungi come: Suffisso Prefisso

Quartile (4) Quintile (5) Decile (10)

Ventile (20) Percentile (100)

N personalizzato

Metodo di espansione: Conteggio dei record Somma dei valori

Pari merito: Aggiungi a successivo Mantieni in corrente Assegna casualmente

Estensione nome blocco. Specificare un'estensione da utilizzare per i campi generati con p-tili standard. L'estensione di default è `_TILE` preceduta da N , dove N è il numero percentile. È inoltre possibile specificare se l'estensione verrà aggiunta all'inizio (Prefisso) o alla fine (Suffisso) del nome del campo. Per esempio, è possibile generare un nuovo campo denominato `reddito_BIN4`.

Estensione blocco personalizzata. Specificare un'estensione da utilizzare per l'intervallo di blocchi personalizzato. L'impostazione di default è `_TILEN`. Si noti che, in questo caso, N non verrà sostituito dal numero personalizzato.

I p-tili disponibili sono:

- **Quartile.** Genera 4 bin, ognuno contenente il 25% dei casi.
- **Quintile.** Genera 5 bin, ognuno contenente il 20% dei casi.
- **Decile.** Genera 10 bin, ognuno contenente il 10% dei casi.
- **Ventile.** Genera 20 bin, ognuno contenente il 5% dei casi.
- **Percentile.** Genera 100 bin, ognuno contenente il 1% dei casi.
- **N personalizzato.** Selezionare questa opzione per specificare il numero di bin. Per esempio, il valore 3 produrrebbe tre categorie suddivise (due punti di taglio), ognuna contenente il 33.3% dei casi.

Si noti che se nei dati è presente un numero di valori discreti inferiore al numero di blocchi specificati, non tutti i blocchi saranno utilizzati. In tali casi, la nuova distribuzione risulterà simile a quella originale dei dati.

Metodo di espansione. Consente di specificare il metodo utilizzato per assegnare i record ai bin.

- **Conteggio dei record.** Cerca di assegnare un numero di record uguale a ogni bin.
- **Somma dei valori.** Cerca di assegnare i record ai bin in modo che la somma dei valori di ogni bin sia uguale. Per esempio, nella targetizzazione dei tentativi di vendita, questo metodo può essere utilizzato per assegnare i potenziali clienti a gruppi di decili in base al valore di ogni record, con i potenziali clienti di valore massimo nel primo bin. Per esempio un'azienda farmaceutica potrebbe classificare i medici in gruppi di decili in base al numero di ricette

compilate. Ogni decile conterrebbe più o meno lo stesso numero di ricette, ma il numero di individui che genera le ricette non sarebbe uguale, poiché gli individui che compilano più ricette sono concentrati nel decile 10. Questo approccio presuppone che tutti i valori siano maggiori di zero e in caso contrario può generare risultati inattesi.

Pari merito. Se i valori a entrambi i lati di un punto di taglio sono identici, si ottiene una condizione di pari merito. Per esempio, nel caso di assegnazione di decili in cui più del 10% dei record ha lo stesso valore per il campo bin, non è possibile che tutti i record possano rientrare nello stesso bin senza forzare in qualche modo la soglia. I record a pari merito possono essere spostati al bin successivo o mantenuti in quello corrente, ma è necessario che la condizione venga risolta affinché tutti i record con valori identici rientrino nello stesso bin, anche se ciò potrebbe comportare per alcuni bin un numero di record superiore al previsto. Di conseguenza, è possibile che anche le soglie dei bin successivi vengano adeguate e che i valori siano assegnati in modo diverso per lo stesso insieme di numeri in base al metodo usato per risolvere le condizioni di pari merito.

- **Aggiungi a successivo.** Selezionare questa opzione per spostare i valori pari merito nel bin successivo.
- **Mantieni in corrente.** Mantiene i valori pari merito nel bin corrente (più basso). Questo metodo può determinare la creazione di un numero totale di bin inferiore.
- **Assegna casualmente.** Selezionare questa opzione per allocare in modo casuale i valori pari merito a un bin. In questo modo si tenta di mantenere uguale il numero di record di ogni bin.

Esempio: Espansione per conteggio dei record

La tabella che segue illustra il modo in cui i valori di campo semplificati vengono classificati come quartili con l'espansione per conteggio dei record. Si noti che i risultati possono variare a seconda dell'opzione di pari merito selezionata.

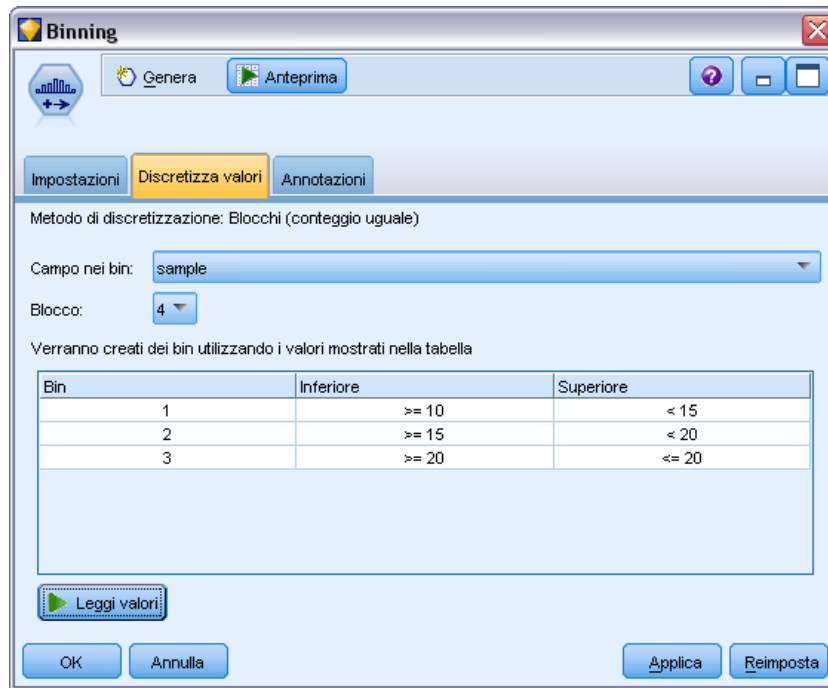
Valori	Aggiungi a successivo	Mantieni in corrente
10	1	1
13	2	1
15	3	2
15	3	2
20	4	3

Il numero di elementi per bin viene calcolato come segue:

numero totale del valore/numero di blocchi

Nel precedente esempio semplificato il numero desiderato di elementi per bin è 1,25 (5 valori/4 quartili). Il valore 13 (ossia il valore numero 2) si trova a cavallo della soglia di conteggio desiderata 1,25 e, quindi, viene trattato diversamente in relazione all'opzione di pari merito selezionata. In modalità Aggiungi a successivo questo valore viene aggiunto al bin 2, mentre in modalità Mantieni in corrente viene lasciato nel bin 1, spostando l'intervallo di valori del bin 4 all'esterno dei valori di dati esistenti. Di conseguenza vengono creati solo tre bin e le soglie di ogni bin vengono opportunamente regolate.

Figura 4-66
Soglie dei bin generati

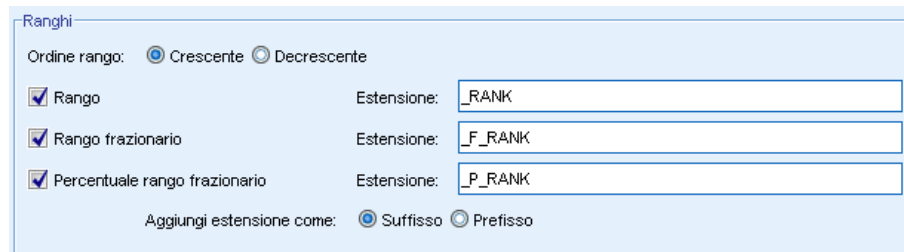


Nota: l'attivazione dell'elaborazione parallela può accelerare il processo di discretizzazione a blocchi. Per ulteriori informazioni, vedere l'argomento [Impostazione delle opzioni di ottimizzazione](#) in il capitolo 12 in *Manuale dell'utente di IBM SPSS Modeler 14.2*.

Classifica casi

Se si sceglie Ranghi come metodo di discretizzazione, nella finestra di dialogo viene visualizzato un nuovo insieme di opzioni.

Figura 4-67
Finestra di dialogo del nodo Discretizza (scheda Impostazioni) con le opzioni relative ai ranghi



L'assegnazione di ranghi crea nuovi campi contenenti ranghi, ranghi frazionari e valori percentili per i campi numerici in base alle opzioni specificate di seguito.

Ordine rango. Selezionare *Crescente* (il valore più basso viene indicato con 1) o *Decrescente* (il valore più alto viene indicato con 1).

Rango. Selezionare il tipo di rango in ordine crescente o decrescente, come indicato in precedenza. L'intervallo dei valori nel nuovo campo corrisponderà a $1-N$, dove N è il numero dei valori discreti nel campo originale. Ai valori pari merito viene assegnata la media del rispettivo rango.

Rango frazionario. Selezionare i tipi di rango in cui il valore del nuovo campo equivale al rango diviso per la somma dei pesi dei casi non mancanti. I ranghi frazionari sono compresi nell'intervallo 0–1.

Percentuale rango frazionario. Ogni rango è diviso per il numero di record con valori validi e moltiplicato per 100. I ranghi frazionari in percentuale sono compresi nell'intervallo 1–100.

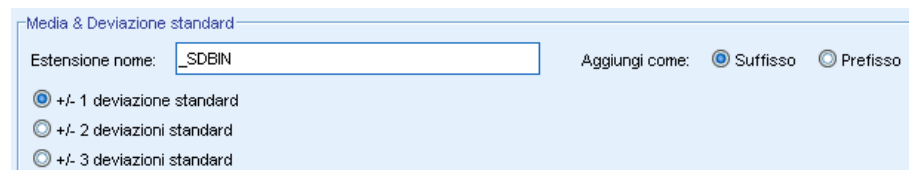
Estensione. Per tutte le opzioni dei ranghi è possibile creare estensioni personalizzate e specificare se l'estensione verrà aggiunta all'inizio (Prefisso) o alla fine (Suffisso) del nome del campo. Per esempio, è possibile generare un nuovo campo denominato *reddito_P_RANGO*.

Deviazione standard/Media

Se si sceglie Deviazione standard/Media come metodo di discretizzazione, nella finestra di dialogo viene visualizzato un nuovo insieme di opzioni.

Figura 4-68

Finestra di dialogo del nodo Discretizza (scheda Impostazioni) con le opzioni relative a deviazione standard/media



Questo metodo genera uno o più campi nuovi con categorie suddivise in base ai valori della deviazione standard e media della distribuzione dei campi specificati. Selezionare il numero di deviazioni da utilizzare di seguito.

Estensione nome. Specificare un'estensione da utilizzare per i campi generati. *_SDBIN* è l'estensione di default. È inoltre possibile specificare se l'estensione verrà aggiunta all'inizio (Prefisso) o alla fine (Suffisso) del nome del campo. Per esempio, è possibile generare un nuovo campo denominato *reddito_SDBIN*.

- **+/- 1 deviazione standard.** Selezionare questa opzione per generare tre bin.
- **+/- 2 deviazione standard.** Selezionare questa opzione per generare cinque bin.
- **+/- 3 deviazione standard.** Selezionare questa opzione per generare sette bin.

Per esempio, se si seleziona +/-1 deviazione standard, si ottengono tre bin calcolati come segue:

Bin 1	Bin 2	Bin 3
$x < (\text{Mean} - \text{Std. Dev})$	$(\text{Mean} - \text{Std. Dev}) \leq x \leq (\text{Mean} + \text{Std. Dev})$	$x > (\text{Mean} + \text{Std. Dev})$

In una distribuzione normale, il 68% dei casi è compreso in una deviazione standard della media, il 95% in due deviazioni standard e il 99% in tre deviazioni standard. Tuttavia, si tenga presente che la creazione di categorie segmentate in base a deviazioni standard può determinare

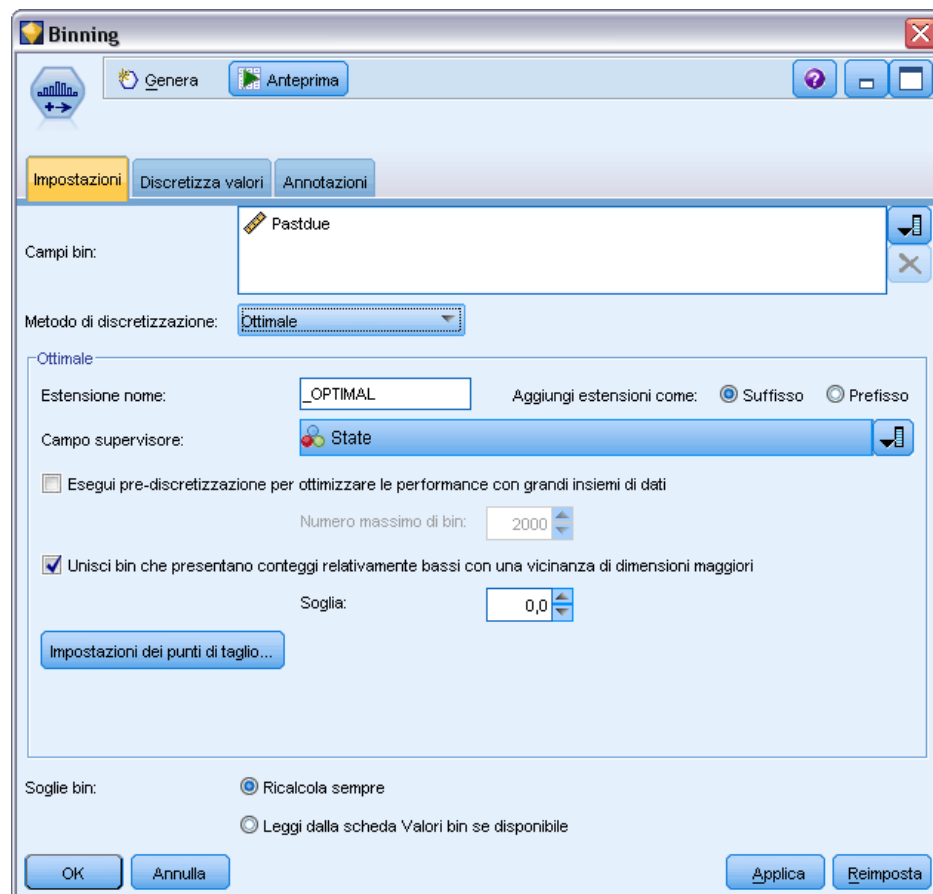
la definizione di alcuni bin all'esterno dell'effettivo intervallo di dati e anche dell'intervallo di possibili valori di dati (per esempio, un intervallo di salari negativo).

Categorizzazione ottimale

Se il campo da discretizzare è fortemente associato a un altro campo categoriale, è possibile selezionare il campo categoriale come campo "supervisore" al fine di creare i bin in modo che venga mantenuta la forza dell'associazione originale tra i due campi.

Per esempio, si supponga di avere utilizzato l'analisi dei cluster per raggruppare diversi stati o regioni in base ai tassi di insolvenza sui mutui residenziali, con i tassi più alti raccolti nel primo cluster. In questo caso, si potrebbe scegliere *Percentuale oltre la scadenza* e *Percentuale pignoramenti* come campi bin e il campo di appartenenza al cluster generato dal modello come campo supervisore.

Figura 4-69
Opzioni per la discretizzazione ottimale o supervisionata



Estensione nome. Specificare un'estensione da utilizzare per il campo o i campi generati e indicare se si desidera aggiungerla all'inizio (Prefisso) o alla fine (Suffisso) del nome del campo. Per esempio, è possibile generare un nuovo campo denominato *scaduti_OTTIMALE* e un altro campo denominato *pignorati_OTTIMALE*.

Campo supervisore. Un campo categoriale utilizzato per creare i bin.

Esegui pre-discretizzazione per ottimizzare le performance con grandi insiemi di dati. Indica se è necessario utilizzare un'elaborazione preliminare per facilitare la discretizzazione ottimale. Il metodo raggruppa i valori di scala in un alto numero di bin utilizzando un semplice metodo di discretizzazione non supervisionato, rappresenta i valori dei singoli bin in base alla media e corregge di conseguenza il peso del caso prima di procedere alla discretizzazione supervisionata. In pratica, questo metodo privilegia la velocità rispetto alla precisione ed è consigliato per insiemi di dati di grosse dimensioni. È anche possibile specificare il numero massimo di bin di ogni variabile dopo la pre-elaborazione quando si utilizza questa opzione.

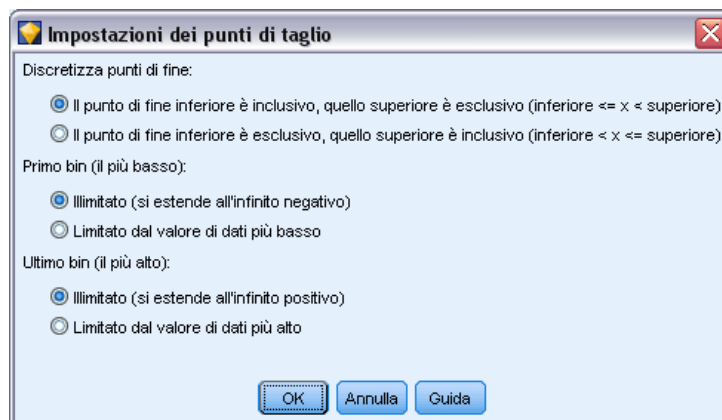
Unisci bin che presentano conteggi relativamente bassi con una vicinanza di dimensioni maggiori. Se attivata, questa opzione indica che un bin viene unito se il rapporto tra le sue dimensioni (numero di casi) e quelle di un bin adiacente è inferiore alla soglia specificata; si noti che soglie più elevate possono determinare un maggior numero di unioni.

Impostazioni dei punti di taglio

La finestra di dialogo Impostazioni dei punti di taglio consente di specificare le opzioni avanzate relative all'algorithm di discretizzazione ottimale. Queste opzioni indicano all'algorithm in che modo calcolare i bin utilizzando il campo obiettivo.

Figura 4-70

Impostazioni dei punti di taglio per discretizzazione ottimale



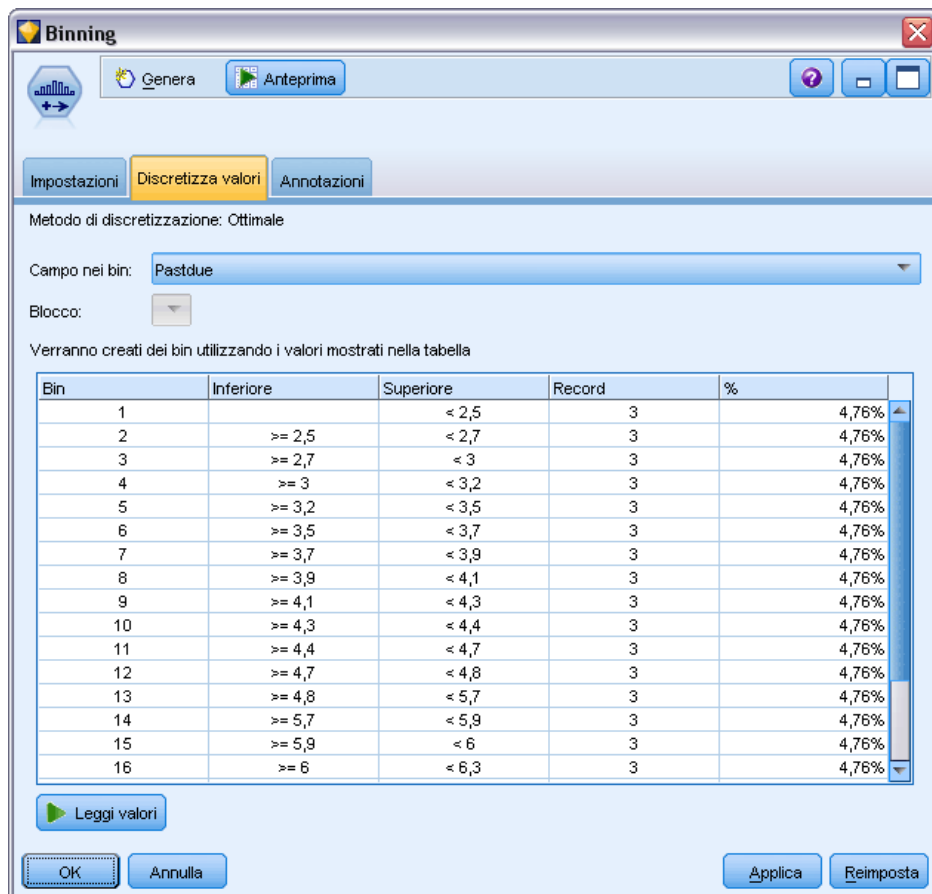
Discretizza punti di fine. Consente di specificare se i punti di fine inferiore e superiore devono essere inclusi (inferiore $\leq x$) o esclusi (inferiore $< x$).

Primo e ultimo bin. È possibile specificare se il primo e l'ultimo bin devono essere illimitati (estendersi verso l'infinito positivo o negativo) oppure limitati dai punti di dati inferiore o superiore.

Anteprima dei bin generati

La scheda Discretizza valori del nodo Discretizza consente di visualizzare le soglie per i bin generati. Mediante il menu Genera, è anche possibile creare un nodo Nuovo campo che può essere utilizzato per applicare queste soglie da un insieme di dati all'altro.

Figura 4-71
Finestra di dialogo del nodo Discretizza, scheda Discretizza valori



Campo nei bin. Utilizzare l'elenco a discesa per selezionare un campo da visualizzare. I nomi di campo mostrati utilizzano il nome originale del campo per maggiore chiarezza.

Blocco. Utilizzare l'elenco a discesa per selezionare un blocco, per esempio 10 o 100, da visualizzare. Questa opzione è disponibile solo quando i bin sono stati generati utilizzando il metodo dei blocchi con conteggio o somma uguale.

Soglie bin. Vengono qui visualizzati i valori di soglia per ogni bin generato, nonché il numero dei record che rientrano in ogni bin. Solo nel caso del metodo di discretizzazione ottimale, il numero dei record dei singoli bin viene espresso come percentuale della quantità totale. Si noti che le soglie non sono applicabili quando si utilizza il metodo di discretizzazione per ranghi.

Leggi valori. Consente di leggere i valori discretizzati dall'insieme di dati. Si noti che anche le soglie vengono sovrascritte quando nuovi dati vengono eseguiti attraverso lo stream.

Generazione di un nodo Nuovo campo

Il menu Genera consente di creare un nodo Nuovo campo sulla base delle soglie correnti. È utile per applicare soglie di bin stabilite da un insieme di dati all'altro. Inoltre, dopo aver individuato i punti di suddivisione, un'operazione Nuovo campo risulta più efficace (più rapida) di un'operazione Discretizza quando si utilizzano insiemi di dati molto estesi.

Nodo Analisi RFM

Il nodo Analisi RFM (Recency, Frequency, Monetary, Passato recente, Frequenza, Monetario) consente di determinare in modo quantitativo i clienti potenzialmente migliori verificando quanto tempo è trascorso dal loro ultimo acquisto (passato recente), con quale frequenza hanno effettuato acquisti (frequenza) e quanto hanno speso per tutte le transazioni (monetario).

L'analisi RFM si basa sull'idea che i clienti che acquistano una volta un prodotto o un servizio hanno maggiori probabilità di acquistarlo nuovamente. I dati sui clienti suddivisi in categorie vengono distribuiti in un certo numero di bin, con i criteri di discretizzazione modificati in funzione delle proprie esigenze. In ogni bin, ai clienti viene assegnato un punteggio; i punteggi vengono quindi combinati per ottenere un punteggio RFM complessivo. Tale punteggio è una rappresentazione dell'appartenenza di un cliente ai bin creati per ogni parametro RFM. Questi dati discretizzati possono essere sufficienti per le proprie esigenze, per esempio perché individuano i clienti di maggior valore e che acquistano con maggiore frequenza; in caso contrario, i dati possono essere passati a uno stream per essere sottoposti a ulteriore modellazione e analisi.

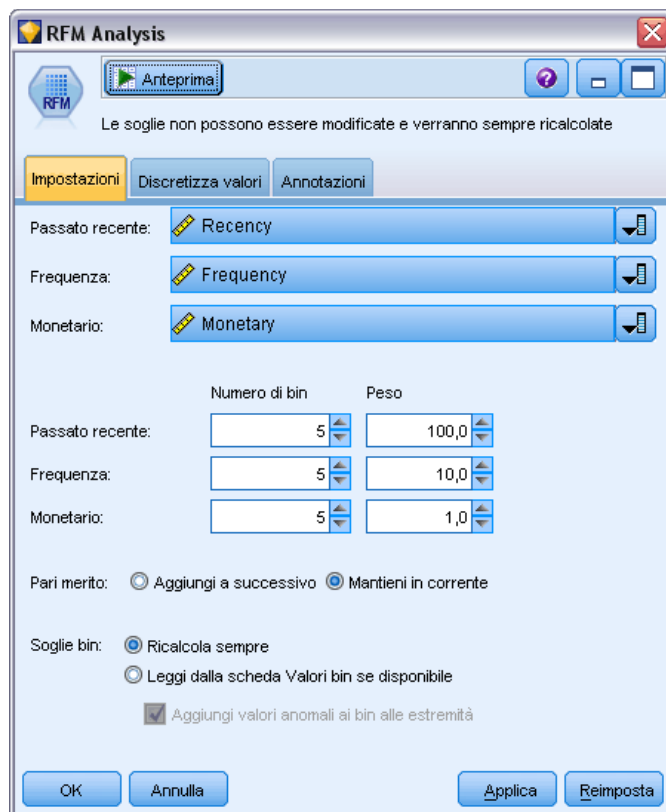
Si noti tuttavia che, sebbene la capacità di analizzare e classificare i punteggi RFM sia uno strumento utile, è necessario tenere presente determinati fattori quando lo si utilizza. Si può essere tentati di scegliere come obiettivo i clienti con le classificazioni più elevate; tuttavia, una eccessiva sollecitazione di questi clienti potrebbe creare fastidio e determinare invece un calo degli acquisti ripetuti. È inoltre bene ricordare che i clienti con punteggio basso non vanno trascurati, bensì coltivati, per farli diventare clienti migliori. Al contrario, a seconda del mercato, un punteggio alto in sé non indica necessariamente un buon cliente potenziale. Per esempio, un cliente allocato al bin 5 per passato recente (cioè un cliente che ha effettuato un acquisto molto di recente) potrebbe non essere il cliente obiettivo migliore per un'azienda che vende prodotti costosi e di maggiore durata quali auto o televisori.

Nota: a seconda del metodo di archiviazione dei dati, può essere necessario far precedere al nodo Analisi RFM un nodo Aggregazione RFM per trasformare i dati in un formato utilizzabile. Per esempio, i dati di input devono essere nel formato del cliente, con una riga per ogni cliente; se i dati dei clienti sono in formato transazionale, utilizzare un nodo Aggregazione RFM a monte per derivare i campi Passato recente, Frequenza e Monetario. [Per ulteriori informazioni, vedere l'argomento Nodo Aggregazione RFM in il capitolo 3 a pag. 75.](#)

I nodi Aggregazione RFM e Analisi RFM di IBM® SPSS® Modeler sono impostati per utilizzare la discretizzazione indipendente, ovvero classificano e discretizzano i dati rispetto a ogni misura di passato recente, frequenza e valore monetario, indipendentemente dai loro valori o dalle altre due misure.

Impostazioni del nodo Analisi RFM

Figura 4-72
Impostazione delle opzioni del nodo Analisi RFM



Passato recente. Il pulsante di selezione dei campi (a destra della casella di testo) permette di selezionare il campo Passato recente il cui valore può essere una data, un timestamp o un semplice numero. Si noti che, quando la data della transazione più recente è rappresentata da una data o da un timestamp, si considera più recente il valore più alto; quando invece è rappresentata da un numero, questo indica il tempo trascorso dalla transazione più recente e si considera più recente il valore più basso.

Nota: se nello stream il nodo Analisi RFM è preceduto da un nodo Aggregazione RFM, i campi Passato recente, Frequenza e Monetario generati dal nodo Aggregazione RFM devono essere selezionati come input nel nodo Analisi RFM.

Frequenza. Selezionare il campo Frequenza da utilizzare con il pulsante di selezione dei campi.

Monetario. Selezionare il campo Monetario da utilizzare con il pulsante di selezione dei campi.

Numero di bin. Selezionare il numero di bin da creare per ciascuno dei tre tipi di output. Il valore di default è 5.

Nota: il numero minimo di bin è 2, il massimo è 9.

Weight. Per default, nel calcolo dei punteggi la massima importanza viene attribuita ai dati relativi al passato recente, seguiti da quelli di frequenza e monetario. Se necessario è possibile modificare la ponderazione che influisce su uno o più di questi dati per cambiare il dato a cui attribuire la massima importanza.

Il punteggio RFM viene calcolato nel modo seguente: (punteggio Passato recente x peso Passato recente) + (punteggio Frequenza x peso Frequenza) + (punteggio Monetario x peso Monetario).

Pari merito. Specifica le modalità di discretizzazione dei punteggi identici (pari merito). Le opzioni disponibili sono:

- **Aggiungi a successivo.** Selezionare questa opzione per spostare i valori pari merito nel bin successivo.
- **Mantieni in corrente.** Mantiene i valori pari merito nel bin corrente (più basso). Questo metodo può determinare la creazione di un numero totale di bin inferiore. Questa è l'impostazione di default.

Soglie bin. Specificare se i punteggi RFM e le allocazioni dei bin vengono sempre rielaborati all'esecuzione del nodo, o se vengono elaborati solo quando è necessario (per esempio in occasione dell'aggiunta di nuovi dati). Se si seleziona Leggi dalla scheda Valori bin se disponibile è possibile modificare i punti di taglio superiore e inferiore dei vari bin nella scheda Discretizza valori.

Quando viene eseguito, il nodo Analisi RFM discretizza i campi grezzi di passato recente, frequenza e monetario e aggiunge i seguenti nuovi campi all'insieme di dati:

- Punteggio data ultimo acquisto. Un rango (valore di discretizzazione) per Passato recente
- Punteggio frequenza. Un rango (valore di discretizzazione) per la Frequenza
- Punteggio valore monetario. Un rango (valore di discretizzazione) per Monetario
- Punteggio RFM. La somma ponderata dei punteggi di passato recente, frequenza e monetario.

Aggiungi valori anomali ai bin alle estremità. Se si seleziona questa casella di controllo, i record di valore inferiore a quello del bin più basso vengono aggiunti a tale bin e quelli di valore superiore a quello del bin più alto vengono aggiunti a tale bin; in caso contrario viene loro assegnato un valore nullo. La casella è disponibile solo se si seleziona Leggi dalla scheda Valori bin se disponibile.

Discretizzazione con il nodo Analisi RFM

Figura 4-73

Impostazione dei valori da discretizzare per il nodo Analisi RFM



La scheda Discretizza valori consente di visualizzare e in alcuni casi modificare le soglie per i bin generati.

Nota: i valori di questa scheda si possono modificare solo se si è selezionata l'opzione Leggi dalla scheda Valori bin se disponibile nella scheda Impostazioni.

Campo nei bin. Utilizzare l'elenco a discesa per selezionare un campo da suddividere in bin. I valori disponibili sono quelli selezionati nella scheda Impostazioni.

Tabella Discretizza valori. I valori delle soglie vengono visualizzati per ogni bin generato. Se si seleziona Leggi dalla scheda Valori bin se disponibile nella scheda Impostazioni è possibile modificare i punti di taglio superiore e inferiore di ogni bin facendo doppio clic sulla cella corrispondente.

Leggi valori. Consente di leggere i valori discretizzati dall'insieme di dati e di popolare la tabella Discretizza valori. Si noti che se si seleziona Ricalcola sempre nella scheda Impostazioni, le soglie dei bin verranno sovrascritte quando vengono elaborati nuovi dati nello stream.

Nodo Partizione

I nodi Partizione vengono utilizzati per generare un campo di partizione che suddivide i dati in sottoinsiemi distinti o campioni destinati alle fasi di addestramento, test e validazione della creazione dei modelli. Utilizzando un campione per generare il modello e un altro campione per sottoporlo a verifica, è possibile ottenere una valida indicazione del modo in cui il modello potrà essere esteso a insiemi di dati di dimensioni maggiori, più simili ai dati correnti.

Il nodo Partizione genera un campo nominale con il ruolo impostato su Partizione. In alternativa, se nei dati esiste già un campo appropriato, è possibile definirlo come una partizione utilizzando un nodo Tipo. In questo caso non è necessario un nodo Partizione distinto. Come partizione è possibile utilizzare qualsiasi campo nominale istanziato con due o tre valori, ma non è possibile utilizzare i campi flag. [Per ulteriori informazioni, vedere l'argomento Impostazione del ruolo del campo a pag. 138.](#)

È possibile definire più campi di partizione in uno stream, ma in questo caso nella scheda Campi di ogni nodo Modelli che utilizza il partizionamento è necessario selezionare un singolo campo di partizione. Se è presente un'unica partizione, verrà utilizzata automaticamente quando si attiva il partizionamento.

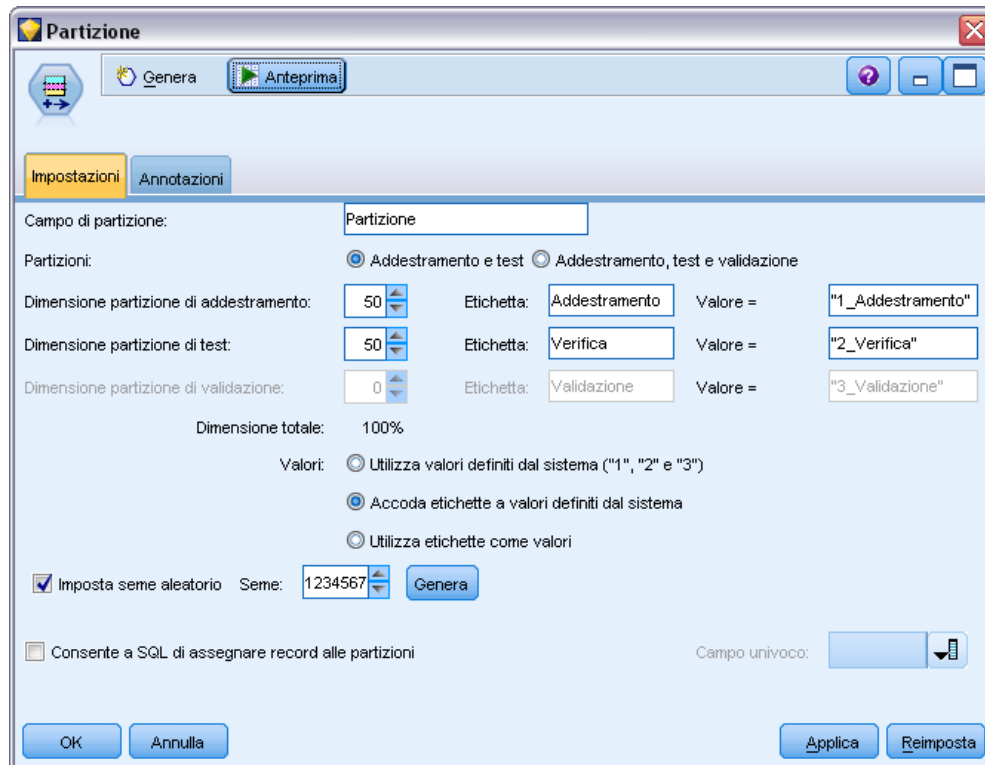
Attivazione del partizionamento. Per utilizzare la partizione in un'analisi, è necessario attivare il partizionamento nella scheda Opzioni modello del nodo di creazione del modello o di analisi appropriato. Se si deseleziona questa opzione, è possibile disattivare il partizionamento senza rimuovere il campo.

Per creare un campo di partizione in base ad altri criteri, per esempio un intervallo di date o una posizione, è inoltre possibile utilizzare un nodo Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Nuovo campo a pag. 154.](#)

Esempio. Durante la creazione di uno stream RFM per individuare i clienti recenti che hanno risposto in modo favorevole alle precedenti campagne di marketing, la divisione marketing di un'azienda utilizza un nodo Partizione per suddividere i dati in partizioni di addestramento e di test.

Opzioni del nodo Partizione

Figura 4-74
Finestra di dialogo del nodo Partizione, scheda Impostazioni



Campo di partizione. Specifica il nome del campo creato dal nodo.

Partizioni. È possibile partizionare i dati in due campioni (addestramento e test) o in tre campioni (addestramento, test e validazione).

- **Addestramento e test.** Partiziona i dati in due campioni, consentendo l'addestramento del modello con un campione e il test con un altro.
- **Addestramento, test e validazione.** Partiziona i dati in tre campioni, consentendo l'addestramento del modello con un campione, il test e il perfezionamento del modello con un secondo campione e la validazione dei risultati con un terzo. In questo modo, tuttavia, le dimensioni di ogni partizione vengono ridotte di conseguenza e ciò risulta particolarmente utile se si utilizzano insiemi di dati di dimensioni molto grandi.

Dimensione delle partizioni. Specifica la dimensione relativa di ogni partizione. Se la somma delle dimensioni di partizione è inferiore al 100%, i record non inclusi in una partizione verranno scartati. Per esempio, se un utente ha 10 milioni di record e ha specificato che il 5% delle dimensioni delle partizioni sono di addestramento e il 10% sono di test, dopo avere eseguito il nodo saranno presenti approssimativamente 500.000 record di addestramento e un milione di record di test, mentre i rimanenti saranno scartati.

Valori. Specifica i valori utilizzati per rappresentare ogni campione di partizione nei dati.

- **Utilizza valori definiti dal sistema ("1", "2" e "3").** Utilizza un intero per rappresentare ogni partizione. Per esempio, tutti i record che fanno parte del campione di addestramento hanno un valore 1 per il campo di partizione. Ciò garantisce che i dati saranno portabili con opzioni internazionali diverse e che l'ordinamento verrà mantenuto, in modo tale che 1 continuerà a rappresentare la partizione di addestramento, se il campo di partizione viene reistanziato in una posizione diversa, per esempio per la lettura di dati da un database. I valori richiedono tuttavia un'interpretazione.
- **Accoda etichette a valori definiti dal sistema.** Combina l'intero con un'etichetta, per esempio i record della partizione di addestramento hanno il valore *I_Addestramento*. In questo modo, un utente che osserva i dati è in grado di identificare i valori e l'ordinamento viene mantenuto. I valori sono tuttavia specifici delle opzioni internazionali impostate.
- **Utilizza etichette come valori.** Utilizza l'etichetta senza interi, per esempio *Addestramento*. Ciò consente di definire i valori mediante la modifica delle etichette. I dati sono tuttavia correlati alle opzioni internazionali specifiche e la reistanziatura di una colonna della partizione determinerà il loro inserimento nel relativo ordinamento naturale, che potrebbe corrispondere al relativo ordine "semantico".

Imposta seme aleatorio. Quando si esegue il campionamento o il partizionamento di record in base a una percentuale casuale, questa opzione consente di duplicare gli stessi risultati in un'altra sessione. Specificando il valore iniziale utilizzato dal generatore di numeri casuali, è possibile garantire che vengano assegnati gli stessi record a ogni esecuzione del nodo. Immettere il valore desiderato, oppure fare clic sul pulsante Genera per generare automaticamente un valore casuale. Se questa opzione non è selezionata, verrà generato un campione diverso ogni volta che si esegue il nodo.

Nota: quando si utilizza l'opzione Imposta seme aleatorio con record letti da un database, potrebbe essere necessario un nodo Ordina prima di eseguire il campionamento, per garantire lo stesso risultato ogni volta che viene eseguito il nodo. Questo si verifica perché il seme aleatorio dipende dall'ordine dei record, per il quale non si ha la garanzia che rimanga invariato in un database relazionale. [Per ulteriori informazioni, vedere l'argomento Nodo Ordina in il capitolo 3 a pag. 78.](#)

Consente a SQL di assegnare record alle partizioni. (Solo per database di Livello 1) Selezionare questa casella se si desidera utilizzare il push back SQL per assegnare record alle partizioni. Dall'elenco a discesa Campo univoco, scegliere un campo con valori univoci (come un campo ID) per fare in modo che i record vengano assegnati in modo casuale ma ripetibile.

I livelli di database sono spiegati nella descrizione del nodo di input Database. [Per ulteriori informazioni, vedere l'argomento Nodo di input Database in il capitolo 2 a pag. 13.](#)

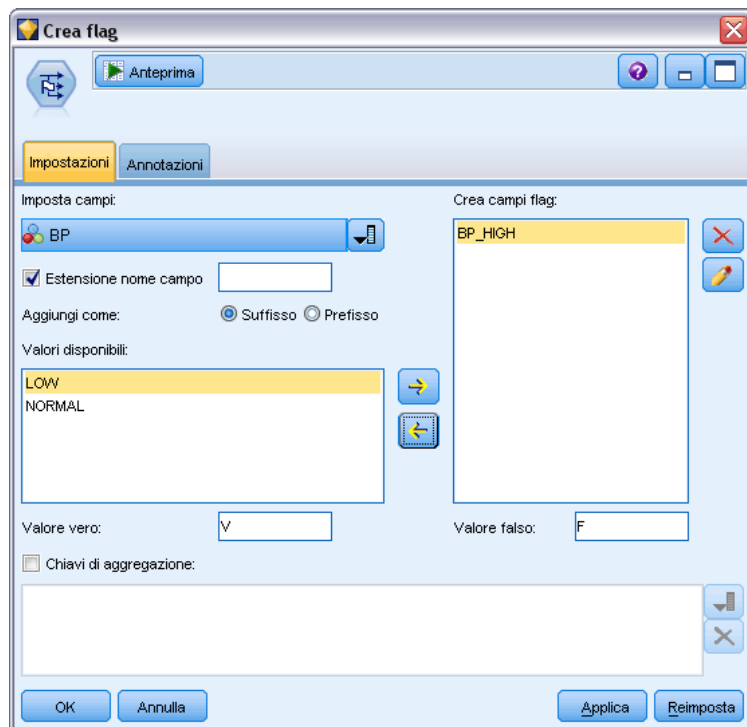
Generazione di nodi Selezione

Mediante il menu Genera del nodo Partizione è possibile generare automaticamente un nodo Selezione per ogni partizione. Per esempio, è possibile selezionare tutti i record della partizione di addestramento per ottenere ulteriori valutazioni o analisi utilizzando solo questa partizione.

Nodo Crea flag

Il nodo Crea flag consente di creare campi flag in base ai valori categoriali di uno o più campi nominali. Per esempio, l'insieme di dati può contenere un campo nominale, *PRESSIONE*, con i valori *Alto*, *Normale* e *Basso*. Per facilitare la manipolazione dei dati, è possibile creare un campo flag per la pressione alta, che indica se il paziente ha una pressione sanguigna elevata oppure no.

Figura 4-75
Creazione di un campo flag per la pressione sanguigna alta



Impostazione delle opzioni relative al nodo Crea flag

Imposta campi. Elenca tutti i campi di dati con livello di misurazione *Nominale* (insieme). Selezionarne uno dall'elenco per visualizzare i valori nell'insieme. È possibile scegliere tra questi valori per creare un campo flag. Si noti che per poter visualizzare i campi nominali disponibili (e i relativi valori) è necessario che i dati vengano completamente istanziati mediante un nodo Tipo o di input a monte. [Per ulteriori informazioni, vedere l'argomento Nodo Tipo a pag. 125.](#)

Estensione nome campo. Selezionare questa opzione per attivare i controlli per la definizione di un'estensione che verrà aggiunta al nuovo campo flag come suffisso o come prefisso. Per default, i nuovi nomi di campo vengono creati automaticamente, mediante la combinazione del nome originale del campo con il valore del campo in un'etichetta, per esempio *Nomecampo_valorecampo*.

Valori disponibili. I valori nell'insieme selezionato sopra vengono visualizzati in questa casella. Selezionare uno o più valori per i quali si desidera generare i flag. Per esempio, se i valori in un campo denominato *pressione_sanguigna* sono *Alto*, *Normale* e *Basso*, è possibile selezionare *Alto*

e aggiungere questo valore all'elenco a destra. In questo modo verrà creato un campo con un flag per i record contenenti un valore che indica un valore elevato di pressione sanguigna.

Crea campi flag. I campi flag creati vengono elencati in questa casella. È possibile specificare opzioni per l'assegnazione di un nome al nuovo campo utilizzando i controlli dell'estensione dei nomi di campo.

Valore vero. Specificare il valore vero utilizzato dal nodo durante l'impostazione di un flag. Per default, questo valore è T.

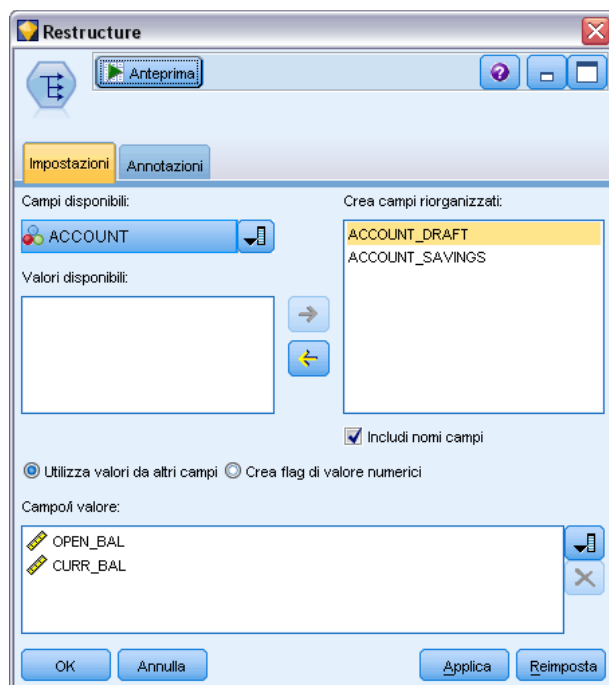
Valore falso. Specificare il valore falso utilizzato dal nodo durante l'impostazione di un flag. Per default, questo valore è F.

Chiavi di aggregazione. Selezionare questa opzione per raggruppare i record in base ai campi chiave specificati sotto. Se si seleziona Chiavi di aggregazione, tutti i campi flag in un gruppo verranno "attivati" se *qualsiasi* record è stato impostato su vero. Utilizzare il pulsante di selezione dei campi per specificare i campi chiave che verranno utilizzati per l'aggregazione dei record.

Nodo Riorganizza

Il nodo Riorganizza può essere utilizzato per generare più campi in base ai valori di un campo nominale o flag. I campi così generati possono contenere valori di altri campi o flag numerici (0 e 1). Il funzionamento di questo nodo è simile a quello del nodo Crea flag, ma offre maggiore flessibilità. Esso consente, infatti, di creare campi di qualunque tipo (anche flag numerici) utilizzando i valori di un altro campo. In seguito è possibile eseguire l'aggregazione o altri tipi di manipolazione con altri nodi situati a valle. Il nodo Crea flag consente di aggregare i campi con un'unica operazione, il che può risultare comodo in caso di creazione di campi flag.

Figura 4-76
Generazione di campi riorganizzati per Conto



Per esempio, l'insieme di dati che segue contiene un campo nominale, *Conto*, con i valori *Risparmio* e *Bozza*. Per ogni conto vengono registrati il saldo d'apertura e il saldo attuale e ad alcuni clienti sono intestati più conti di ciascun tipo. Si supponga di voler sapere se ogni cliente ha un determinato tipo di conto e, in caso affermativo, quanto denaro è presente in ogni tipo di conto. Con il nodo Riorganizza, viene generato un campo per ogni singolo valore del *Conto* e si seleziona *Saldo_attuale* come valore. Ogni nuovo campo conterrà il saldo attuale relativo al record corrispondente.

Tabella 4-2
Dati di esempio prima della riorganizzazione

IDclie	Conto	Saldo_apert	Saldo_att
12701	Bozza	1000	1005.32
12702	Risparmio	100	144.51
12703	Risparmio	300	321.20
12703	Risparmio	150	204.51
12703	Bozza	1200	586.32

Tabella 4-3
Dati di esempio dopo la riorganizzazione

IDclie	Conto	Saldo_apert	Saldo_att	Conto_Bozza_Saldo_att	Conto_risparmio_Saldo_att
12701	Bozza	1000	1005.32	1005.32	\$null\$
12702	Risparmio	100	144.51	\$null\$	144.51
12703	Risparmio	300	321.20	\$null\$	321.20

IDclie	Conto	Saldo_apert	Saldo_att	Conto_Bozza_Saldo_att	Conto_risparmio_Saldo_att
12703	Risparmio	150	204.51	\$null\$	204.51
12703	Bozza	1200	586.32	586.32	\$null\$

Utilizzo del nodo Riorganizza con il nodo Aggregazione

In molti casi può essere consigliabile abbinare il nodo Riorganizza a un nodo Aggregazione. Nell'esempio precedente, un cliente (con il codice identificativo 12703) è titolare di tre conti. È possibile utilizzare un nodo Aggregazione per calcolare il saldo totale di ogni tipo di conto. Il campo chiave è *IDclie* e i campi aggregati sono i nuovi campi riorganizzati, *Conto_Bozza_Saldo_att* e *Conto_Risparmio_Saldo_att*. I risultati sono illustrati nella tabella seguente.

Tabella 4-4

Dati di esempio dopo la riorganizzazione e l'aggregazione

IDclie	Conteggio_reco	Conto_Bozza_Somma_Saldo_att	Conto_Risparmio_Somma_Saldo_att
12701	1	1005.32	\$null\$
12702	1	\$null\$	144.51
12703	3	586.32	525.71

Impostazione delle opzioni relative al nodo Riorganizza

Campi disponibili. Elenca tutti i campi di dati con livello di misurazione *Nominale* (insieme) o *Flag*. Selezionare un campo per visualizzare i valori dell'insieme o del flag, quindi scegliere i valori per la creazione dei campi riorganizzati. Si noti che per poter visualizzare i campi disponibili (e i relativi valori), è necessario che i dati vengano completamente istanziati mediante un nodo Tipo o di input a monte. [Per ulteriori informazioni, vedere l'argomento Nodo Tipo a pag. 125.](#)

Valori disponibili. I valori nell'insieme selezionato sopra vengono visualizzati in questa casella. Selezionare uno o più valori per i quali si desidera generare i campi riorganizzati. Per esempio, se i valori in un campo denominato *Pressione sanguigna* sono *Alto*, *Normale* e *Basso*, è possibile selezionare *Alto* e aggiungere questo valore all'elenco a destra. Questo determinerà la creazione di un campo con il valore specificato (vedere sotto) per i record con valore *Alto*.

Crea campi riorganizzati. I campi riorganizzati creati vengono elencati in questa casella. Per default, i nuovi nomi di campo vengono creati automaticamente, mediante la combinazione del nome originale del campo con il valore del campo in un'etichetta, per esempio *Nomecampo_valorecampo*.

Includi nomi campi. Deselezionare questa opzione per rimuovere il nome del campo originale apposto come prefisso dai nomi dei nuovi campi.

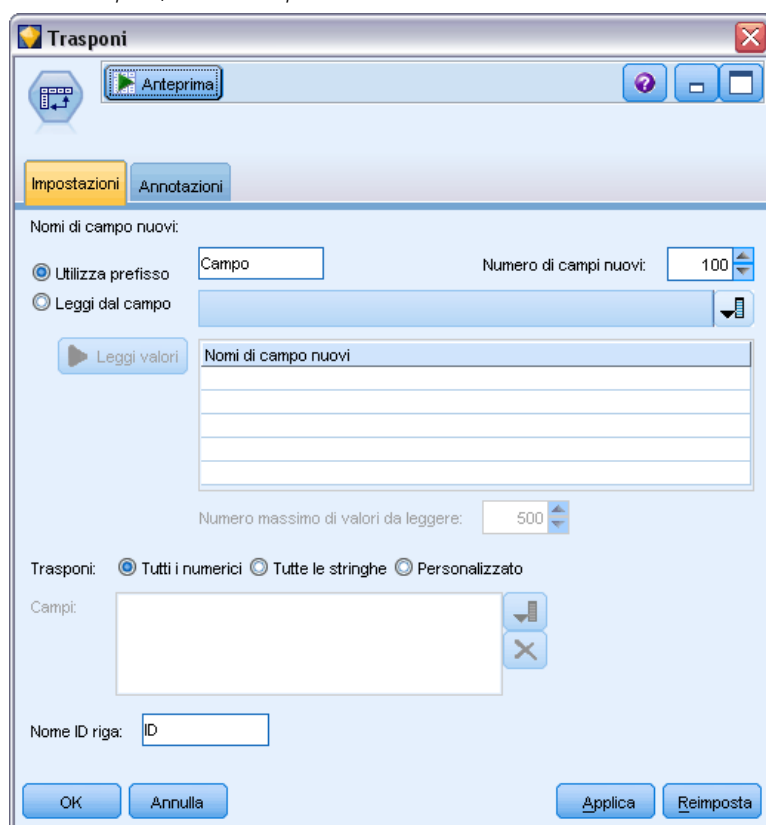
Utilizza valori da altri campi. Consente di specificare uno o più campi i cui valori saranno utilizzati per popolare i campi riorganizzati. Utilizzare il pulsante Selezione campi per selezionare uno o più campi. Per ogni campo scelto viene creato un nuovo campo. Il nome campo del valore viene accodato al nome del campo riorganizzato: per esempio, *Pressione_Alto_Età* o *Pressione_Basso_Età*. Ogni nuovo campo eredita il tipo di campo del valore originale.

Crea flag di valore numerici. Selezionare questa opzione per popolare i nuovi campi con flag di valore numerici (0 = falso e 1 = vero) anziché utilizzare un valore preso da un altro campo.

Nodo Trasponi

Per default, le colonne sono campi e le righe sono record o osservazioni. Se necessario, è possibile utilizzare un nodo Trasponi per scambiare i dati delle righe e delle colonne in modo da trasformare i campi in record e i record in campi. Per esempio, se nei dati di serie storica di cui si dispone ogni serie è una riga anziché una colonna, è possibile trasporre i dati prima di analizzarli.

Figura 4-77
Nodo Trasponi, scheda Impostazioni



Impostazione delle opzioni del nodo Trasponi

Nomi di campo nuovi

È possibile generare automaticamente i nomi dei nuovi campi in base a un prefisso specificato, oppure leggerli da un campo esistente nei dati.

Utilizza prefisso. Questa opzione genera automaticamente i nomi dei nuovi campi in base al prefisso specificato (*Campo1*, *Campo2* e così via). Il prefisso può essere personalizzato in funzione delle proprie esigenze. Con questa opzione è necessario indicare il numero di campi da

creare, indipendentemente dal numero di righe dei dati originali. Per esempio, se Numero di campi nuovi è impostato su 100, tutti i dati oltre le prime 100 righe saranno scartati. Se i dati originali contengono meno di 100 righe, alcuni campi conterranno valori nulli. Il numero dei campi può essere aumentato a piacere, ma lo scopo di questa impostazione è evitare la trasposizione di un milione di record in un milione di campi, che genererebbe un risultato impossibile da gestire.

Si supponga per esempio di disporre di dati con serie suddivise per righe e di un campo separato (colonna) per ogni mese. Questi dati possono essere trasposti in modo che ogni serie si trovi in un campo separato, con una riga per ogni mese.

Figura 4-78
Dati originali con serie in righe

	Jan	Feb	Mar	Apr
1	1	3	5	7
2	2	4	6	8

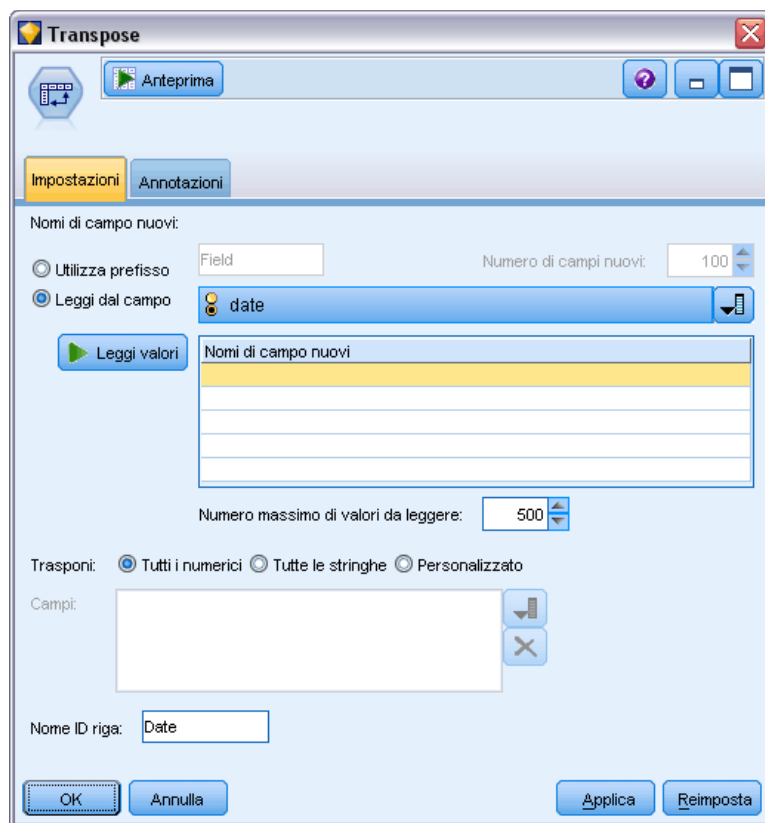
Figura 4-79
Dati trasposti con serie in colonne

	Month	Field1	Field2
1	Jan	1	2
2	Feb	3	4
3	Mar	5	6
4	Apr	7	8

Nota: per ottenere i risultati mostrati, l'opzione Numero di campi nuovi è stata cambiata da 100 a 2 e il nome ID della riga è stato modificato da ID a Mese (vedi sotto).

Leggi dal campo. Legge i nomi dei campi da un campo esistente. Con questa opzione, il numero dei nuovi campi è determinato dai dati, fino al limite massimo specificato. Ogni valore del campo selezionato diventa un nuovo campo nei dati di output. Il campo selezionato può avere qualsiasi tipo di archiviazione (numero intero, stringa, data e così via) ma per evitare che due campi abbiano lo stesso nome, ogni valore del campo selezionato deve essere univoco (in altre parole, il numero dei valori deve corrispondere al numero delle righe). Se vengono rilevati campi con nomi uguali, viene visualizzato un avviso.

Figura 4-80
Lettura dei nomi dei campi da un campo esistente



- **Leggi valori.** Se il campo selezionato non è stato istanziato, selezionare questa opzione per creare l'elenco dei nuovi nomi di campi. Se il campo è già stato istanziato, questa operazione non è necessaria.
- **Numero massimo di valori da leggere.** Quando i nomi dei campi vengono letti dai dati viene indicato un limite superiore per evitare di creare un numero eccessivo di campi (come notato in precedenza, la trasposizione di un milione di record in un milione di campi genererebbe un risultato impossibile da gestire).

Per esempio, se la prima colonna nei dati specifica il nome di ciascuna serie, è possibile usare questi valori come nomi di campi nei dati trasposti.

Figura 4-81
Dati originali con serie in un'unica riga

	date	1949-01-01	1949-02-01	1949-04-01	1949-05-01	1949-06-01	1949-07-01	1949-08-01
1	passengers	112.000	118.000	129.000	121.000	135.000	148.000	148.000

Figura 4-82
Dati trasposti con serie in colonne

	Date	passengers
1	1949-01-01	112.000
2	1949-02-01	118.000
3	1949-04-01	129.000
4	1949-05-01	121.000
5	1949-06-01	135.000
6	1949-07-01	148.000
7	1949-08-01	148.000
8	1949-09-01	136.000
9	1949-10-01	119.000
10	1949-11-01	104.000
11	1949-12-01	118.000
12	1950-01-01	115.000
13	1950-02-01	126.000
14	1950-03-01	141.000
15	1950-04-01	135.000
16	1950-05-01	125.000
17	1950-06-01	149.000
18	1950-07-01	170.000
19	1950-08-01	170.000
20	1950-09-01	158.000

Trasponi. Per default, vengono trasposti solo i campi continui (intervalli numerici) (archiviazione di numeri interi o reali). Se lo si desidera, è possibile scegliere un sottoinsieme di campi numerici o trasporre invece solo campi stringa. Tuttavia, i campi trasposti devono avere tutti lo stesso tipo di archiviazione (numerica o stringa, ma non entrambi), poiché se si mescolano i campi di input vengono generati valori misti all'interno di ogni colonna di output, in violazione della regola secondo cui tutti i valori di un campo devono avere lo stesso tipo di archiviazione. Gli altri tipi di archiviazione (data, ora, timestamp) non possono essere trasposti.

- **Tutti i numerici.** Traspone tutti i campi numerici (archiviazione di numeri interi o reali). Il numero delle righe di output corrisponde al numero di campi numerici presente nei dati originali.
- **Tutte le stringhe.** Traspone tutti i campi stringa.
- **Personalizzato.** Consente di selezionare un sottoinsieme di campi numerici. Il numero delle righe di output corrisponde al numero dei campi selezionati. *Nota:* Questa opzione è disponibile solo per i campi numerici.

Nome ID riga. Specifica il nome del campo ID riga creato dal nodo. I valori di questo campo sono determinati dai nomi dei campi dei dati originali.

Suggerimento: Quando si esegue la trasposizione dei dati di serie storica da righe a colonne, se i dati originali includono una riga, per esempio data, mese o anno, che etichetta il periodo relativo a ogni misurazione, assicurarsi di leggere queste etichette in IBM® SPSS® Modeler come nomi di campi (come dimostrato negli esempi precedenti, che mostrano il mese o la data come nomi di campi nei dati originali) invece di includere l'etichetta nella prima riga di dati. In questo modo si eviterà di mescolare etichette e valori nelle singole colonne (in quel caso, i numeri verrebbero letti come stringhe dato che non ci possono essere più tipi di archiviazione all'interno di una stessa colonna).

Nodo Intervalli di tempo

Il nodo Intervalli di tempo consente di specificare gli intervalli e generare le etichette per i dati di serie storica da utilizzare in una creazione di modelli di serie storica o in un nodo Plot tempo per realizzare stime o previsioni. È supportata una gamma completa di intervalli temporali, dai secondi agli anni. Per esempio, se si dispone di una serie di misurazioni giornaliere che iniziano il 3 gennaio 2005, è possibile assegnare delle etichette ai record a partire da quella data; la seconda riga sarà quella del 4 gennaio, e così via. È inoltre possibile specificare la periodicità: per esempio, cinque giorni alla settimana oppure otto ore al giorno.

Inoltre, è possibile specificare l'intervallo dei record da utilizzare per le stime. Si può scegliere di escludere i primi record della serie e di specificare i record di controllo. In questo modo sarà possibile testare il modello utilizzando come controllo i record più recenti della serie storica per confrontare i loro valori noti con i valori stimati per quei periodi.

È anche possibile specificare quanti periodi di tempo futuri si desidera prevedere e i valori futuri da utilizzare nelle previsioni da parte dei nodi Modelli di serie storica a valle.

Il nodo Intervalli di tempo genera un campo *EtichettaTempo* in un formato appropriato per l'intervallo e il periodo indicati, insieme a un campo *IndiceTempo* che assegna un numero intero univoco a ogni record. È possibile che vengano generati anche diversi campi aggiuntivi, a seconda del periodo o dell'intervallo selezionato (per esempio, il minuto o il secondo in cui avviene una misurazione).

È possibile integrare o aggregare i valori in funzione delle proprie esigenze per fare in modo che le misurazioni siano ripartite in modo uniforme. I metodi di modellazione dei dati di serie storica richiedono un intervallo uniforme tra una misurazione e l'altra, con gli eventuali valori mancanti indicati da righe vuote. Se i dati a disposizione dell'utente non possiedono già queste caratteristiche, il nodo è in grado di trasformarli nel formato richiesto.

Commenti

- È possibile che gli intervalli periodici non corrispondano al tempo reale. Per esempio, una serie basata su una settimana lavorativa standard di cinque giorni considererebbe l'interruzione fra venerdì e lunedì come un solo giorno.
- Il nodo Intervalli di tempo presuppone che ogni serie si trovi in un campo o in una colonna, con una riga per ogni misurazione. Se necessario, è possibile trasporre i dati per soddisfare questo requisito. [Per ulteriori informazioni, vedere l'argomento Nodo Trasponi a pag. 200.](#)
- Per le serie non ripartite in modo uniforme, è possibile specificare un campo che identifica la data o l'ora di ogni misurazione. Si noti che sono richiesti un campo data, ora o timestamp nel formato appropriato da utilizzare come input. Se necessario, è possibile convertire un campo esistente (come un campo etichetta stringa) in questo formato utilizzando un nodo Riempimento. [Per ulteriori informazioni, vedere l'argomento Conversione della classe di archiviazione mediante il nodo Riempimento a pag. 168.](#)
- Quando vengono visualizzati i dettagli dei campi delle etichette e dell'indice che sono stati generati, in molti casi può essere utile attivare la visualizzazione delle etichette dei valori. Per esempio, quando si esamina una tabella contenente dei valori generati per dati mensili, è possibile fare clic sull'icona delle etichette dei valori nella barra degli strumenti per visualizzare *gennaio, febbraio, marzo* e così via anziché *1, 2, 3* e così via.

Figura 4-83

Icona etichette valori



Indicazione degli intervalli di tempo

La scheda Intervalli consente di indicare l'intervallo e la periodicità per la creazione della serie o l'assegnazione di etichette alla stessa. Le impostazioni specifiche dipendono dall'intervallo selezionato. Per esempio, se si sceglie Ore al giorno, è possibile indicare il numero di giorni alla settimana, il giorno d'inizio di ogni settimana, il numero di ore al giorno e l'ora d'inizio di ogni giorno. [Per ulteriori informazioni, vedere l'argomento Intervalli supportati a pag. 213.](#)

Figura 4-84
Impostazioni degli intervalli di tempo per una serie oraria

Assegnazione di etichette o creazione della serie

È possibile identificare i record consecutivamente o creare le serie in base a un campo data, timestamp o ora specifico.

- Iniziare assegnando un'etichetta al primo record.** Indicare la data e/o l'ora d'inizio per assegnare l'etichetta ai record consecutivi. Per assegnare l'etichetta Ore al giorno, per esempio, è necessario specificare la data e l'ora d'inizio della serie, che deve proseguire con un solo record per ogni ora successiva. A parte l'aggiunta delle etichette, questo metodo non modifica i dati originali. Esso presuppone invece che i record siano già ripartiti in modo equivalente, con un intervallo uniforme tra tutte le misurazioni. Le eventuali misurazioni mancanti devono essere indicate nei dati da righe vuote.
- Crea dai dati.** Per le serie non ripartite uniformemente, è possibile specificare un campo che identifica la data o l'ora di ogni misurazione. Si noti che sono richiesti un campo data, ora o timestamp nel formato appropriato da utilizzare come input. Per esempio, se si dispone di un campo stringa con valori come *Gen 2000*, *Feb 2000* e così via, è possibile convertirlo in campo data mediante un nodo Riempimento. [Per ulteriori informazioni, vedere l'argomento Conversione della classe di archiviazione mediante il nodo Riempimento a pag. 168.](#) L'opzione Crea dai dati, inoltre, trasforma i dati per adeguarli all'intervallo specificato integrando o aggregando opportunamente i record, per esempio “facendo confluire” le settimane in mesi, o sostituendo i record mancanti con spazi vuoti o valori estrapolati. Le funzioni da utilizzare per integrare o aggregare i record possono essere impostate nella scheda

Creazione. [Per ulteriori informazioni, vedere l'argomento Opzioni di creazione degli intervalli di tempo a pag. 207.](#)

Nuova estensione nome di campo. Consente di specificare un prefisso o un suffisso da applicare a tutti i campi generati dal nodo. Per esempio, se si utilizza il prefisso di default *\$TI_*, i campi creati dal nodo sarebbero denominati *\$TI_IndiceTempo*, *\$TI_EtichettaTempo* e così via.

Formato data. Indica il formato del campo *EtichettaTempo* creato dal nodo applicabile all'intervallo corrente. La disponibilità di questa opzione dipende dall'elemento selezionato.

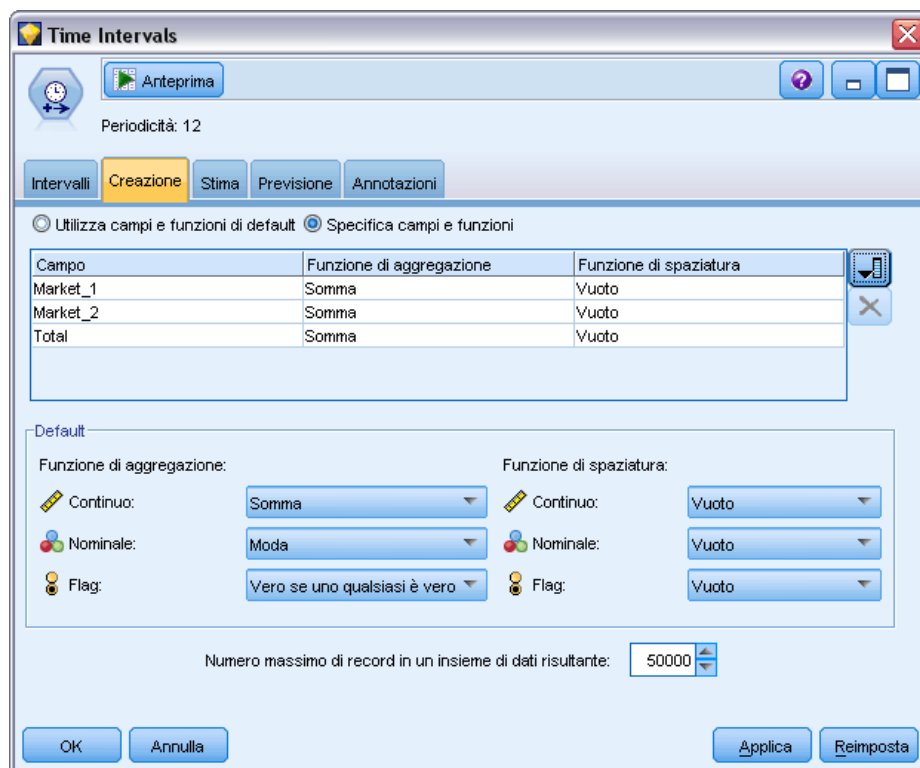
Formato ora. Indica il formato del campo *EtichettaTempo* creato dal nodo applicabile all'intervallo corrente. La disponibilità di questa opzione dipende dall'elemento selezionato.

Opzioni di creazione degli intervalli di tempo

La scheda Creazione del nodo Intervalli di tempo consente di selezionare le opzioni per l'aggregazione o l'integrazione dei campi al fine di farli corrispondere all'intervallo specificato. Queste impostazioni sono valide solo quando nella scheda Intervalli è selezionata l'opzione Crea dai dati. Per esempio, se si dispone di un insieme eterogeneo di dati settimanali e mensili, è possibile aggregare o "accorpare" i valori settimanali per ottenere un intervallo mensile uniforme. In alternativa è possibile impostare un intervallo settimanale e integrare la serie inserendo valori vuoti per tutte le settimane mancanti o estrapolando i valori mancanti mediante una funzione di integrazione specifica.

In caso di integrazione o aggregazione dei dati, tutti i campi data o timestamp esistenti vengono sostituiti a tutti gli effetti dai campi *EtichettaTempo* e *IndiceTempo* generati e vengono esclusi dall'output. Anche i campi senza tipo vengono eliminati. I campi che misurano il tempo come durata vengono conservati (per esempio un campo che misura la lunghezza di un intervento di assistenza anziché l'ora di inizio dell'intervento) a condizione che vengano archiviati internamente come campi ora e non come campi timestamp. [Per ulteriori informazioni, vedere l'argomento Impostazione dell'archiviazione e della formattazione dei campi in il capitolo 2 a pag. 27.](#) Gli altri campi vengono aggregati a seconda delle opzioni specificate nella scheda Creazione.

Figura 4-85
Nodo Intervalli di tempo, scheda Creazione



- Utilizza campi e funzioni di default.** Indica che tutti i campi devono essere opportunamente aggregati o integrati, fatta eccezione per i campi data, timestamp e per quelli senza tipo come notato in precedenza. La funzione di default viene applicata in base al livello di misurazione: per esempio, per l'aggregazione dei campi continui viene utilizzata la media, mentre per i campi nominali viene impiegata la moda. L'impostazione di default di uno o più livelli di misurazione può essere modificata nella parte inferiore della finestra di dialogo.
- Specifica campi e funzioni.** Consente di specificare i campi da integrare o da aggregare e la funzione da utilizzare per ciascuna operazione. I campi non selezionati vengono eliminati dall'output. Aggiungere o eliminare i campi dalla tabella mediante le icone a destra, oppure fare clic nella cella all'interno della colonna desiderata per modificare la funzione di aggregazione o di integrazione utilizzata per un determinato campo in modo da ignorare l'impostazione di default. I campi senza tipo sono esclusi dall'elenco e non possono essere aggiunti alla tabella.

Default. Indica le funzioni di aggregazione e di integrazione utilizzate per default per i vari tipi di campi. Queste impostazioni di default vengono applicate quando è selezionato Utilizza impostazioni di default e anche come impostazioni di default iniziali per tutti i nuovi campi aggiunti alla tabella. La modifica delle impostazioni di default non modifica le impostazioni esistenti della tabella ma è valida per tutti i campi aggiunti in seguito.

Funzione di aggregazione. Sono disponibili le seguenti funzioni di aggregazione:

- Continuo.** Le funzioni disponibili per i campi continui sono Media, Somma, Moda, Min e Max.

- **Nominale.** Le opzioni disponibili sono Moda, Primi e Ultimo. Primi indica il primo valore non nullo (ordinato per data) del gruppo di aggregazione; Ultimo indica l'ultimo valore non nullo del gruppo.
- **Flag.** Le opzioni disponibili sono Vero se uno qualsiasi è vero, Moda, Primi e Ultimo.

Funzione di spaziatura. Sono disponibili le seguenti funzioni di integrazione:

- **Continuo.** Le opzioni disponibili sono Vuoto e Media dei punti recenti, ovvero la media dei tre valori non nulli più recenti precedenti al periodo di tempo che sarà creato. Se non sono disponibili tre valori, il nuovo valore è vuoto. I valori recenti comprendono solo i valori effettivi; i valori integrati creati in precedenza non vengono considerati nella ricerca di un valore non nullo.
- **Nominale.** Vuoto e Valore più recente. "Più recente" si riferisce al valore non nullo più recente precedente al periodo di tempo che sarà creato. Anche in questo caso, nella ricerca di un valore recente vengono considerati solo i valori effettivi.
- **Flag.** Le opzioni disponibili sono Vuoto, Vero e Falso.

Numero massimo di record in un insieme di dati risultante. Consente di impostare un limite superiore al numero di record creati che potrebbe altrimenti aumentare a dismisura, soprattutto quando l'intervallo di tempo è impostato (deliberatamente o meno) sui secondi. Per esempio, una serie comprendente solo due valori (1 gennaio 2000 e 1 gennaio 2001) genererebbe 31.536.000 record se venisse integrata in base ai secondi (60 secondi x 60 minuti x 24 ore x 365 giorni). Quando si supera il limite massimo impostato, il sistema interrompe l'elaborazione e visualizza un avviso.

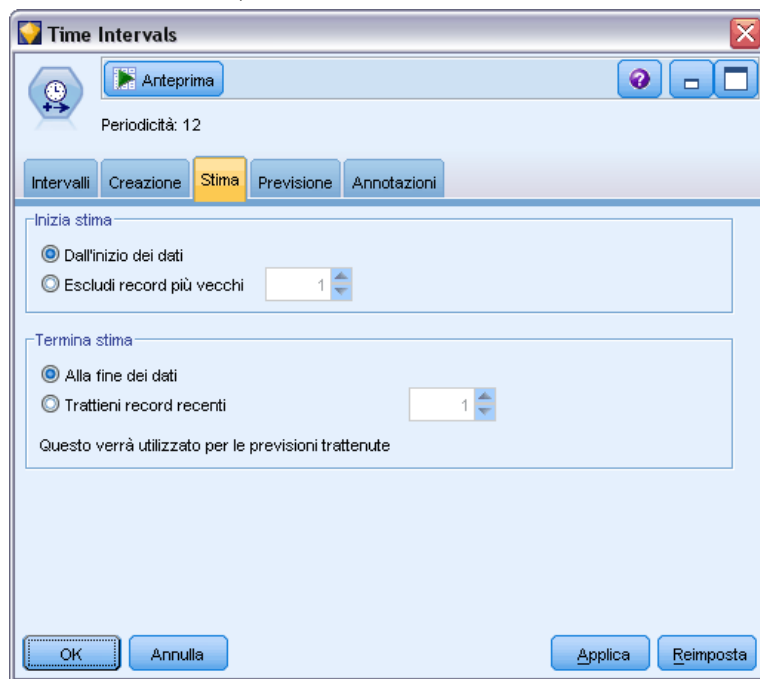
Campo Conteggio

Quando i valori vengono aggregati o integrati, il sistema crea un nuovo campo denominato *Conteggio* che indica il numero di record utilizzati per determinare il nuovo record. Per esempio, se in un solo mese vengono aggregati quattro valori settimanali il conteggio sarebbe pari a 4. Per un record integrato, il conteggio è 0. Il nome del campo è *Conteggio* più il prefisso o il suffisso specificato nella scheda Intervallo.

Periodo di stima

La scheda Stima del nodo Intervalli di tempo consente di specificare la gamma di record utilizzati nella stima del modello, oltre agli eventuali record di controllo. Queste impostazioni possono essere sovrascritte nei nodi Modelli a valle, ma la specifica delle stesse a questo punto potrebbe essere più comoda che la specifica a livello dei singoli nodi.

Figura 4-86
Nodo Intervalli di tempo, scheda Stima



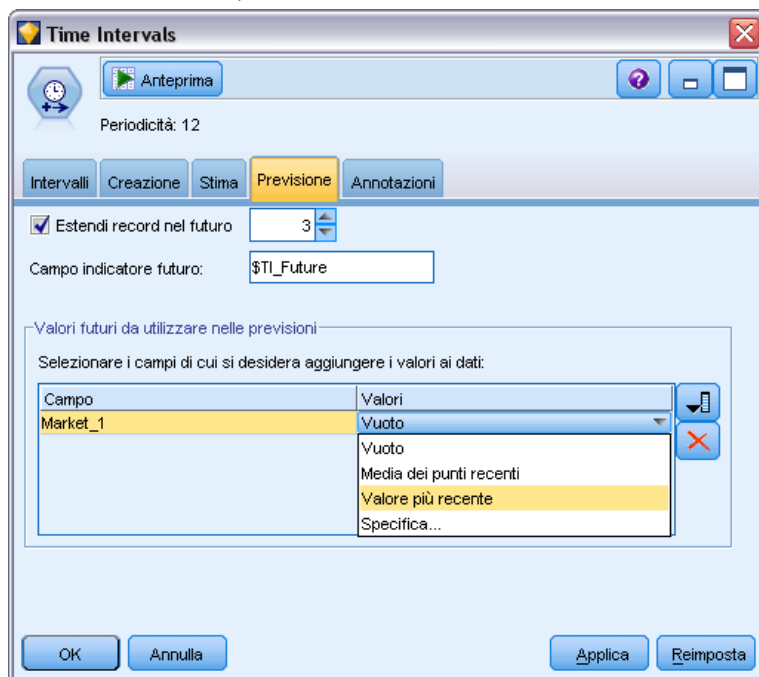
Inizia stima. È possibile far partire il periodo di stima all'inizio dei dati o escludere i valori meno recenti che potrebbero risultare poco utili ai fini della previsione. A seconda dei dati, la riduzione del periodo di stima potrebbe migliorare la performance e ridurre il tempo necessario alla preparazione dei dati, senza comportare una significativa diminuzione della precisione della previsione.

Termina stima. È possibile eseguire una stima del modello utilizzando tutti i record fino alla fine dei dati oppure utilizzare i record più recenti come record di controllo per valutare il modello. Nel secondo caso, si sta eseguendo la "previsione" di valori già noti, il che consente di confrontare i valori osservati e predittivi per misurare l'efficacia del modello.

Previsioni

La scheda Previsione del nodo Intervalli di tempo consente di specificare il numero dei record che si desidera prevedere e specificare i valori futuri da utilizzare nelle previsioni con i nodi Modelli di serie storica. Queste impostazioni possono essere sovrascritte nei nodi Modelli a valle, ma la specifica delle stesse a questo punto potrebbe essere più comoda che la specifica a livello dei singoli nodi.

Figura 4-87
Nodo Intervalli di tempo, scheda Previsione



Estendi record nel futuro. Specifica il numero dei record da prevedere oltre il periodo di stima. Si noti che questi record possono essere “previsti” o meno, a seconda del numero dei record di controllo specificati nella scheda Stima.

Campo indicatore futuro. Etichetta del campo generato che indica se un record contiene i dati di previsione o meno. Il valore di default per l’etichetta è *\$TI_Futuro*.

Valori futuri da utilizzare nelle previsioni. Per ciascun record che si desidera prevedere (ad eccezione dei record di controllo), se si utilizzano campi predittori (con il ruolo impostato su *Input*), è necessario specificare i valori stimati per il periodo di previsione per ogni predittore. Questi valori possono essere specificati manualmente oppure scelti da un elenco.

- **Campo.** Fare clic sul pulsante di selezione dei campi e scegliere gli eventuali campi che potrebbero essere utilizzati come predittori. Si noti che i campi selezionati qui possono essere utilizzati nella modellazione oppure no; per utilizzare effettivamente un campo come predittore, è necessario selezionarlo in un nodo Modelli a valle. Questa finestra di dialogo rappresenta semplicemente un metodo conveniente per specificare i valori futuri in modo che possano essere condivisi da più nodi Modelli a valle senza che sia necessario specificarli separatamente nei singoli nodi. Si noti anche che l’elenco dei campi disponibili può essere limitato a seconda delle selezioni effettuate nella scheda Creazione. Per esempio, se nella scheda Creazione è stata selezionata l’opzione Specifica campi e funzioni, i campi non aggregati o integrati vengono esclusi dallo stream e non possono essere utilizzati nella modellazione.

Nota: se vengono specificati valori futuri per un campo che non è più disponibile nello stream (perché è stato escluso oppure a causa di nuove selezioni eseguite nella scheda Creazione), tale campo viene visualizzato in rosso nella scheda Previsione.

- **Valori.** Per ciascun campo, è possibile scegliere da un elenco di funzioni o fare clic su *Specifica* per immettere i valori manualmente o selezionare i valori predefiniti dall'elenco. Se i campi predittori sono correlati a elementi sotto il controllo dell'utente o comunque conoscibili in anticipo, immettere i valori manualmente. Per esempio, se si sta eseguendo la previsione delle entrate di un hotel per il mese prossimo sulla base del numero delle camere prenotate, è possibile specificare il numero delle prenotazioni effettive di cui già si dispone per quel periodo. Al contrario, se un campo predittore si riferisce a un elemento al di fuori del controllo dell'utente, per esempio al prezzo di un titolo azionario, è possibile utilizzare una funzione, quale il valore più recente oppure la media dei punti recenti.

Le funzioni disponibili dipendono dal livello di misurazione del campo.

Livello di misurazione	Funzioni
Campo continuo o nominale	Blank Media dei punti recenti Valore più recente Specify
Campo flag	Blank Valore più recente True False Specify

Media dei punti recenti— Calcola il valore futuro in base alla media degli ultimi tre punti dei dati.

Valore più recente— Imposta il valore futuro sul valore del punto dei dati più recente.

Vero/Falso— Imposta il valore futuro di un campo flag su Vero o Falso, come specificato.

Specifica— Apre una finestra di dialogo per specificare i valori futuri manualmente oppure per consentire di selezionarli da un elenco predefinito.

Figura 4-88
Specifica dei valori futuri per i predittori



Valori futuri

Qui è possibile specificare i valori futuri da utilizzare nelle previsioni con i nodi Modelli di serie storica a valle. Queste impostazioni possono essere sovrascritte nei nodi Modelli a valle, ma la specifica delle stesse a questo punto potrebbe essere più comoda che la specifica a livello dei singoli nodi.

È possibile immettere i valori manualmente, oppure fare clic sul pulsante di selezione a destra della finestra di dialogo per scegliere da un elenco di valori definiti per il campo corrente. [Per ulteriori informazioni, vedere l'argomento Visualizzazione o selezione di valori in il capitolo 7 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Il numero dei valori futuri che possono essere specificati corrisponde al numero dei record in base ai quali si sta estendendo la serie storica nel futuro.

Intervalli supportati

Il nodo Intervalli di tempo supporta una gamma di intervalli completa, dai secondi agli anni, nonché periodi ciclici (per esempio, le stagioni) e non ciclici. L'intervallo viene specificato nel campo Intervallo di tempo nella scheda Intervalli.

Periods

Selezionare l'opzione Periodi per assegnare un'etichetta a una serie esistente non ciclica che non corrisponde ad alcuno degli altri intervalli specificati. La serie deve già essere nell'ordine esatto, con un intervallo uniforme tra una misurazione e l'altra. Quando è selezionato questo intervallo, l'opzione Crea dai dati non è disponibile.

Figura 4-89
Impostazioni degli intervalli di tempo per i periodi non ciclici

Output campione

L'assegnazione delle etichette ai record avviene in modo incrementale, in base al valore iniziale specifico (*Periodo 1*, *Periodo 2* e così via). I nuovi campi vengono creati nel modo seguente:

\$TI_IndiceTempo (Intero)	\$TI_EtichettaTempo (Stringa)	\$TI_Periodo (Intero)
1	Periodo 1	1
2	Periodo 2	2
3	Periodo 3	3

\$TI_IndiceTempo (Intero)	\$TI_EtiquetaTempo (Stringa)	\$TI_Periodo (Intero)
4	Periodo 4	4
5	Periodo 5	5

Periodi ciclici

Selezionare Periodi ciclici per assegnare un'etichetta a una serie esistente caratterizzata da un ciclo ripetitivo che non corrisponde ad alcuno degli intervalli standard. Questa opzione si potrebbe utilizzare, per esempio, se un anno fiscale comprende solo 10 mesi. La serie deve già essere nell'ordine esatto, con un intervallo uniforme tra una misurazione e l'altra. Quando è selezionato questo intervallo, l'opzione Crea dai dati non è disponibile.

Figura 4-90

Impostazioni degli intervalli di tempo per i periodi ciclici

Intervallo di tempo: **Periodi ciclici**

Numero di periodi per ciclo: **12**

Avvia l'assegnazione di etichette dal primo record Crea dai dati

Ciclo: **1**

Periodo: **1**

Nuova estensione nome di campo: **\$TI_** Aggiungi come: Prefisso Suffixo

Output campione

L'assegnazione delle etichette ai record avviene in modo incrementale, in base al ciclo iniziale e al periodo indicato (*Ciclo 1, Periodo 1, Ciclo 1, Periodo 2* e così via). Per esempio, se il numero di periodi per ogni ciclo è impostato su 3, i nuovi campi vengono creati nel modo seguente:

\$TI_IndiceTempo (Intero)	\$TI_EtiquetaTempo (Stringa)	\$TI_Ciclo (Intero)	\$TI_Periodo (Intero)
1	Ciclo 1, Periodo 1	1	1
2	Ciclo 1, Periodo 2	1	2
3	Ciclo 1, Periodo 3	1	3
4	Ciclo 2, Periodo 1	2	1
5	Ciclo 2, Periodo 2	2	2

Years

Per gli anni è possibile indicare l'anno di inizio per identificare una serie di record consecutivi, oppure selezionare Crea dai dati per specificare un campo timestamp o data che identifichi l'anno di ogni singolo record.

Figura 4-91
 Impostazioni degli intervalli di tempo per una serie annuale

Intervallo di tempo: Anni

Avvia l'assegnazione di etichette dal primo record Crea dai dati

Anno: 2000

Nuova estensione nome di campo: \$TI_ Aggiungi come: Prefisso Suffisso

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Stringa)	\$TI-Anno (Intero)
1	2000	2000
2	2001	2001
3	2002	2002
4	2003	2003
5	2004	2004

Quarters

Per le serie trimestrali è possibile specificare il mese in cui inizia l'anno fiscale. È possibile inoltre indicare il trimestre e l'anno di inizio (per esempio, T1 2000) per identificare una serie di record consecutivi, oppure selezionare Crea dai dati per scegliere un campo timestamp o data che identifichi il trimestre e l'anno di ogni singolo record.

Figura 4-92
 Impostazioni degli intervalli di tempo per le serie trimestrali

Intervallo di tempo: Trimestri

L'anno fiscale inizia: Gennaio

Avvia l'assegnazione di etichette dal primo record Crea dai dati

Anno: 2000 Trimestre: 1

Nuova estensione nome di campo: \$TI_ Aggiungi come: Prefisso Suffisso

Output campione

Per un anno fiscale che inizia in gennaio, i nuovi campi sarebbero creati e compilati nel modo seguente:

\$TI-Indice Tempo (Intero)	\$TI-Etichetta Tempo (Stringa)	\$TI-Anno (Intero)	\$TI-Trimestre (Intero con etichette)
1	T1 2000	2000	1 (T1)
2	T2 2000	2000	2 (T2)
3	T3 2000	2000	3 (T3)
4	T4 2000	2000	4 (T4)
5	T1 2001	2001	1 (T1)

Se l'anno inizia in un mese diverso da gennaio, i nuovi campi saranno come quelli riportati di seguito (presupponendo un anno fiscale che inizia in luglio). Per visualizzare le etichette che identificano i mesi di ogni trimestre, attivare la visualizzazione delle etichette dei valori facendo clic sull'icona della barra degli strumenti.

Figura 4-93
Icona etichette valori



\$TI-Indice Tempo (Intero)	\$TI-Etichetta Tempo (Stringa)	\$TI-Anno (Intero)	\$TI-Trimestre (Intero con etichette)
1	T1 2000/2001	1	1 (T1 Lug-Set)
2	T2 2000/2001	1	2 (T2 Ott-Dic)
3	T3 2000/2001	1	3 (T3 Gen-Mar)
4	T4 2000/2001	1	4 (T4 Apr-Giu)
5	T1 2001/2002	2	1 (T1 Lug-Set)

Months

È possibile selezionare l'anno e il mese d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o data che indichi il mese di ogni singolo record.

Figura 4-94
Impostazioni degli intervalli di tempo per una serie mensile

Intervallo di tempo:

Avvia l'assegnazione di etichette dal primo record Crea dai dati

Anno: Mese:

Nuova estensione nome di campo: Aggiungi come: Prefisso Suffisso

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Data)	\$TI-Anno (Intero)	\$TI-Mesi (Intero con etichette)
1	Gen 2000	2000	1 (Gennaio)
2	Feb 2000	2000	2 (Febbraio)
3	Mar 2000	2000	3 (Marzo)
4	Apr 2000	2000	4 (Aprile)
5	Mag 2000	2000	5 (Maggio)

Settimane (non periodiche)

Per una serie settimanale, è possibile selezionare il giorno della settimana in cui inizia il ciclo.

Si noti che le settimane possono essere soltanto non periodiche perché i diversi mesi, trimestri e anche gli anni non hanno necessariamente lo stesso numero di settimane. Tuttavia, i dati con timestamp possono essere facilmente aggregati o integrati a un livello settimanale per i modelli non periodici.

Figura 4-95

Impostazioni degli intervalli di tempo per una serie settimanale

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Data)	\$TI-Settimana (Intero)
1	1999-12-27	1
2	2000-01-03	2
3	2000-01-10	3
4	2000-01-17	4
5	2000-01-24	5

Il campo *\$TI-EtichettaTempo* per una settimana visualizza il primo giorno di quella settimana. Nella tabella precedente, l'utente inizia l'etichettatura dal primo gennaio del 2000. Tuttavia, la settimana inizia di lunedì mentre il primo gennaio 2000 è un sabato. Pertanto, la settimana che comprende il primo gennaio inizia il 27 dicembre 1999 e rappresenta l'etichetta del primo punto.

Il formato data determina le stringhe prodotte per il campo *\$TI-EtichettaTempo*.

Giorni a settimana

Per le misurazioni quotidiane che avvengono in un ciclo settimanale, è possibile specificare il numero di giorni a settimana e il giorno d'inizio di ogni settimana. È possibile indicare una data d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o data che indichi la data di ogni singolo record.

Figura 4-96

Impostazioni degli intervalli di tempo per una serie quotidiana

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Data)	\$TI-Settimana (Intero)	\$TI-Giorno (Intero con etichette)
1	5 Gen 2005	1	3 (Mercoledì)
2	6 Gen 2005	1	4 (Giovedì)
3	7 Gen 2005	1	5 (Venerdì)
4	10 Gen 2005	2	1 (Lunedì)
5	11 Gen 2005	2	2 (Martedì)

Nota: La settimana inizia sempre da 1 per il primo periodo di tempo e non segue il ciclo del calendario. Pertanto, la 52° settimana è seguita dalla 53°, dalla 54° e così via. La settimana non corrisponde alla settimana dell'anno, ma solo al numero di incrementi settimanali della serie.

Giorni (non periodici)

Scegliere i giorni non periodici in presenza di misurazioni quotidiane che non rientrano in un normale ciclo settimanale. È possibile indicare una data d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o data che indichi la data di ogni singolo record.

Figura 4-97

Impostazioni degli intervalli di tempo per una serie quotidiana (non periodici)

Output campione

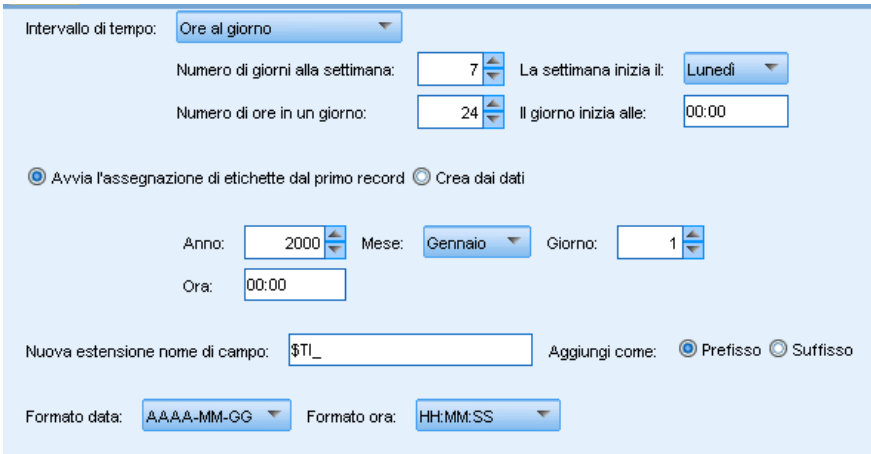
I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Data)
1	5 Gen 2005
2	6 Gen 2005
3	7 Gen 2005
4	8 Gen 2005
5	9 Gen 2005

Ore al giorno

Per le misurazioni orarie che rientrano in un ciclo giornaliero è possibile specificare il numero di giorni a settimana, il numero di ore al giorno (per esempio una giornata lavorativa di otto ore), il giorno d'inizio della settimana e l'ora d'inizio di ogni giorno. Le ore possono essere indicate con precisione al minuto in base al formato 24 ore (per esempio, 14:05 = 2:05 p.m.)

Figura 4-98
 Impostazioni degli intervalli di tempo per una serie oraria



È possibile indicare la data e l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp che identifichi la data e l'ora di ogni singolo record.

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Timestamp)	\$TI-Giorno (Intero con etichette)	\$TI-Ora (Intero con etichette)
1	5 Gen 2005 08:00:00	3 (Mercoledì)	8 (8:00)
2	5 Gen 2005 09:00:00	3 (Mercoledì)	9 (9:00)
3	5 Gen 2005 10:00:00	3 (Mercoledì)	10 (10:00)
4	5 Gen 2005 11:00:00	3 (Mercoledì)	11 (11:00)
5	5 Gen 2005 12:00	3 (Mercoledì)	12 (12:00)

Ore (non periodiche)

Scegliere questa opzione in presenza di misurazioni orarie che non rientrano in un normale ciclo quotidiano. È possibile indicare l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o ora che indichi l'ora di ogni singolo record.

Figura 4-99
 Impostazioni degli intervalli di tempo per dati annuali

Le ore si basano sul formato 24 ore (13:00 = 1:00 p.m.) e non ripartono dall'inizio (l'ora 24 è seguita dall'ora 25).

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Stringa)	\$TI-Ora (Intero con etichette)
1	8:00	8 (8:00)
2	9:00	9 (9:00)
3	10:00	10 (10:00)
4	11:00	11 (11:00)
5	12:00	12 (12:00)

Minuti al giorno

Per le misurazioni al minuto che rientrano in un ciclo quotidiano, è possibile specificare il numero di giorni alla settimana, il giorno d'inizio della settimana, il numero di ore al giorno e l'ora d'inizio del giorno. Le ore vengono indicate in base al formato 24 ore e si possono specificare con una precisione di minuti e secondi utilizzando i due punti (per esempio, 2:05:17 p.m. = 14:05:17). È possibile anche specificare il numero di minuti per gli incrementi (ogni minuto, ogni due minuti e così via; l'incremento deve essere un valore divisibile per 60).

Figura 4-100
 Impostazioni degli intervalli di tempo per i minuti al giorno

È possibile indicare la data e l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp che identifichi la data e l'ora di ogni singolo record.

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Timestamp)	\$TI-Minuto
1	2005-01-05 08:00:00	0
2	2005-01-05 08:01:00	1
3	2005-01-05 08:02:00	2
4	2005-01-05 08:03:00	3
5	2005-01-05 08:04:00	4

Minuti (non periodici)

Scegliere questa opzione in presenza di misurazioni al minuto che non rientrano in un normale ciclo quotidiano. È possibile specificare il numero di minuti per gli incrementi (ogni minuto, ogni due minuti e così via; il valore indicato deve essere un numero divisibile per 60).

Figura 4-101
Impostazioni degli intervalli di tempo per i minuti (non periodici)

È possibile indicare l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o ora che identifichi l'ora di ogni singolo record.

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Stringa)	\$TI-Minuto
1	8:00	0
2	8:01	1
3	8:02	2
4	8:03	3
5	8:04	4

- La stringa *EtichettaTempo* viene creata utilizzando i due punti per separare le ore dai minuti. L'ora non riparte dall'inizio: l'ora 24 è seguita dall'ora 25.
- I minuti vengono incrementati in base al valore impostato nella finestra di dialogo. Per esempio, se l'incremento è 2, il valore di *EtichettaTempo* sarà 8:00, 8:02 e così via, e i minuti saranno 0, 2 e così via.

Secondi al giorno

Per gli intervalli di secondi che rientrano in un ciclo quotidiano, è possibile specificare il numero di giorni alla settimana, il giorno d'inizio della settimana, il numero di ore al giorno e l'ora d'inizio del giorno. Le ore vengono indicate in base al formato 24 ore e si possono specificare con una precisione di minuti e secondi utilizzando i due punti (per esempio, 2:05:17 p.m. = 14:05:17). È possibile anche specificare il numero di secondi per gli incrementi (ogni secondo, ogni due secondi e così via; il valore indicato deve essere un numero divisibile per 60).

Figura 4-102
 Impostazioni degli intervalli di tempo per i secondi al giorno

È possibile specificare la data e l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp che indichi la data e l'ora di ogni singolo record.

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Timestamp)	\$TI-Minuto	\$TI-Secondo
1	2005-01-05 08:00:00	0	0
2	2005-01-05 08:00:01	0	1
3	2005-01-05 08:00:02	0	2
4	2005-01-05 08:00:03	0	3
5	2005-01-05 08:00:04	0	4

Secondi (non periodici)

Scegliere questa opzione in presenza di misurazioni eseguite al secondo che non rientrano in un normale ciclo quotidiano. È possibile specificare il numero di secondi per gli incrementi (ogni secondo, ogni due secondi e così via; il valore indicato deve essere un numero divisibile per 60).

Figura 4-103
Impostazioni degli intervalli di tempo per i secondi (non periodici)

Specificare l'ora d'inizio per identificare una serie consecutiva di record oppure selezionare Crea dai dati per scegliere un campo timestamp o ora che identifichi l'ora di ogni singolo record.

Output campione

I nuovi campi vengono creati nel modo seguente:

\$TI-IndiceTempo (Intero)	\$TI-EtichettaTempo (Stringa)	\$TI-Minuto	\$TI-Secondo
1	8:00:00	0	0
2	8:00:01	0	1
3	8:00:02	0	2
4	8:00:03	0	3
5	8:00:04	0	4

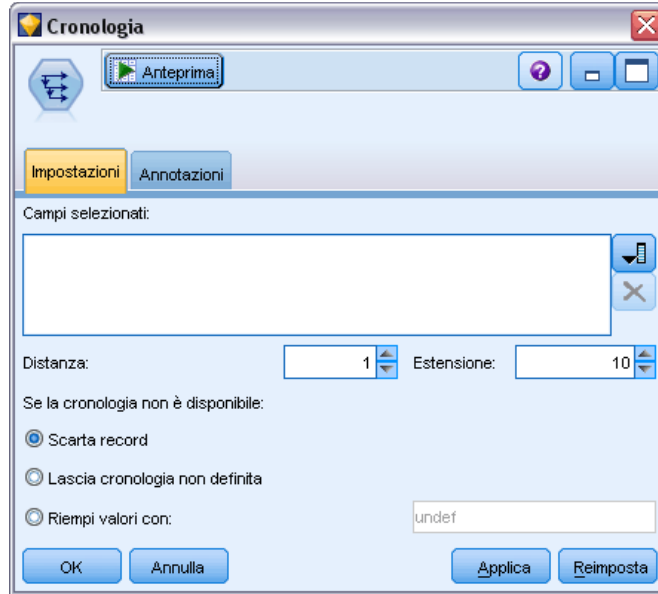
- La stringa *EtichettaTempo* viene creata utilizzando i due punti per separare le ore dai minuti e i minuti dai secondi. L'ora non riparte dall'inizio: l'ora 24 è seguita dall'ora 25.
- I secondi vengono incrementati del numero indicato come incremento. Se l'incremento è 2, il valore di *EtichettaTempo* sarà 08:00:00, 08:00:02 e così via, e i secondi saranno 0, 2 e così via.

Nodo Cronologia

I nodi Cronologia sono utilizzati in genere per dati sequenziali, per esempio per dati di serie storica. Consentono di creare nuovi campi contenenti dati dei campi di record precedenti. Quando si utilizza un nodo Cronologia, è possibile che si desideri preordinare i dati in base a un determinato campo. A questo scopo, è possibile utilizzare un nodo Ordina.

Impostazione delle opzioni relative al nodo Cronologia

Figura 4-104
Finestra di dialogo del nodo Cronologia



Campi selezionati. Il pulsante di selezione dei campi (a destra della casella di testo) permette di selezionare i campi per i quali si desidera esaminare la cronologia. Ogni campo selezionato viene utilizzato per creare nuovi campi per tutti i record nell'insieme di dati.

Offset. Specificare il record che precede quello corrente dal quale si desidera estrarre i valori di campo cronologici. Se Distanza è impostata su 3, per esempio, al passaggio di ciascun record in questo nodo verranno inclusi nel record corrente i valori di campo del terzo record precedente. Utilizzare le impostazioni Estensione per specificare il punto a partire dal quale verranno estratti i record. Utilizzare le frecce per modificare il valore della distanza.

Estensione. Specificare il numero di record precedenti a partire dal quale si desidera estrarre i valori. Per esempio, se Distanza è impostata su 3 ed Estensione su 5, a ciascun record che passa per il nodo verranno aggiunti cinque campi per ciascun campo specificato nell'elenco dei campi selezionati. Ciò significa che quando il nodo elabora il record 10, verranno aggiunti i campi dal record 7 al record 3. Utilizzare le frecce per modificare il valore dell'estensione.

Se la cronologia non è disponibile. Selezionare una delle opzioni seguenti per gestire i record senza valori cronologici. In genere si tratta dei record presenti all'inizio dell'insieme di dati, per i quali non esistono record precedenti da utilizzare come cronologia.

- **Scarta record.** Selezionare questa opzione per scartare i record nel caso in cui per il campo selezionato non esista un valore cronologico.

- **Lascia cronologia non definita.** Selezionare questa opzione per mantenere i record in cui non sia disponibile un valore cronologico. Il campo cronologico verrà riempito con un valore non definito, visualizzato come `$null$`.
- **Riempi valori con.** Specificare il valore o la stringa da utilizzare per i record in cui non sia disponibile un valore cronologico. Il valore sostitutivo di default è `undef`, il valore nullo di sistema. I valori nulli sono visualizzati utilizzando la stringa `$null$`.

Quando si seleziona un valore sostitutivo, per garantire un'esecuzione corretta è necessario tenere presente le seguenti regole:

- I campi selezionati devono essere dello stesso tipo di archiviazione.
- Se tutti i campi selezionati includono un'archiviazione numerica, il valore sostitutivo deve essere analizzato come numero intero.
- Se tutti i campi selezionati includono un'archiviazione reale, il valore sostitutivo deve essere analizzato come numero reale.
- Se tutti i campi selezionati includono un'archiviazione simbolica, il valore sostitutivo deve essere analizzato come stringa.
- Se tutti i campi selezionati includono un'archiviazione di data/ora, il valore sostitutivo deve essere analizzato come campo data/ora.

Se qualcuna delle condizioni sopra riportate non viene soddisfatta, durante l'esecuzione del nodo Cronologia verrà generato un errore.

Nodo Ordina campi

Il nodo Ordina campi consente di definire l'ordine naturale utilizzato per visualizzare i campi a valle. Tale ordine incide sulla visualizzazione dei campi in vari contesti, quali tabelle, elenchi e Selezione campi. Questa operazione risulta utile, per esempio, se si desidera evidenziare i campi più interessanti negli insiemi di dati di grandi dimensioni.

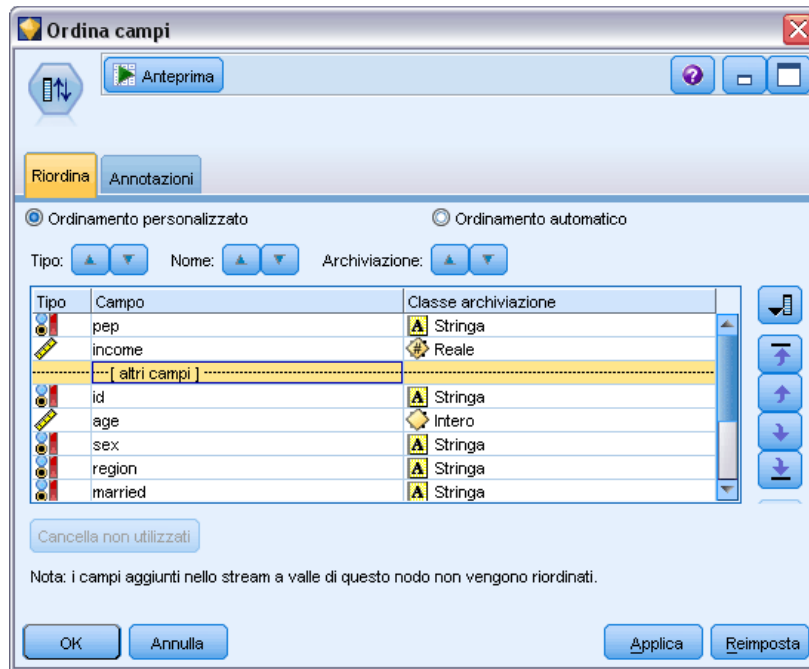
Impostazione delle opzioni Ordina campi

I campi possono essere riordinati in due modi: ordinamento personalizzato e ordinamento automatico.

Ordinamento personalizzato

Selezionare Ordinamento personalizzato per attivare una tabella con i nomi e i tipi di campo in cui è possibile visualizzare tutti i campi e utilizzare i pulsanti con le frecce per creare un ordine personalizzato.

Figura 4-105
 Riordinamento per visualizzare prima i campi desiderati



Per riordinare i campi:

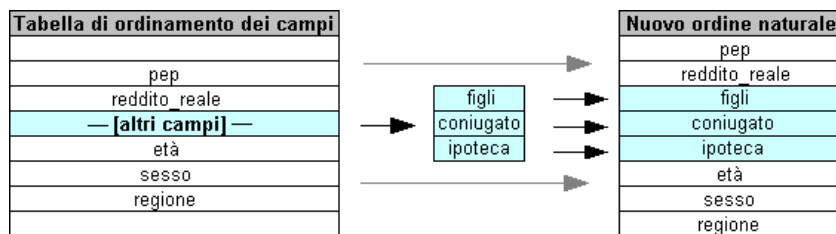
- ▶ Selezionare un campo nella tabella. Utilizzare il metodo Ctrl-clic per selezionare più campi.
- ▶ Utilizzare i pulsanti con le frecce per spostare i campi verso l'alto o verso il basso di una riga.
- ▶ Utilizzare i pulsanti con le frecce e una riga per spostare i campi all'inizio o alla fine dell'elenco.
- ▶ Specificare l'ordine dei campi non inclusi in questo contesto spostando la riga di divisione verso l'alto o verso il basso, indicata come [altri campi].

Altri campi. La riga di divisione [altri campi] consente di suddividere la tabella in due parti.

- I campi visualizzati al di sopra della riga di divisione verranno posizionati (in base alla visualizzazione nella tabella) prima di tutti gli ordini naturali utilizzati per visualizzare i campi a valle di questo nodo.
- I campi visualizzati al di sotto della riga di divisione verranno posizionati (in base alla visualizzazione nella tabella) dopo tutti gli ordini naturali utilizzati per visualizzare i campi a valle di questo nodo.

Figura 4-106

Diagramma che mostra il modo in cui gli "altri campi" vengono incorporati nel nuovo ordine dei campi



- Tutti gli altri campi non presenti nella tabella di ordinamento dei campi verranno visualizzati tra i campi "superiore" e "inferiore" come indicato dalla posizione della riga di divisione.

Sono disponibili altre opzioni di ordinamento personalizzate, tra cui:

- Ordinare i campi in modo crescente o decrescente facendo clic sulle frecce poste sopra ogni intestazione di colonna (Tipo, Nome e Classe archiviazione). Quando si ordina in base alle colonne, i campi non specificati in questo punto (indicati dalla riga [altri campi]) vengono ordinati per ultimi in base all'ordine naturale.
- Fare clic su Cancella non utilizzati per eliminare tutti i campi non utilizzati dal nodo Ordina campi. I campi non utilizzati vengono visualizzati nella tabella con un carattere rosso. Ciò indica che il campo è stato eliminato in operazioni a monte.
- Specificare l'ordinamento per tutti i campi nuovi (contrassegnati da una saetta per indicare un campo nuovo o non specificato). Se si fa clic su OK o Applica, l'icona scompare.

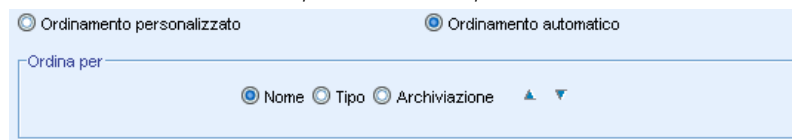
Nota: se vengono aggiunti campi a monte dopo l'applicazione di un ordine personalizzato, i nuovi campi verranno accodati alla fine dell'elenco personalizzato.

Ordinamento automatico

Selezionare Ordinamento automatico per specificare un parametro per l'ordinamento. Le opzioni della finestra di dialogo cambiano in modo dinamico per mostrare opzioni relative all'ordinamento automatico.

Figura 4-107

Riordinamento di tutti i campi mediante le opzioni di ordinamento automatico



Ordina per. Selezionare uno dei tre modi di ordinare i campi disponibili nel nodo Riordina. I pulsanti con le frecce indicano se l'ordine sarà crescente o decrescente. Selezionare uno dei modi per eseguire una modifica.

- Name
- Type
- Storage

I campi aggiunti a monte del nodo Ordina campi dopo l'applicazione dell'ordinamento automatico verranno disposti automaticamente nella posizione corretta in base al tipo di ordinamento selezionato.

Nodi Grafico

Caratteristiche comuni dei nodi Grafico

Numerose fasi del processo di data mining utilizzano grafici e diagrammi per esaminare i dati immessi in IBM® SPSS® Modeler. Per esempio, è possibile connettere un nodo Plot o Distribuzione a una sorgente dati per ottenere informazioni su distribuzioni e tipi di dati ed eseguire quindi manipolazioni su record e campi per preparare i dati per operazioni di creazione di modelli a valle. I grafici vengono comunemente utilizzati anche per controllare la distribuzione e le relazioni tra i nuovi campi derivati.

Dopo avere configurato le opzioni relative a un nodo Grafico, è possibile eseguirlo dall'interno della finestra di dialogo o come parte di uno stream. Nella finestra del grafico così generata, è possibile generare nodi Nuovo campo (Insieme e Flag) e nodi Seleziona in base a una selezione o a un'area di dati, dividendo effettivamente i dati in "sottoinsiemi". Questa potente funzione può essere utilizzata, per esempio, per individuare ed escludere i valori anomali.

La palette Grafici contiene i seguenti nodi:



Il nodo Lavagna grafica offre numerosi tipi di grafici diversi in un unico nodo. Con questo nodo è possibile scegliere i campi di dati da esplorare e selezionare quindi un grafico fra quelli disponibili per i dati selezionati. Il nodo esclude automaticamente tutti i tipi di grafici non adatti ai campi selezionati. [Per ulteriori informazioni, vedere l'argomento Nodo Lavagna grafica a pag. 239.](#)



Il nodo Plot mostra la relazione tra campi numerici. È possibile creare un grafico plot utilizzando punti (un grafico a dispersione) oppure linee. [Per ulteriori informazioni, vedere l'argomento Nodo Plot a pag. 252.](#)



Il nodo Distribuzione mostra l'occorrenza di valori simbolici (categoriali), per esempio tipo o genere di ipoteca. In genere è possibile utilizzare un nodo Distribuzione per mostrare sbilanciamenti nei dati, che possono essere successivamente corretti con un nodo Bilanciamento prima di creare un modello. [Per ulteriori informazioni, vedere l'argomento Nodo Distribuzione a pag. 261.](#)



Il nodo Istogramma mostra l'occorrenza dei valori per i campi numerici. Viene spesso utilizzato per analizzare i dati prima delle manipolazioni e della generazione del modello. Come il nodo Distribuzione, anche il nodo Istogramma viene frequentemente utilizzato per rivelare sbilanciamenti nei dati. [Per ulteriori informazioni, vedere l'argomento Scheda Plot del nodo Istogramma a pag. 267.](#)



Il nodo Raccolta mostra la distribuzione dei valori di un campo numerico in relazione ai valori di un altro, ovvero crea grafici simili a istogrammi. È utile per illustrare una variabile o un campo i cui valori vengono modificati nel tempo. La grafica 3-D consente inoltre di includere un asse simbolico che visualizza le distribuzioni per categoria. [Per ulteriori informazioni, vedere l'argomento Scheda Plot del nodo Raccolta a pag. 271.](#)



Un nodo Plot multiplo crea un grafico che consente di visualizzare più campi Y in un singolo campo X . I campi Y sono rappresentati come linee colorate e ognuno di essi equivale a un nodo Plot con lo Stile impostato su Linea e la Modalità X impostata su Ordina. I plot multipli sono utili quando si desidera esplorare la fluttuazione di numerose variabili nel tempo. [Per ulteriori informazioni, vedere l'argomento Nodo Plot multiplo a pag. 275.](#)



Il nodo Web illustra l'intensità della relazione tra valori di due o più campi simbolici (categoriali). Il grafico utilizza linee di spessore diverso per indicare l'intensità della connessione. Un nodo Web può essere utilizzato, per esempio, per analizzare la relazione tra l'acquisto di vari oggetti in un sito di e-commerce. [Per ulteriori informazioni, vedere l'argomento Nodo Web a pag. 280.](#)



Il nodo Plot tempo visualizza uno o più insiemi di dati di serie storica. In genere, si utilizza prima un nodo Intervalli di tempo per creare un campo *EtichettaTempo*, che viene utilizzato per attribuire un'etichetta all'asse x . [Per ulteriori informazioni, vedere l'argomento Nodo Plot tempo a pag. 291.](#)



Il nodo Valutazione facilita la valutazione e il confronto di modelli predittivi. Il grafico di valutazione mostra il comportamento dei modelli nella previsione di particolari risultati. Ordina i record in base al valore previsto e alla confidenza della previsione, quindi li suddivide in gruppi di uguale dimensione (**quantili**) e infine rappresenta il valore del criterio commerciale per ciascun quantile, dal più alto al più basso. I modelli multipli sono mostrati nel grafico come linee separate. [Per ulteriori informazioni, vedere l'argomento Nodo Valutazione a pag. 295.](#)

Dopo aver aggiunto un nodo Grafico a uno stream, è possibile fare doppio clic su di esso per aprire una finestra di dialogo in cui specificare le opzioni. La maggior parte dei grafici contiene opzioni univoche, presentate in una o più schede. Esistono inoltre diverse opzioni comuni a tutti i grafici. Ulteriori informazioni sulle opzioni comuni sono disponibili in sezioni seguenti.

Dopo avere configurato le opzioni relative a un nodo Grafico, è possibile eseguirlo dall'interno della finestra di dialogo o come parte di uno stream. Nella finestra del grafico così generata, è possibile generare nodi Nuovo campo (Insieme e Flag) e nodi Seleziona in base a una selezione o a un'area di dati, dividendo effettivamente i dati in "sottoinsiemi". Questa potente funzione può essere utilizzata, per esempio, per individuare ed escludere i valori anomali.

Effetti estetici, sovrapposizioni, riquadri e animazioni

Sovrapposizioni ed effetti estetici

Gli elementi estetici (e le sovrapposizioni) aggiungono dimensionalità alle visualizzazioni. L'effetto di un elemento estetico (raggruppamento o sovrapposizione) dipende dal tipo di visualizzazione, dal tipo di campo (variabile) e dal tipo e dalle statistiche dell'elemento grafico. Per esempio, per il colore è possibile utilizzare un campo categoriale per raggruppare i punti di un grafico a dispersione o per creare le barre di un grafico a barre sovrapposte. Per il colore è inoltre possibile utilizzare un intervallo numerico continuo per indicare i valori dell'intervallo per ogni punto di un grafico a dispersione.

Per trovare l'elemento estetico e la sovrapposizione adatti alle proprie esigenze, occorre eseguire delle prove. Le seguenti descrizioni possono facilitare la scelta dell'elemento adeguato.

Nota: non tutti gli elementi estetici o sovrapposizioni sono disponibili per tutti i tipi di visualizzazione.

- **Colore.** Quando il colore è definito da un campo categoriale, suddivide la visualizzazione in base alle singole categorie, un colore per ciascuna categoria. Quando il colore è un intervallo numerico continuo, varia in base al valore del campo intervallo. Se l'elemento grafico (per esempio una barra o una casella) rappresenta più di un record/caso e per il colore viene utilizzato un campo, il colore varia in base alla *media* del campo intervallo.
- **Forma.** La forma è definita da un campo categoriale che suddivide la visualizzazione in elementi di forme diverse, una per ciascuna categoria.
- **Trasparenza.** Quando la trasparenza è definita da un campo/una variabile categoriale, suddivide la visualizzazione in base alle singole categorie, un livello di trasparenza per ciascuna categoria. Quando la trasparenza è un intervallo numerico continuo, varia in base al valore del campo/della variabile intervallo. Se l'elemento grafico (per esempio una barra o una casella) rappresenta più di un record/caso e per la trasparenza viene utilizzato un campo/una variabile intervallo, il colore varia in base alla *media* del campo/della variabile intervallo. Al valore massimo, gli elementi grafici sono completamente trasparenti, mentre a quello minimo sono completamente opachi.
- **Etichetta dati.** Le etichette dati sono definite da qualsiasi tipo di campo i cui valori sono utilizzati per creare etichette che sono collegate agli elementi grafici.
- **Dimensione.** Quando la dimensione è definita da un campo categoriale, suddivide la visualizzazione in base alle singole categorie, una dimensione per ciascuna categoria. Quando la dimensione è un intervallo numerico continuo, varia in base al valore del campo intervallo. Se l'elemento grafico, per esempio una barra o una casella, rappresenta più di un record/caso e per la dimensione viene utilizzato un campo intervallo, la dimensione varia in base alla *media* del campo intervallo.

Figura 5-1

Grafico con elemento estetico di sovrapposizione di colori

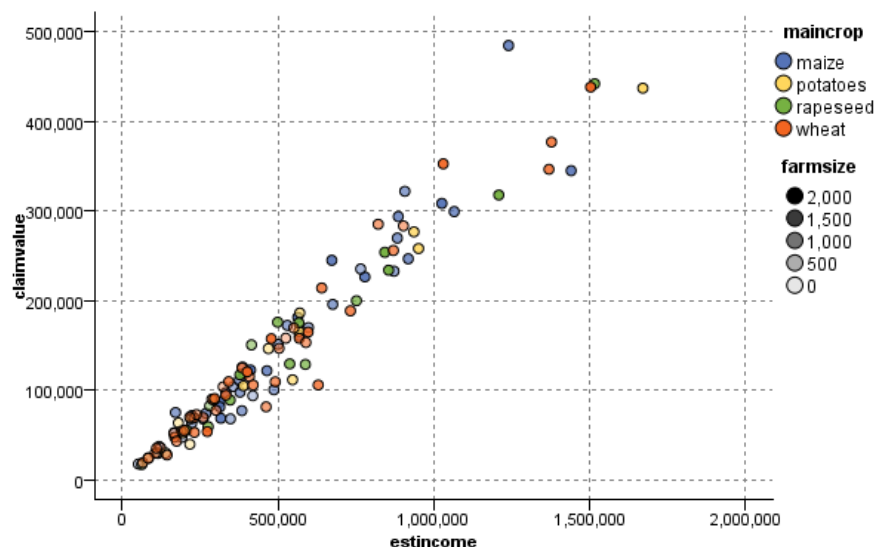
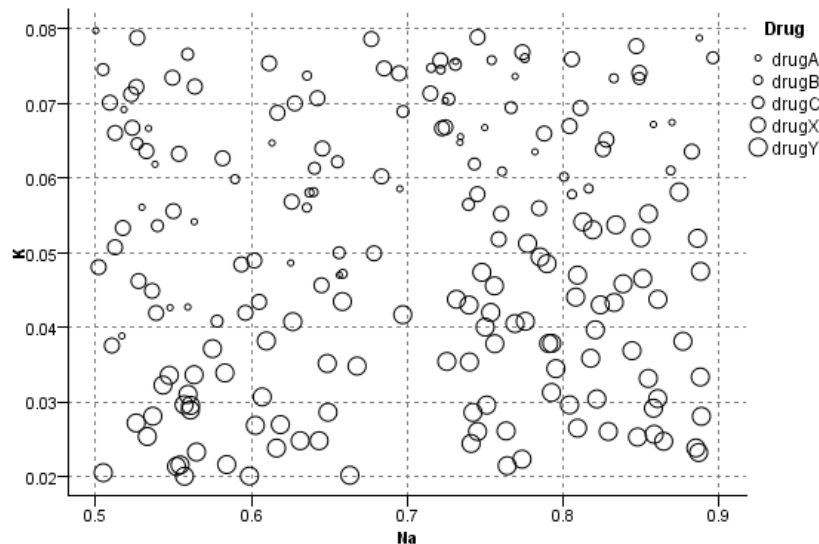


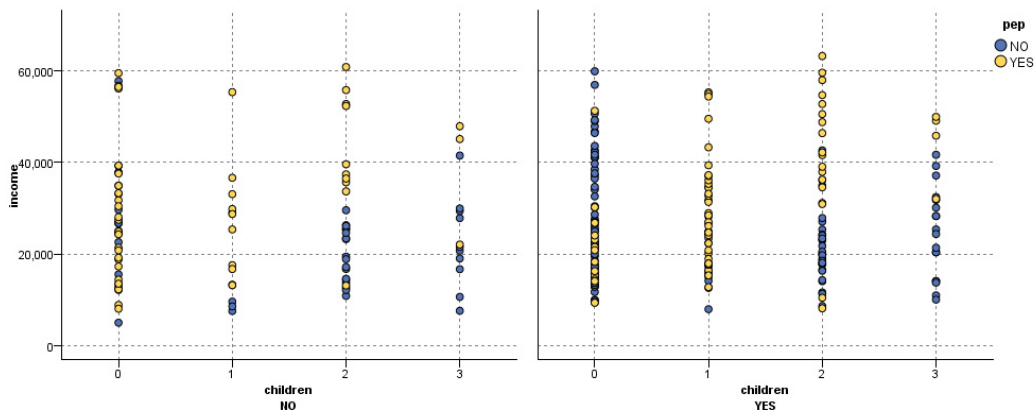
Figura 5-2
Grafico con elemento estetico di sovrapposizione di dimensioni



Utilizzo di riquadri e animazioni

Definizione dei riquadri. I pannelli, denominati anche “facet”, creano una tabella di grafici. Per ogni categoria dei campi/delle variabili dei pannelli viene generato un grafico, ma tutti i pannelli vengono visualizzati contemporaneamente. I pannelli sono utili per verificare se la visualizzazione è soggetta alle condizioni dei campi/delle variabili dei pannelli. Per esempio, è possibile creare pannelli in un istogramma in base al sesso per stabilire se le distribuzioni della frequenza sono uguali per maschi e femmine. In altre parole, è possibile verificare se lo stipendio è soggetto a differenze tra i sessi. Selezionare una variabile/un campo categoriale per la creazione dei pannelli.

Figura 5-3
Grafico con pannelli per stato civile coniugato (SÌ/NO)



Animazione. L'animazione assomiglia ai pannelli per il fatto che vengono creati più grafici dai valori del campo/della variabile di animazione, ma tali grafici non vengono mostrati insieme. È invece possibile utilizzare i controlli in modalità Esplora per animare l'output e scorrere una sequenza di singoli grafici. Inoltre, a differenza dei pannelli, l'animazione non richiede un campo/una variabile categoriale. È possibile specificare un campo/una variabile continuo/a i cui valori vengono suddivisi automaticamente in intervalli. È possibile variare la dimensione dell'intervallo con i controlli dell'animazione in modalità Esplora. Non tutte le visualizzazioni offrono la funzionalità di animazione.

Figura 5-4

Grafico plot animato mediante l'utilizzo di una variabile con tre categorie; dispositivo di scorrimento su pressione sanguigna bassa

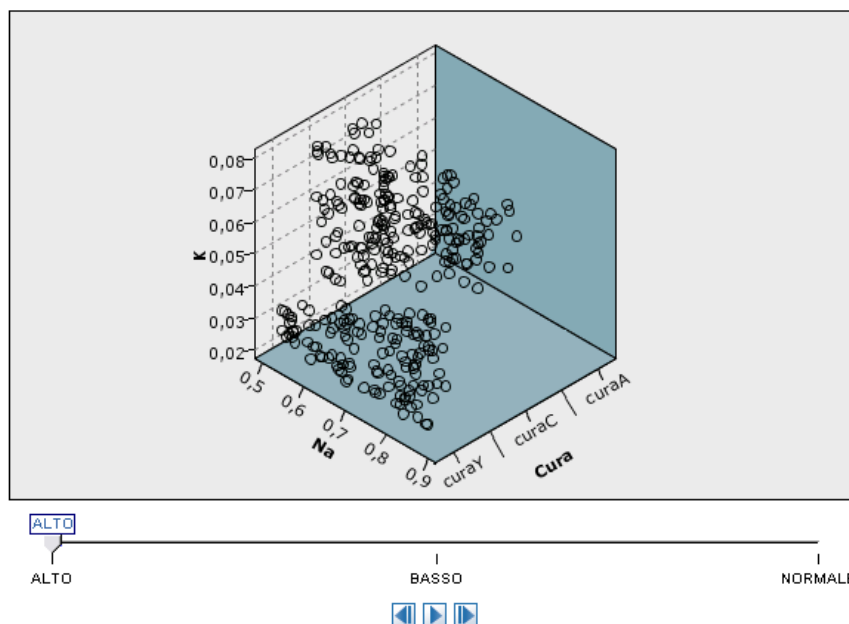


Figura 5-5

Grafico plot animato mediante l'utilizzo di una variabile con tre categorie; dispositivo di scorrimento su pressione sanguigna normale

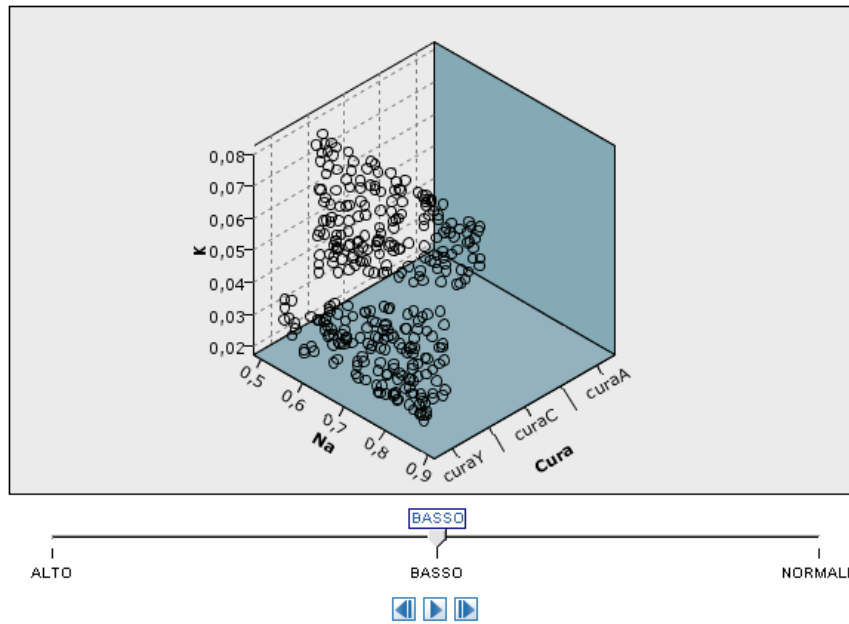
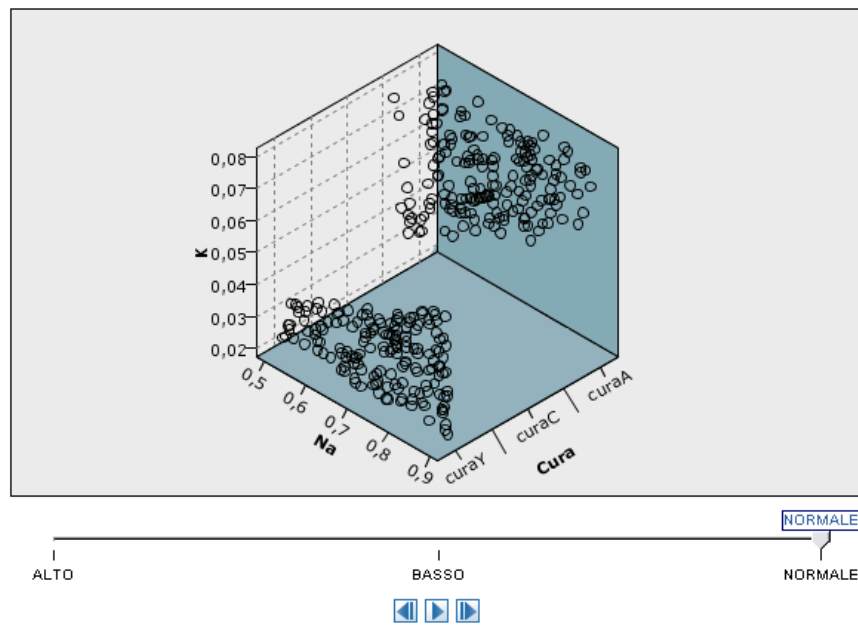


Figura 5-6

Grafico plot animato mediante l'utilizzo di una variabile con tre categorie; dispositivo di scorrimento su pressione sanguigna alta



Utilizzo della scheda Output

Per tutti i tipi di grafici è possibile specificare le seguenti opzioni relative al nome di file e alla visualizzazione dei grafici generati.

Nota: i grafici del nodo Distribuzione dispongono di impostazioni aggiuntive.

Nome output. Specifica il nome del grafico prodotto durante l'esecuzione del nodo. Automatico sceglie un nome in base al nodo che genera l'output. Se necessario, è possibile selezionare Personalizzato per specificare un nome diverso.

Output su schermo. Selezionare questa opzione per generare e visualizzare il grafico in una nuova finestra.

Output su file. Selezionare questa opzione per salvare l'output come file.

- **Grafico di output.** Selezionare questa opzione per generare l'output in formato grafico. Disponibile solo nei nodi Distribuzione.
- **Tabella di output.** Selezionare questa opzione per generare l'output in formato tabella. Disponibile solo nei nodi Distribuzione.
- **Nome file.** Specificare un nome di file utilizzato per il grafico o la tabella generati. Utilizzare il pulsante con i puntini di sospensione (...) per specificare un file e una posizione specifici.
- **Tipo di file.** Selezionare il tipo di file dall'elenco a discesa. Per tutti i nodi Grafico, fatta eccezione per il nodo Distribuzione con l'opzione Tabella di output, i tipi di file disponibili per i grafici sono:

- | | | |
|-----------------|----------------|---|
| - Bitmap (.bmp) | - PNG (.png) | - Oggetto di output (.cou) |
| - JPEG (.jpg) | - HTML (.html) | - Documento ViZml (.xml) da utilizzare in altre applicazioni IBM® SPSS® Statistics. |

Per l'opzione Tabella di output del nodo Distribuzione, i tipi di file disponibili sono:

- | | | | |
|---|-------------------------------------|----------------|----------------------------|
| - Dati delimitati da tabulazione (.tab) | - Dati delimitati da virgola (.csv) | - HTML (.html) | - Oggetto di output (.cou) |
|---|-------------------------------------|----------------|----------------------------|

Impagina output. Quando si salva l'output come HTML, questa opzione è attivata per permettere di controllare le dimensioni delle singole pagine HTML (valida solo per il nodo Distribuzione).

Righe per pagina. Quando viene selezionata l'opzione Impagina output, questa opzione è attivata per permettere di determinare la lunghezza di ogni pagina HTML. L'impostazione di default è 400 righe (valida solo per il nodo Distribuzione).

Utilizzo della scheda Annotazioni.

Utilizzata per tutti i nodi, questa scheda fornisce opzioni per rinominare i nodi, uno strumento personalizzato e la possibilità di archiviare annotazioni estese. [Per ulteriori informazioni, vedere l'argomento Annotazioni in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

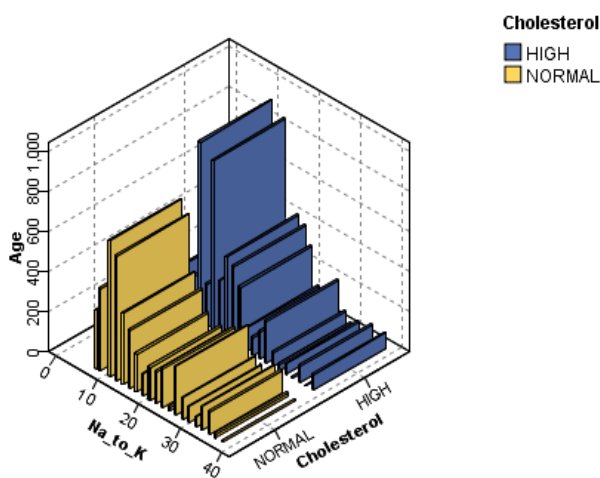
Grafici 3-D

In IBM® SPSS® Modeler, i grafici del nodo Plot e del nodo Raccolta consentono di visualizzare le informazioni su un terzo asse. Questa funzione offre all'utente una maggiore flessibilità in fase di visualizzazione dei dati, in quanto consente di selezionare sottoinsiemi o derivare nuovi campi per la creazione di modelli.

Dopo che è stato creato un grafico 3-D, è possibile fare clic su di esso e trascinarlo in modo da ruotarlo e visualizzarlo da tutte le angolature.

Figura 5-7

Grafico del nodo Raccolta con assi x, y e z



In SPSS Modeler esistono due modalità di creazione per i grafici 3-D: tramite rappresentazione delle informazioni su un terzo asse (grafici 3-D effettivi) o tramite visualizzazione di grafici con effetti 3-D. Entrambi i metodi sono disponibili per i grafici plot e di raccolta.

Per rappresentare le informazioni su un terzo asse

- ▶ Nella finestra di dialogo del nodo Grafico, fare clic sulla scheda Plot.
- ▶ Fare clic sul pulsante 3D per attivare le opzioni per l'asse z.
- ▶ Utilizzare il pulsante Selezione campi per selezionare un campo per l'asse z. In alcuni casi, per l'asse z è possibile utilizzare solo campi simbolici. Selezione campi visualizzerà i campi appropriati.

Per aggiungere effetti 3-D a un grafico

- ▶ Dopo aver creato un grafico, fare clic sulla scheda Grafico nella finestra di output.
- ▶ Fare clic sul pulsante 3-D per visualizzare un grafico tridimensionale.

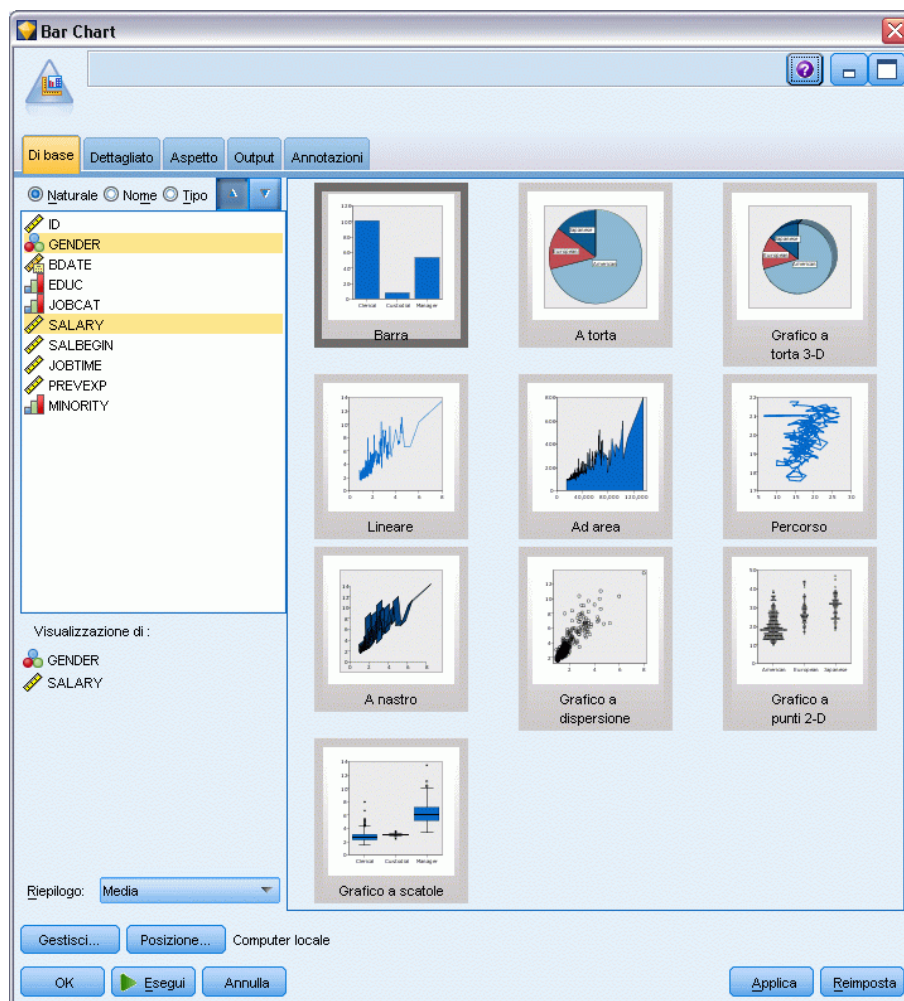
Nodo Lavagna grafica

Il nodo Lavagna grafica consente di scegliere tra numerosi tipi di output grafico (grafici a barre, grafici a torta, istogrammi, grafici a dispersione, mappe termiche, ecc.) in un singolo nodo. Per iniziare, scegliere i campi di dati da analizzare nella prima scheda. Il nodo visualizza un elenco di tipi di grafici adatti ai dati prescelti. Il nodo esclude automaticamente tutti i tipi di grafici non adatti ai campi selezionati. È possibile definire opzioni dettagliate o più avanzate per i grafici nella scheda Dettagliato.

Nota: per modificare il nodo Lavagna grafica o selezionare dei tipi di grafici è necessario collegare il nodo Lavagna grafica a uno stream contenente dei dati.

Figura 5-8

Finestra di dialogo del nodo Lavagna grafica, scheda Di base



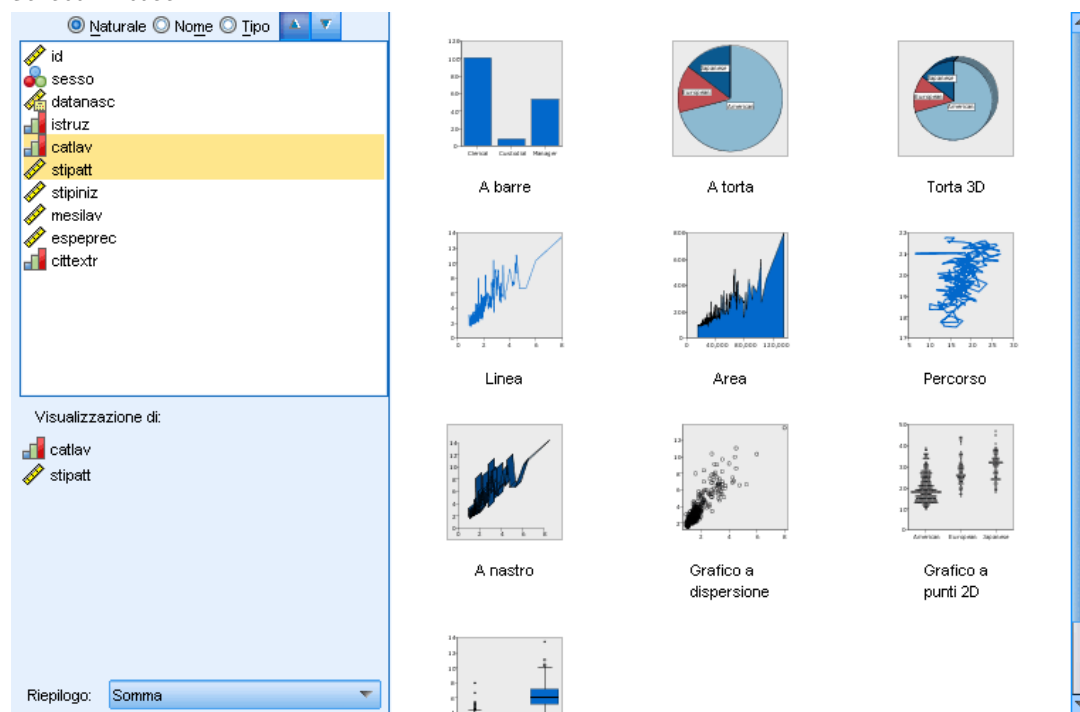
Due pulsanti consentono di controllare quali modelli (e fogli di stile) di visualizzazione sono disponibili:

Gestisci. Gestire modelli e i fogli di stile di visualizzazione sul computer. È possibile importare, esportare, rinominare ed eliminare modelli e i fogli di stile di visualizzazione sul computer locale. Per ulteriori informazioni, vedere l'argomento [Gestione dei modelli, dei fogli di stile e dei file delle mappe](#) a pag. 251.

Posizione. Modificare la posizione in cui sono archiviati i modelli e i fogli di stile di visualizzazione. La posizione corrente viene elencata alla destra del pulsante. Per ulteriori informazioni, vedere l'argomento [Impostazione della posizione di modelli, fogli di stile e mappe](#) a pag. 249.

Scheda Di base del nodo Lavagna grafica

Figura 5-9
Scheda Di base



Se non si è certi di quale sia il tipo di visualizzazione ottimale per rappresentare i propri dati, utilizzare la scheda Di base. Quando si selezionano i dati, viene visualizzato un sottoinsieme dei tipi di visualizzazione adatti per rappresentarli.

- Selezionare uno o più campi (variabili) nell'elenco. Per selezionare più campi, tenere premuto il tasto Ctrl e fare clic con il mouse.

Si noti che il livello di misurazione del campo determina i tipi di visualizzazione disponibili. Per cambiare il livello di misurazione, fare clic con il pulsante destro del mouse sul campo nell'elenco e scegliere un'opzione. Per maggiori informazioni sui tipi di livelli di misurazione disponibili, vedere [Tipi di campo \(variabile\)](#) a pag. 243.

- ▶ Selezionare un tipo di visualizzazione.
- ▶ Per alcune visualizzazioni è possibile selezionare una statistica riassuntiva. Sono disponibili diversi sottoinsiemi di statistiche a seconda del fatto che la statistica sia basata sul conteggio o calcolata a partire da un campo continuo. Le statistiche disponibili dipendono anche dal modello. La fase successiva è seguita da un elenco completo delle statistiche che potrebbero essere disponibili.
- ▶ Se si desidera definire altre opzioni, quali gli elementi estetici facoltativi e i campi del pannello, fare clic su [Dettagliato](#). [Per ulteriori informazioni, vedere l'argomento Lavagna grafica Scheda Dettagliato](#) a pag. 245.

Statistiche riassuntive calcolate a partire da un campo continuo

- **Mean.** Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.
- **Median.** È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.
- **Mode.** Il valore o i valori che ricorrono più frequentemente. Se più valori condividono la maggiore ricorrenza, ognuno di essi è una moda.
- **Minimo.** Il valore più basso assunto da una variabile numerica.
- **Massimo.** Il valore più alto di una variabile numerica.
- **Intervallo.** Differenza tra i valori minimi e massimi.
- **Intervallo medio.** Il punto medio dell'intervallo, vale a dire, il valore la cui differenza rispetto al punto minimo è uguale alla differenza rispetto al punto massimo.
- **Sum.** La somma o il totale di tutti i valori non mancanti di tutti i casi.
- **Somma cumulata.** La somma cumulata dei valori. Ogni elemento grafico mostra la somma di un sottogruppo più la somma totale di tutti i gruppi precedenti.
- **Somma percentuale.** La percentuale all'interno di ogni sottogruppo basata su un campo sommato confrontato alla somma di tutti i gruppi.
- **Somma percentuale cumulata.** La percentuale cumulativa all'interno di ogni sottogruppo basata su un campo sommato confrontato alla somma di tutti i gruppi. Ogni elemento grafico mostra la percentuale di un sottogruppo più la percentuale totale di tutti i gruppi precedenti.
- **Variance.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.

- **Deviazione standard.** La radice quadrata della varianza. La deviazione standard è una misura della dispersione intorno alla media espressa nella stessa unità di misura delle osservazioni. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Errore standard.** Una misura di quanto il valore di una statistica del test può variare da campione a campione. È la deviazione standard della distribuzione di campionamento di una statistica. L'errore standard della media, ad esempio, è la deviazione standard delle medie campionarie.
- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che, rispetto a una distribuzione normale, le osservazioni sono più raggruppate intorno al centro della distribuzione e hanno code più sottili fino ai valori estremi della distribuzione; a quel punto, le code della distribuzione leptocurtica sono più spesse rispetto a una distribuzione normale. Una curtosi negativa indica che, rispetto a una distribuzione normale, le osservazioni sono meno raggruppate e hanno code più spesse fino ai valori estremi della distribuzione; a quel punto, le code della distribuzione platocurtica sono più sottili rispetto a una distribuzione normale.
- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Le seguenti statistiche di regione potrebbero restituire più elementi grafici per ogni sottogruppo. Quando si utilizzano gli elementi grafici intervallo, area o limite, una statistica di regione restituisce un elemento grafico che mostra l'intervallo. Tutti gli altri elementi grafici danno come risultato due elementi separati, uno che mostra l'inizio dell'intervallo e l'altro che mostra la fine dell'intervallo.

- **Regione: Intervallo.** L'intervallo di valori tra il valore minimo e il valore massimo.
- **Regione: Intervallo di confidenza per la media al 95%.** Un intervallo di valori con il 95% di probabilità di includere la media della popolazione.
- **Regione: Intervallo di confidenza per il singolo al 95%.** Un intervallo di valori con il 95% di probabilità di includere il valore previsto dato il caso singolo.
- **Regione: 1 Deviazione standard sopra/sotto la media.** Un intervallo di valori compreso tra 1 deviazione standard sopra e sotto la media.
- **Regione: 1 Errore standard sopra/sotto la media.** Un intervallo di valori compreso tra 1 errore standard sopra e sotto la media.

Statistiche riassuntive basate sul conteggio

- **Conteggio.** Il numero di righe/casi.
- **Conteggio cumulato.** Il numero cumulato di righe/casi. Ogni elemento grafico mostra il conteggio riferito a un sottogruppo più il conteggio totale di tutti i gruppi precedenti.

- **Percentuale del conteggio.** La percentuale di righe/casi in ogni sottogruppo rispetto al numero totale di righe/casi.
- **Percentuale cumulata del conteggio.** La percentuale cumulata di righe/casi in ogni sottogruppo rispetto al numero totale di righe/casi. Ogni elemento grafico mostra la percentuale di un sottogruppo più la percentuale totale di tutti i gruppi precedenti.

Tipi di campo (variabile)

Negli elenchi di campi, accanto ai campi appaiono delle icone che indicano il tipo di campo e il tipo di dati. Le icone identificano anche gli insiemi di risposte multiple.

Livello di misurazione	Tipo di dati			
	Numeriche	Stringa	Data	Ora
Continuo		n/d		
Insieme ordinato				
Set				

Insieme di risposte multiple, categorie multiple	
Insieme di risposte multiple, dicotomie multiple	

Livello di misurazione

Il livello di misurazione di un campo è importante quando si crea una visualizzazione. Di seguito viene fornita una descrizione dei livelli di misurazione. Per cambiare temporaneamente un livello di misurazione, fare clic con il pulsante destro del mouse su un campo in un elenco di campi e scegliere un'opzione. Nella maggior parte dei casi, è necessario tenere in considerazione solo le due classificazioni più ampie di campi, vale a dire, categoriali e continui:

Categoriali. Dati con un numero limitato di categorie o valori distinti (per esempio, genere o religione). I campi categoriali possono essere stringhe (alfanumerici) o campi numerici che utilizzano codici numerici per rappresentare categorie (per esempio, 0 = *maschio* e 1 = *femmina*). Sono definiti anche dati qualitativi. Insiemi, insiemi ordinati e flag sono tutti campi categoriali.

-
-
-

Continua. Dati misurati su una scala di intervalli o rapporti, dove i valori dei dati indicano sia l'ordine dei valori che la distanza tra i valori. Per esempio, uno stipendio di € 72.195 è superiore a uno stipendio di € 52.398 e la distanza tra i due valori è € 19.797. Sono definiti anche dati quantitativi, di scala o di intervallo numerico.

I campi categoriali definiscono le categorie nella visualizzazione, tipicamente per creare elementi grafici distinti o per raggruppare gli elementi grafici. I campi continui sono spesso riepilogati all'interno di categorie di campi categoriali. Per esempio, in una visualizzazione predefinita relativa al reddito per le categorie Sesso verranno visualizzati il reddito medio dei maschi e il reddito medio delle femmine. I valori grezzi relativi ai campi continui possono anche essere rappresentati graficamente, come nel caso dei grafici a dispersione. Per esempio, un grafico a dispersione potrebbe mostrare lo stipendio attuale e lo stipendio iniziale per ogni caso. Un campo categoriale potrebbe essere usato per raggruppare i casi in base al sesso.

Tipi di dati

Il livello di misurazione non è l'unica proprietà che determina il tipo di un campo. Il campo viene anche memorizzato come tipo di dati specifico, per esempio, stringhe (dati non numerici, come le lettere), valori numerici (numeri reali) e date. A differenza del livello di misurazione, il tipo di dati di un campo non può essere temporaneamente modificato. È necessario modificare il modo in cui i dati sono memorizzati nell'insieme di dati originale.

Insiemi a risposta multipla

Alcuni file di dati supportano un tipo di “campo” speciale denominato **insieme a risposta multipla**. Gli insiemi di risposte multiple non sono in realtà “campi” in senso normale. Gli insiemi di risposte multiple utilizzano campi multipli per registrare le risposte alle domande per cui l'intervistato può fornire più di una risposta. Vengono trattati come campi categoriali e le operazioni supportate dai campi categoriali possono, per la maggior parte, essere eseguite anche con gli insiemi di risposte multiple.

Gli insiemi di risposte multiple possono essere insiemi dicotomici multipli o insiemi di categorie multiple.

Insiemi a dicotomie multiple. Un insieme dicotomico multiplo tipicamente è composto da più campi dicotomici, vale a dire, campi con due soli valori possibili, quali sì/no, presente/assente, selezionata/non selezionata. Anche se i campi potrebbero non essere tutti rigorosamente dicotomici, tutti i campi nell'insieme sono codificati allo stesso modo.

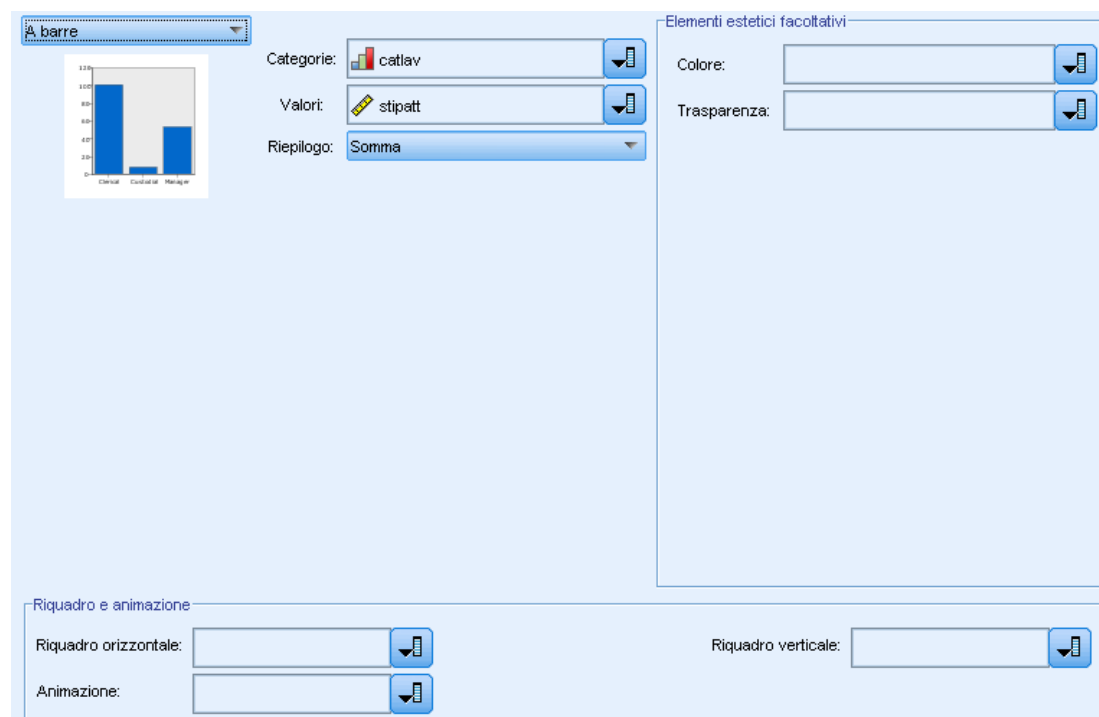
Per esempio, si immagina un sondaggio che offre cinque risposte possibili alla domanda, “Quale delle seguenti fonti di informazioni utilizzi per tenerti informato”? Il rispondente può indicare più risposte selezionando la casella accanto a ogni opzione. Le cinque risposte diventano cinque campi nel file di dati, codificate con 0 per *No* (non selezionata) e 1 per *Sì* (selezionata).

Insiemi di categorie multiple. Un insieme di categorie multiple è formato da più campi, tutti codificati allo stesso modo, spesso con molte categorie di risposta possibili. Per esempio, si supponga la seguente domanda in un sondaggio “Indica un massimo di tre nazionalità che meglio descrivono la tua appartenenza etnica”. Potrebbero esserci centinaia di risposte possibili ma, per esigenze di codificazione, l'elenco è limitato alle 40 nazionalità più comuni, mentre tutte le altre

possibilità sono incluse nella categoria “altro”. Nel file di dati, le tre scelte diventeranno tre campi, ognuno con 41 categorie (40 nazionalità codificate e una categoria “altro”).

Lavagna grafica Scheda Dettagliato

Figura 5-10
Scheda Dettagliato



Utilizzare la scheda Dettagliato quando si conosce il tipo di visualizzazione da creare o quando si desidera aggiungere elementi estetici facoltativi, pannelli e/o effetti di animazione a una visualizzazione.

- ▶ Verrà visualizzato il tipo di visualizzazione scelto nella scheda Di base, se ne è stato selezionato uno; in caso contrario, sceglierne uno dall’elenco a discesa.
- ▶ Subito a destra dell’anteprima della visualizzazione si trovano i comandi per specificare i campi (le variabili) necessari per il tipo di visualizzazione. È necessario specificare tutti questi campi.
- ▶ Per alcune visualizzazioni è possibile selezionare una statistica riassuntiva. In taluni casi, per esempio nei grafici a barre, è possibile utilizzare una di queste opzioni riassuntive per l’effetto estetico della trasparenza. Per le descrizioni delle statistiche riassuntive, vedere [Scheda Di base del nodo Lavagna grafica](#) a pag. 240.
- ▶ È possibile selezionare uno o più degli elementi estetici facoltativi. Essi aggiungono dimensionalità consentendo l’inclusione di altri campi nella visualizzazione. Per esempio, è possibile utilizzare un campo per variare la dimensione dei punti in un grafico a dispersione. Per ulteriori informazioni sugli elementi estetici facoltativi, vedere [Effetti estetici, sovrapposizioni, riquadri e animazioni](#) a pag. 232. Si noti che l’effetto estetico della trasparenza non è supportato tramite gli script.

- ▶ Se si sta creando una visualizzazione delle mappe, il gruppo File mappa mostra il file o i file delle mappe che saranno utilizzati. Viene visualizzato l'eventuale file delle mappe di default. Per modificare il file delle mappe, fare clic su Seleziona un file mappa per visualizzare la finestra di dialogo Seleziona mappe. In questa finestra di dialogo è possibile anche specificare il file delle mappe di default. [Per ulteriori informazioni, vedere l'argomento Selezione dei file delle mappe per le visualizzazioni delle mappe a pag. 246.](#)
- ▶ È possibile selezionare una o più opzioni relative ai pannelli o all'animazione. Per ulteriori informazioni sulle opzioni relative ai pannelli e all'animazione, vedere [Effetti estetici, sovrapposizioni, riquadri e animazioni](#) a pag. 232.

Selezione dei file delle mappe per le visualizzazioni delle mappe

Se si seleziona un modello di visualizzazione delle mappe, è necessario un file delle mappe che definisca le informazioni geografiche per disegnare la mappa. Se è presente un file delle mappe di default, viene utilizzato per la visualizzazione delle mappe. Per scegliere un file delle mappe differente, fare clic su Seleziona un file mappa nella scheda Dettagliato per visualizzare la finestra di dialogo Seleziona mappe.

La finestra di dialogo Seleziona mappe consente di selezionare un file delle mappe principale e un file delle mappe di riferimento. I file delle mappe definiscono le informazioni geografiche per disegnare le mappe. L'applicazione è installata con diversi file delle mappe standard. Se si dispone di altri shapefile ESRI che si desidera utilizzare, è necessario convertire prima gli shapefile in file SMZ. Dopo aver convertito la mappa, fare clic su Gestisci... nella finestra di dialogo Selezionatore modelli per importare la mappa nel sistema di gestione in modo che sia disponibile nella finestra di dialogo Seleziona mappe.

Si riportano di seguito alcuni punti da considerare quando si specificano file delle mappe:

- Tutti i modelli delle mappe richiedono almeno un file delle mappe.
- Il file delle mappe collega generalmente un attributo chiave della mappa a una chiave di dati.
- Se il modello non necessita di una chiave della mappa che si colleghi a una chiave dei dati, necessita di un file e dei campi delle mappe di riferimento che specifichino le coordinate (ad esempio longitudine e latitudine) per disegnare gli elementi nella mappa di riferimento.
- I modelli di mappe di sovrapposizione necessitano di due mappe: un file delle mappe principale e un file delle mappe di riferimento. Il file delle mappe di riferimento viene disegnato per primo in modo che si trovi dietro al file delle mappe principale.

File mappa. È possibile selezionare qualunque file delle mappe presente nel sistema di gestione. I file comprendono i file delle mappe preinstallati che sono stati importati. Per ulteriori informazioni sulla gestione dei file delle mappe, vedere [Gestione dei modelli, dei fogli di stile e dei file delle mappe](#) a pag. 251.

Chiave mappa. Specificare l'attributo che si vuole utilizzare come chiave che collega il file delle mappe alla chiave dei dati.





Salva questo file mappa e le impostazioni come valori predefiniti. Selezionare questa casella di controllo se si vuole utilizzare il file delle mappe selezionato come default. Se si è specificato un file delle mappe di default, non è necessario specificare un file delle mappe ogni volta che si crea una visualizzazione delle mappe.

Chiave dati. Questo comando elenca gli stessi valori che appaiono nella scheda Dettagliato del Selezionatore modelli. È presentato qui per comodità se si desidera cambiare la chiave in funzione dello specifico file delle mappe che si è scelto.

Visualizza tutte le caratteristiche della mappa nella visualizzazione. Quando questa opzione è selezionata, si esegue il rendering di tutte le funzioni della mappa nella visualizzazione, anche se non vi sono valori corrispondenti nella chiave di dati. Se si desidera vedere solo le caratteristiche per le quali si dispone di dati, deselezionare questa opzione.

Confronta la mappa e i valori dei dati. La chiave della mappa e la chiave dei dati si collegano tra loro per creare una visualizzazione della mappa. I valori di queste due chiavi devono corrispondere, in caso contrario non è possibile creare una visualizzazione della mappa. Fare clic su Confronta per verificare se i valori della chiave dei dati e della chiave della mappa corrispondono. L'icona visualizzata informa sullo stato del confronto. Le icone sono descritte di seguito. Se è stato effettuato un confronto e vi sono chiavi dei dati senza corrispondenti chiavi della mappa, i valori delle chiavi dei dati sono mostrati nell'elenco Chiavi non corrispondenti.

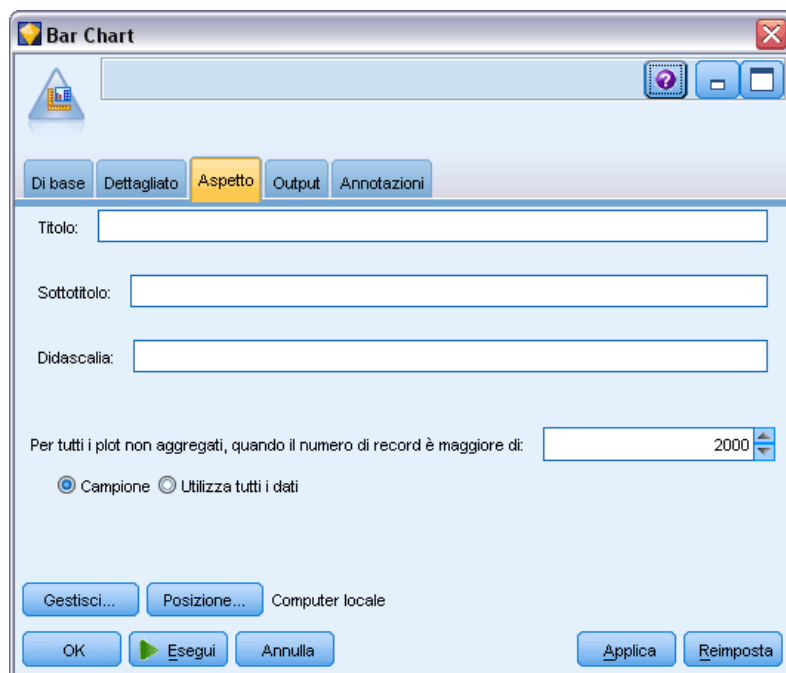
Tabella 5-1
Icane di confronto

Icone	Descrizione
	Non è stato eseguito alcun confronto. Questo è lo stato di default prima di fare clic su Confronta. Si consiglia di procedere con cautela perché non è noto se i valori delle chiavi dei dati e delle chiavi delle mappe corrispondono.
	È stato effettuato un confronto, e i valori della chiave dei dati e della chiave della mappa corrispondono. Per ogni valore della chiave dei dati è stata individuata una caratteristica corrispondente nella chiave della mappa.
	È stato effettuato un confronto, e alcuni valori della chiave dei dati e della chiave della mappa non corrispondono. Per alcuni valori della chiave dei dati non è stata individuata una caratteristica corrispondente nella chiave della mappa. Procedere con attenzione. Se si prosegue, la visualizzazione della mappa non conterrà tutti i valori dei dati.
	È stato effettuato un confronto, e nessuno dei valori della chiave dei dati e della chiave della mappa corrisponde. Si consiglia di selezionare una chiave dei dati o una chiave della mappa diversa perché se si prosegue non sarà eseguito il rendering di alcuna mappa.

Scheda Aspetto del nodo Lavagna grafica

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-12
Impostazioni della scheda Aspetto per un nodo Lavagna grafica



Opzioni generali della scheda Aspetto

Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Campionamento. Specificare un metodo per rappresentare insiemi di dati di notevoli dimensioni. È possibile specificare una dimensione massima per l'insieme di dati oppure utilizzare il numero di record di default. Per gli insiemi di dati di grandi dimensioni le prestazioni risultano migliori se si seleziona l'opzione Campione. In alternativa, è possibile rappresentare tutti i punti dei dati selezionando Utilizza tutti i dati, tuttavia occorre tenere presente che in questo modo le prestazioni del software potrebbero risultare notevolmente ridotte.

Opzioni dei fogli di stile della scheda Aspetto

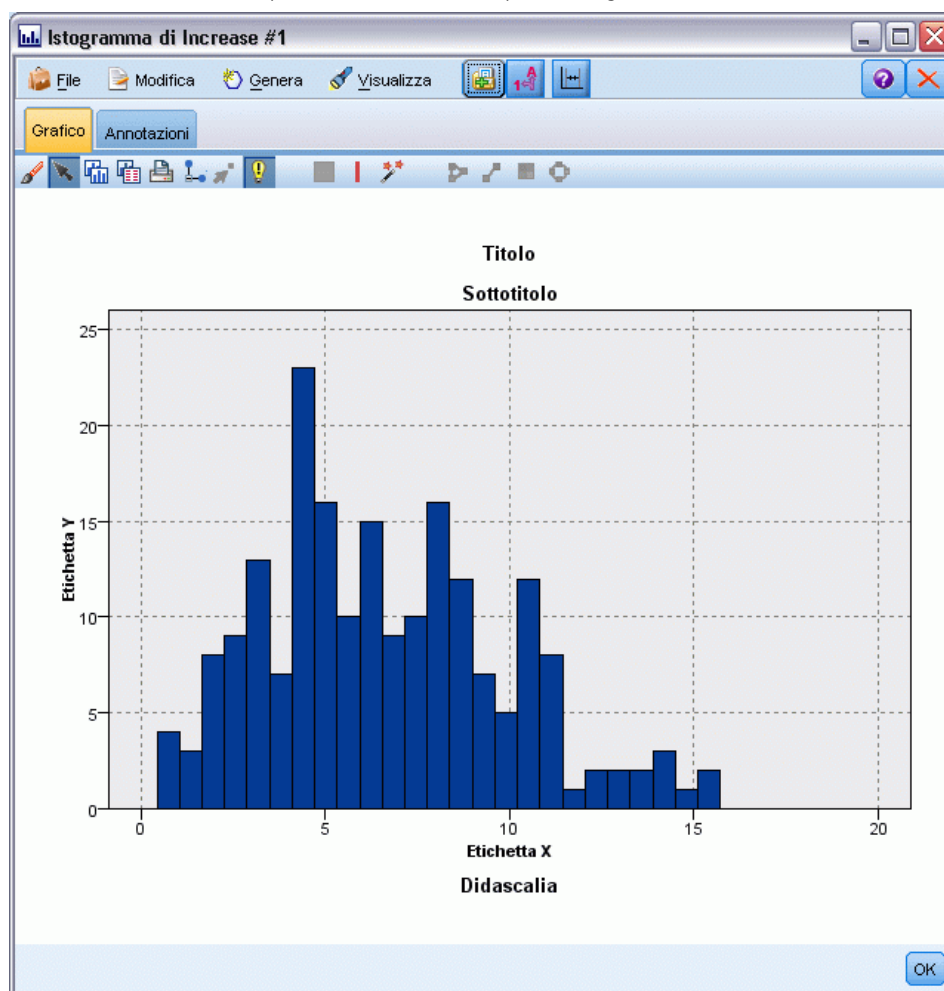
Due pulsanti consentono di controllare quali modelli (e fogli di stile) di visualizzazione sono disponibili:

Gestisci. Gestire modelli e i fogli di stile di visualizzazione sul computer. È possibile importare, esportare, rinominare ed eliminare modelli e i fogli di stile di visualizzazione sul computer locale. [Per ulteriori informazioni, vedere l'argomento Gestione dei modelli, dei fogli di stile e dei file delle mappe a pag. 251.](#)

Posizione. Modificare la posizione in cui sono archiviati i modelli e i fogli di stile di visualizzazione. La posizione corrente viene elencata alla destra del pulsante. [Per ulteriori informazioni, vedere l'argomento Impostazione della posizione di modelli, fogli di stile e mappe a pag. 249.](#)

Il seguente esempio mostra dove vengono posizionate le opzioni della scheda Aspetto su un grafico. (Nota: non tutti i grafici utilizzano tutte queste opzioni.)

Figura 5-13
Posizione delle diverse opzioni della scheda Aspetto sui grafici



Impostazione della posizione di modelli, fogli di stile e mappe

I modelli di visualizzazione, i fogli di stile di visualizzazione e i file delle mappe sono archiviati in una cartella locale specifica o nel IBM® SPSS® Collaboration and Deployment Services Repository. Quando si selezionano i modelli, i fogli di stile e le mappe, sono visualizzati solo quelli integrati in questo percorso. Il fatto di archiviare tutti i modelli, i fogli di stile e i file delle mappe nello stesso punto, consente alle applicazioni di IBM Corp. di potervi accedere più

facilmente. Per informazioni sull'aggiunta di modelli, fogli di stile e file delle mappe in questa posizione, vedere [Gestione dei modelli, dei fogli di stile e dei file delle mappe](#) a pag. 251.

Come impostare la posizione di modelli, fogli di stile e file delle mappe

- ▶ Nella finestra di dialogo di un modello o foglio di stile, fare clic su Posizione... per visualizzare la finestra di dialogo Modelli, fogli di stile e mappe.
- ▶ Selezionare un'opzione per impostare il percorso predefinito di modelli, fogli di stile e file delle mappe:

Computer locale. I modelli, i fogli di stile e i file delle mappe si trovano in una cartella specifica del computer locale. In Windows XP, si tratta della cartella *C:\Documents and Settings\<utente>\Application Data\SPSSInc\Graphboard*. Non è possibile modificare questa cartella.

IBM® SPSS® Collaboration and Deployment Services Repository. I modelli, i fogli di stile e i file delle mappe si trovano in una cartella specificata dall'utente nel IBM SPSS Collaboration and Deployment Services Repository. Per individuare la cartella specifica, fare clic su Cartella. Per maggiori informazioni, vedere [Utilizzo di IBM SPSS Collaboration and Deployment Services Repository come posizione per modelli, fogli di stile e file delle mappe](#) a pag. 250.

- ▶ Fare clic su OK.

Utilizzo di IBM SPSS Collaboration and Deployment Services Repository come posizione per modelli, fogli di stile e file delle mappe

I modelli e fogli di stile di visualizzazione possono essere memorizzati in IBM® SPSS® Collaboration and Deployment Services Repository. Si tratta di una cartella specifica in IBM SPSS Collaboration and Deployment Services Repository. Se tale percorso è impostato come posizione di default, è possibile selezionare tutti i modelli, i fogli di stile e i file delle mappe presenti al suo interno.

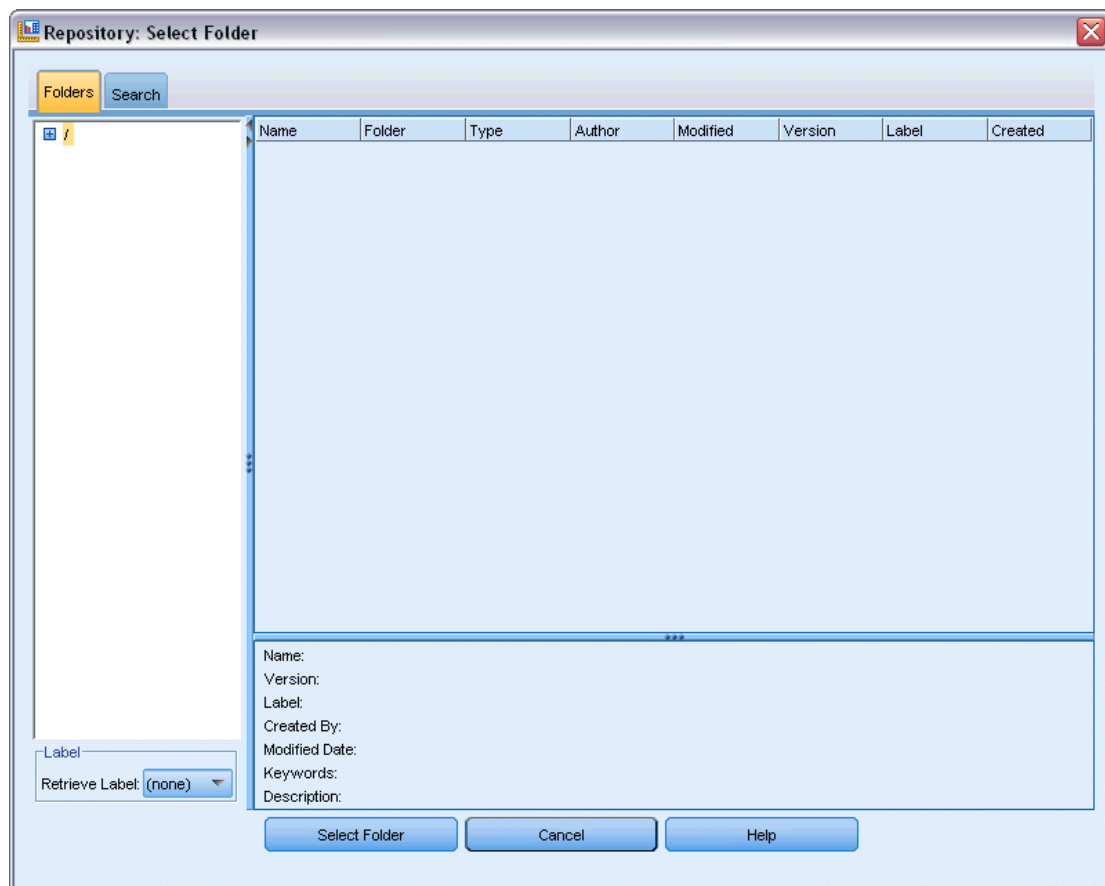
Come impostare una cartella in IBM SPSS Collaboration and Deployment Services Repository come percorso per i modelli, i fogli di stile e i file delle mappe

- ▶ Nelle finestre di dialogo con il pulsante Posizione, fare clic su Posizione....
- ▶ Selezionare IBM® SPSS® Collaboration and Deployment Services Repository.
- ▶ Fare clic su Cartella.

Nota: se non è ancora stata stabilita una connessione al IBM SPSS Collaboration and Deployment Services Repository, verranno richieste le informazioni per la connessione.

- ▶ Nella finestra di dialogo Seleziona cartella, selezionare la cartella in cui sono memorizzati i modelli, i fogli di stile e i file delle mappe.

Figura 5-14
Finestra di dialogo *Seleziona cartella*



- ▶ Se lo si desidera, selezionare un'etichetta da *Recupera etichetta*. Verranno visualizzati solo i modelli, i fogli di stile e i file delle mappe con l'etichetta selezionata.
- ▶ Se si cerca una cartella contenente un determinato modello, foglio di stile o file delle mappe, è possibile effettuare una ricerca di questi elementi nella scheda *Cerca*. La finestra di dialogo *Seleziona cartella* consente di selezionare automaticamente la cartella del modello, del foglio di stile o del file delle mappe trovato.
- ▶ Fare clic su *Seleziona cartella*.

Gestione dei modelli, dei fogli di stile e dei file delle mappe

I modelli, i fogli di stile e i file delle mappe possono essere gestiti dalla posizione locale del computer mediante la finestra di dialogo *Gestisci modelli, fogli di stile e mappe*. Questa finestra di dialogo consente di importare, esportare, rinominare ed eliminare i modelli, i fogli di stile di visualizzazione e i file delle mappe che si trovano nella posizione locale del computer.

- ▶ Fare clic su *Gestisci...* in una delle finestre di dialogo in cui si selezionano i modelli, i fogli di stile o le mappe.

Finestra di dialogo Gestisci modelli, fogli di stile e mappe

Nella scheda Modello sono elencati tutti i modelli locali. Nella scheda Foglio di stile sono elencati tutti i fogli di stile locali e le visualizzazioni di esempio contenenti dei dati campione. È possibile selezionare uno dei fogli di stile e applicare i relativi stili alle visualizzazioni di esempio. [Per ulteriori informazioni, vedere l'argomento Applicazione di fogli di stile a pag. 342.](#) La scheda Mappa elenca tutti i file delle mappe locali. Questa scheda visualizza anche le chiavi delle mappe compresi valori campione, un commento se ne è stato specificato uno durante la creazione della mappa e un'anteprima della mappa.

I pulsanti seguenti funzionano in qualsiasi scheda attiva in un determinato momento.

Importa. Consente di importare un modello, un foglio di stile di visualizzazione o un file mappa dal file system. Dopo l'importazione, un modello, un foglio di stile o un file delle mappe sono disponibili nell'applicazione IBM Corp.. Se si riceve un modello, un foglio di stile o un file delle mappe da un altro utente, è necessario importarlo prima di poterlo usare nella propria applicazione.

Esporta. Consente di esportare un modello, un foglio di stile di visualizzazione o un file mappa nel file system. Se si desidera inviare un modello, foglio di stile o un file delle mappe a un altro utente, è necessario esportarlo.

Rinomina. Consente di modificare il nome di un modello, foglio di stile di visualizzazione o file delle mappe. Non è consentito usare un nome già esistente.

Elimina. Consente di eliminare i modelli, i fogli di stile di visualizzazione o i file delle mappe selezionati. È possibile selezionare più modelli, fogli di stile o file delle mappe con Ctrl-clic. L'eliminazione non può essere annullata, pertanto si consiglia di agire con cautela.

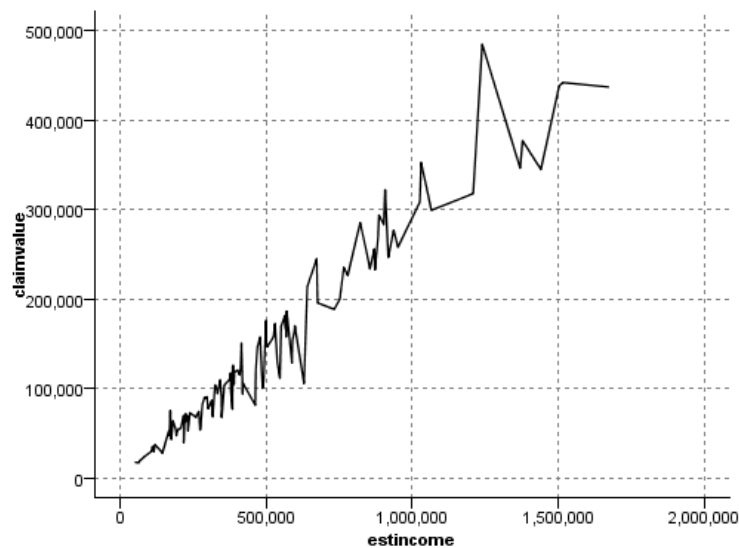
Nodo Plot

I nodi Plot mostrano la relazione tra campi numerici. Per creare un grafico plot è possibile utilizzare punti (in modo da ottenere un grafico a dispersione) o linee. L'impostazione della Modalità X nella finestra di dialogo consente di creare tre tipi di grafici plot lineari.

Modalità X = Ordina

Se la Modalità X viene impostata su Ordina, i dati vengono ordinati in base ai valori per il campo rappresentati sull'asse x . Questa impostazione produce un'unica linea che attraversa il grafico da sinistra a destra. L'utilizzo di un campo nominale come sovrapposizione produce più linee di diverse tonalità che attraversano il grafico da sinistra a destra.

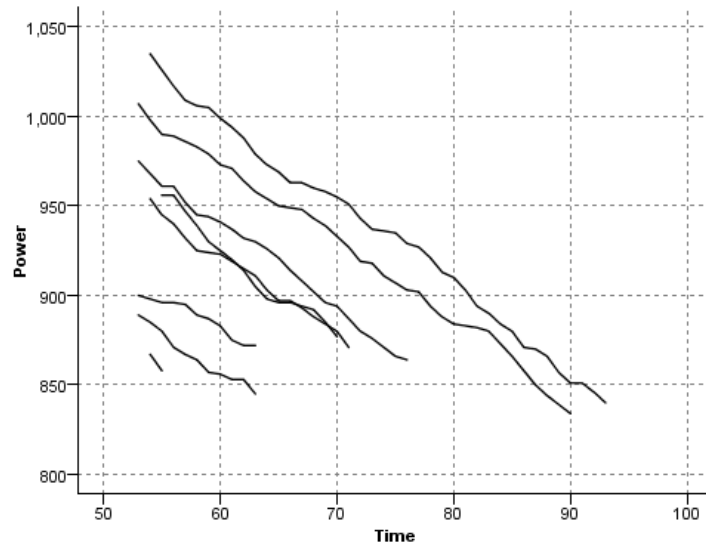
Figura 5-15
Grafico plot lineare con la Modalità X impostata su Ordina



Modalità X = Sovrapposizione

L'impostazione della Modalità X su Sovrapposizione crea più grafici lineari nello stesso grafico. I dati di un grafico plot di sovrapposizione non vengono ordinati. Se i valori sull'asse x aumentano, i dati verranno rappresentati su un'unica linea. Se i valori diminuiscono, viene aggiunta una nuova linea. Per esempio, quando x aumenta da 0 a 100 i valori y vengono rappresentati su un'unica linea. Quando x diminuisce fino a un valore inferiore a 100, nel grafico viene rappresentata una nuova linea in aggiunta alla linea esistente. La versione definitiva del grafico plot può includere diversi grafici utili per confrontare più serie di valori y . Questo tipo di grafico può essere utilizzato per dati con una componente temporale periodica, per esempio il consumo di energia elettrica in periodi successivi di 24 ore.

Figura 5-16
Grafico plot lineare con la Modalità X impostata su Sovrapposizione

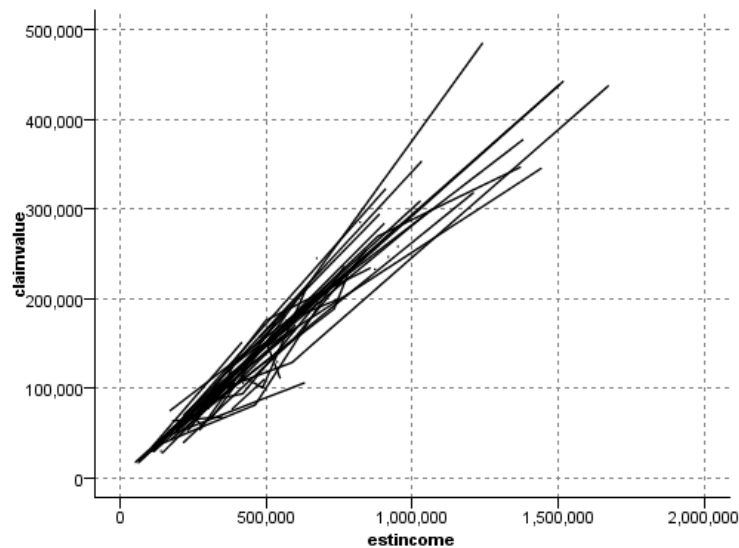


Modalità X = Ordine di lettura

L'impostazione della Modalità X su Ordine di lettura consente di rappresentare i valori x e y così come vengono letti dalla sorgente dati. Questa opzione è utile per dati con una componente di serie storiche utilizzati per analizzare le tendenze o gli schemi che dipendono dall'ordine dei dati. Potrebbe essere necessario ordinare i dati prima di creare questo tipo di grafico. Inoltre, potrebbe essere utile confrontare due grafici plot simili con la Modalità X impostata su Ordina e su Ordine di lettura, in modo da stabilire fino a che punto uno schema dipenda dall'ordinamento.

Figura 5-17

Grafico plot lineare mostrato in precedenza con l'impostazione Ordina, eseguito nuovamente con la Modalità X impostata su Ordine di lettura

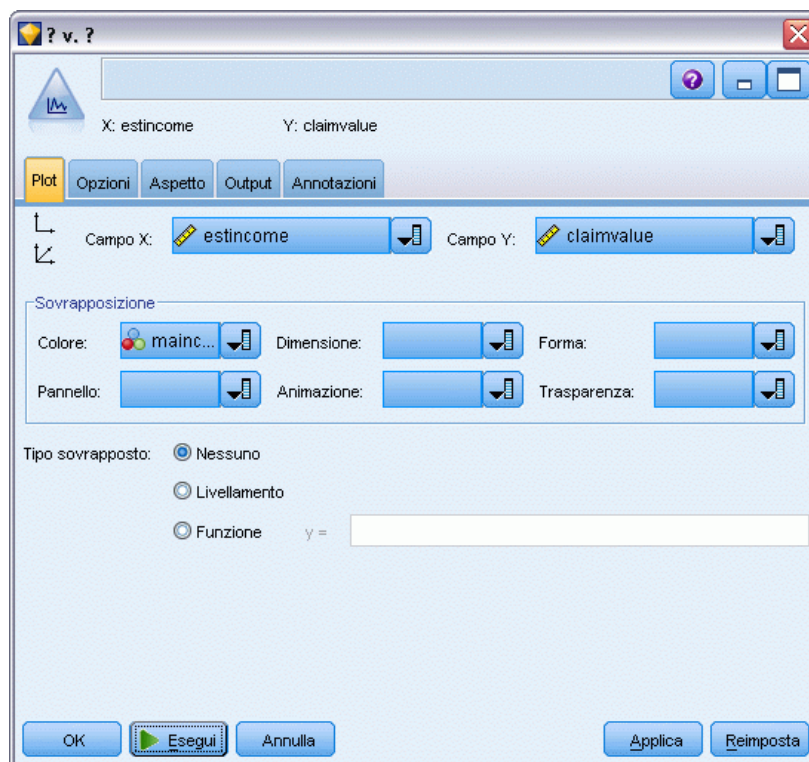


Per generare grafici a dispersione e grafici plot lineari è anche possibile utilizzare il nodo Lavagna grafica. Tuttavia, il nodo qui illustrato dispone di un numero maggiore di opzioni da cui scegliere.

Scheda Nodo Plot

I nodi Plot mostrano il confronto tra i valori di un campo Y e i valori di un campo X . Spesso questi campi corrispondono rispettivamente a una variabile dipendente e a una variabile indipendente.

Figura 5-18
Impostazioni della scheda Plot per un nodo Plot



Campo X. Selezionare dall’elenco il campo da visualizzare lungo l’asse x orizzontale.

Campo Y. Selezionare dall’elenco il campo da visualizzare lungo l’asse y verticale.

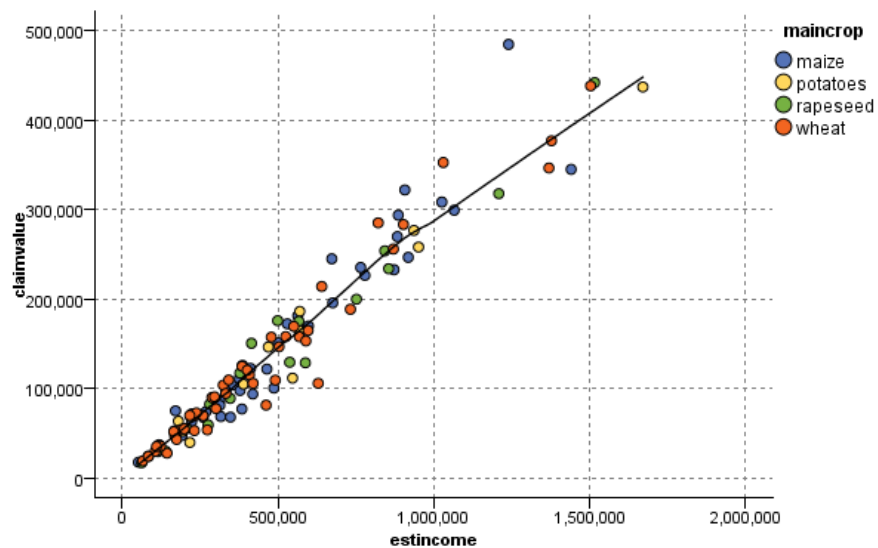
Campo Z. Quando si fa clic sul pulsante del grafico 3-D è possibile scegliere dall’elenco un campo da visualizzare sull’asse z .

Sovrapposizione. Esistono diverse modalità di rappresentazione delle categorie per i valori dei dati. Ad esempio, è possibile utilizzare *raccolto_principale* come sovrapposizione di colore per indicare i valori *stima_reddito* e *valore_domanda* per il raccolto principale coltivato da coloro che hanno presentato domanda di indennizzo. [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Tipo sovrapposto. Specifica se viene visualizzata una funzione di sovrapposizione o di livellamento. Le funzioni di livellamento e sovrapposizione sono sempre calcolate come funzione di y .

- **Nessuna.** Non vengono visualizzate sovrapposizioni.
- **Livellamento.** Visualizza una curva di interpolazione calcolata con il metodo LOESS (locally-weighted iterative robust least squares regression, regressione iterativa di minimi quadrati ponderati localmente). Questo metodo calcola in modo efficiente una serie di regressioni, ciascuna concentrata su una piccola area del plot. In questo modo viene generata una serie di linee di regressione “locali” che vengono in seguito unite in modo da creare una curva.

Figura 5-19
Plot con sovrapposizione del livellamento LOESS



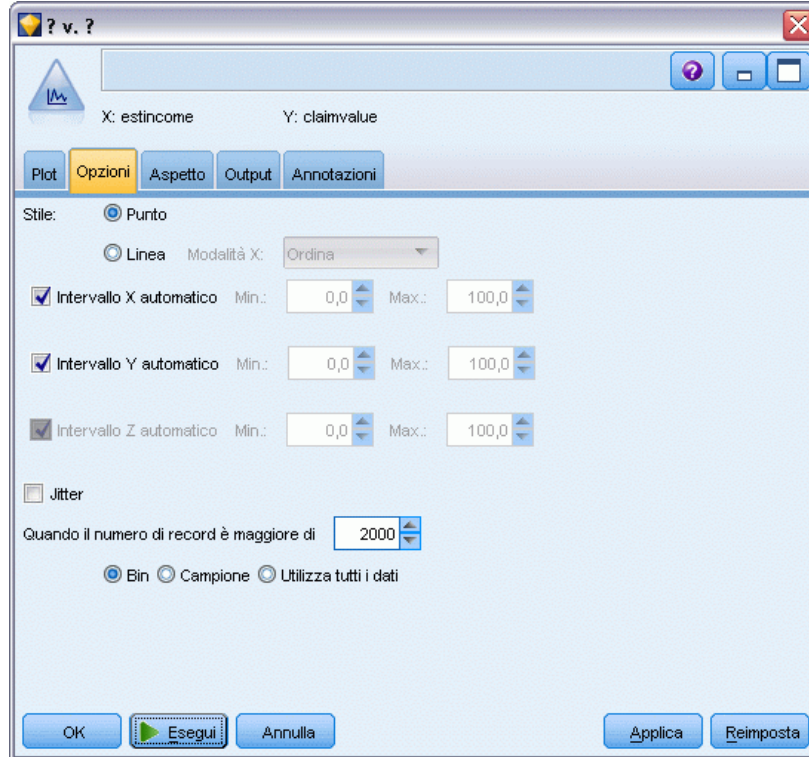
- **Funzione.** Selezionare questa opzione per specificare una funzione conosciuta da confrontare con i valori effettivi. Ad esempio, per confrontare valori effettivi con valori previsti, è possibile rappresentare la funzione $y = x$ come una sovrapposizione. Specificare una funzione nella casella di testo $y =$. La funzione di default è $y = x$, tuttavia è possibile specificare una funzione qualsiasi, ad esempio una funzione quadratica o un'espressione arbitraria, al posto di x .

Nota: le funzioni sovrapposte non sono disponibili per un grafico contenente pannelli o animazione.

Dopo aver impostato le opzioni per un grafico plot, è possibile eseguire il grafico direttamente dalla finestra di dialogo facendo clic su Esegui. Tuttavia, è possibile che si desideri utilizzare la scheda Opzioni per impostare specifiche aggiuntive, ad esempio discretizzazione, Modalità X e stile.

Scheda Opzioni del nodo Plot

Figura 5-20
Impostazioni della scheda Opzioni per un nodo Plot



Stile. Selezionare Punto o Linea per lo stile del grafico. La selezione di Linea attiva il controllo Modalità X. Se si seleziona Punto, come forma di default del punto viene utilizzato il segno più (+). Dopo aver creato il grafico, è possibile modificare la forma del punto e la dimensione.

Modalità X. Per i grafici lineari è necessario selezionare un'opzione della Modalità X in modo da definire lo stile del grafico. Selezionare Ordina, Sovrapposizione o Ordine di lettura. Per Sovrapposizione o Ordine di lettura è necessario specificare la dimensione massima dell'insieme di dati utilizzata per campionare i primi n record, altrimenti verrà utilizzata l'impostazione di default di 2.000 record.

Intervallo X automatico. Selezionare questa opzione per utilizzare l'intero intervallo di valori nei dati lungo l'asse. Deselezionare questa opzione per utilizzare un sottoinsieme esplicito di valori compresi tra i valori Min e Max specificati. Digitare i valori direttamente oppure utilizzare le frecce. Per default, gli intervalli vengono selezionati automaticamente in modo da consentire una più rapida creazione dei grafici.

Intervallo Y automatico. Selezionare questa opzione per utilizzare l'intero intervallo di valori nei dati lungo l'asse. Deselezionare questa opzione per utilizzare un sottoinsieme esplicito di valori compresi tra i valori Min e Max specificati. Digitare i valori direttamente oppure utilizzare le frecce. Per default, gli intervalli vengono selezionati automaticamente in modo da consentire una più rapida creazione dei grafici.

Intervallo Z automatico. Solo quando nella scheda Plot è selezionato un grafico 3-D. Selezionare questa opzione per utilizzare l'intero intervallo di valori nei dati lungo l'asse. Deselezionare questa opzione per utilizzare un sottoinsieme esplicito di valori compresi tra i valori Min e Max specificati. Digitare i valori direttamente oppure utilizzare le frecce. Per default, gli intervalli vengono selezionati automaticamente in modo da consentire una più rapida creazione dei grafici.

Jitter. Anche noto come **agitazione**, è utile per grafici a punti di un insieme di dati in cui molti valori sono ripetuti. Per visualizzare più chiaramente la distribuzione dei valori, è possibile utilizzare l'opzione jitter per distribuire i punti in modo casuale attorno al valore effettivo.

Nota per gli utenti delle precedenti versioni di SPSS Modeler: il valore di jitter usato in un grafico plot utilizza un tipo di misurazione diversa in questa versione di IBM® SPSS® Modeler. Nelle versioni precedenti il valore era un numero effettivo, ora è una percentuale della dimensione della cornice. Ciò significa che i valori di agitazione degli stream delle versioni precedenti risulteranno probabilmente maggiori rispetto a quelli di questa versione. In questa versione, qualsiasi valore di agitazione diverso da zero verrà convertito nel valore 0,2.

Numero massimo di record per il plot. Specifica un metodo per rappresentare insiemi di dati di notevoli dimensioni. È possibile specificare una dimensione massima per l'insieme di dati oppure utilizzare l'impostazione di default di 2.000 record. Per gli insiemi di dati di grandi dimensioni le performance risultano migliori se si seleziona l'opzione Bin o Campione. In alternativa, è possibile rappresentare tutti i punti dei dati selezionando Utilizza tutti i dati, tuttavia occorre tenere presente che in questo modo le prestazioni del software potrebbero risultare notevolmente ridotte.

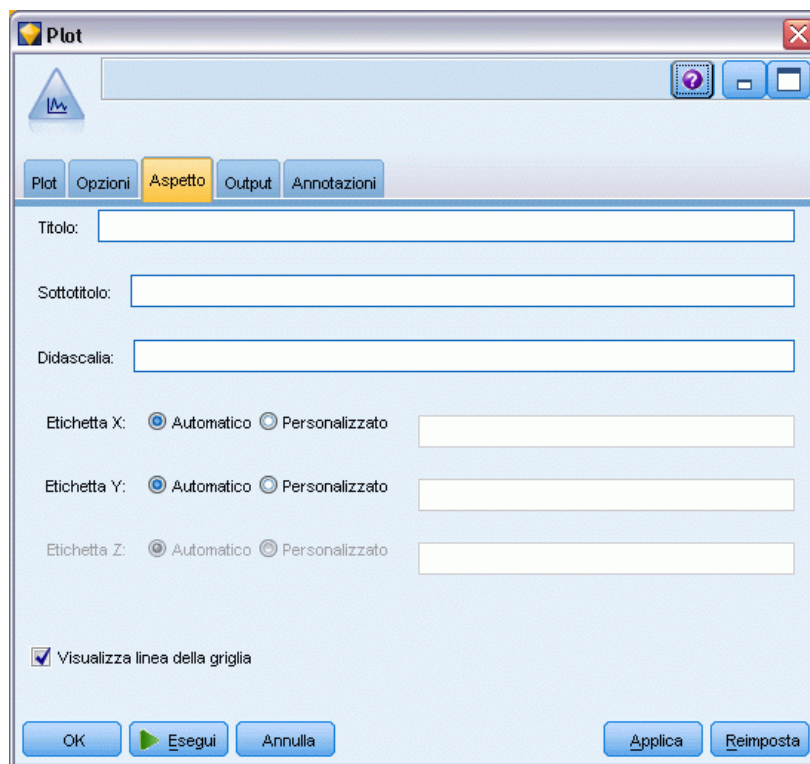
Nota: quando la Modalità X viene impostata su Sovrapposizione o su Ordine di lettura, queste opzioni non sono disponibili e vengono utilizzati solo i primi n record.

- **Bin.** Selezionare questa opzione per attivare gli intervalli quando l'insieme di dati contiene un numero di record maggiore di quello specificato. L'attivazione degli intervalli determina la divisione del grafico in griglie fini prima della rappresentazione effettiva e il conteggio del numero di punti che verrebbero visualizzati in ogni cella della griglia. Nel grafico finale viene rappresentato un unico punto per cella sul baricentro del bin (media di tutte le posizioni dei punti nel bin). La dimensione dei simboli rappresentati indica il numero di punti in quell'area (a meno che la dimensione non sia stata utilizzata come sovrapposizione). Grazie all'utilizzo del baricentro e della dimensione per la rappresentazione del numero di punti, il grafico a intervalli offre una modalità di rappresentazione eccellente per insiemi di dati di grandi dimensioni, in quanto impedisce la sovrarappresentazione in aree dense (masse di colore non differenziate) e riduce il rischio di rappresentazioni non valide (a causa di schemi di densità falsati). Le rappresentazioni non valide si verificano quando alcuni simboli (in particolare il segno più [+]) collidono in modo da produrre aree dense che non sono presenti nei dati grezzi.
- **Esempio.** Selezionare questa opzione per eseguire un campionamento casuale dei dati sul numero di record immesso nel campo di testo. Il valore di default è 2,000.

Scheda Aspetto del nodo Plot

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-21
Impostazioni della scheda Aspetto per un nodo Plot



Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Z. Disponibile solo per i grafici 3-D. Accettare l'etichetta dell'asse z generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

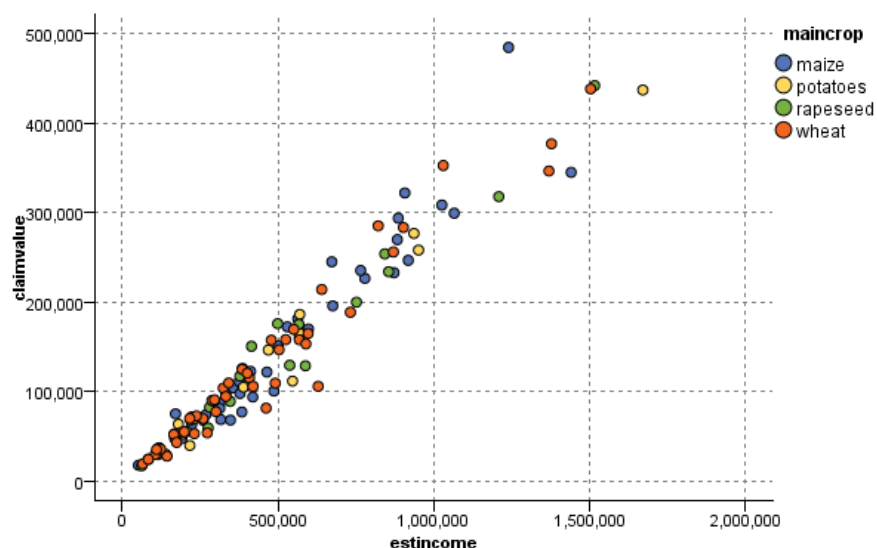
Utilizzo del grafico di un nodo Plot

I nodi Plot e Plot multiplo creano essenzialmente grafici in cui X viene rappresentato rispetto a Y . Per esempio, se si sta analizzando una potenziale frode nelle richieste di concessioni agricole (come illustrato in *fraud.str* nella cartella *Demos* dell'installazione di IBM® SPSS® Modeler), si

potrebbe voler rappresentare il reddito dichiarato nella domanda a fronte di quello stimato da una rete neurale. Una sovrapposizione, ad esempio il tipo di raccolto, consentirà di vedere se esiste una relazione tra domande (valore o numero) e tipo di raccolto.

Figura 5-22

Grafico della relazione tra reddito stimato e valore della domanda con il tipo raccolto principale come sovrapposizione



Dal momento che i grafici prodotti dai nodi Plot, Plot multiplo e Valutazione sono visualizzazioni bidimensionali di Y rispetto a X , è facile interagire con essi selezionando le aree, contrassegnando elementi o persino disegnando sezioni mediante il mouse. È inoltre possibile generare nodi per i dati rappresentati da queste aree, sezioni o elementi. [Per ulteriori informazioni, vedere l'argomento Esplorazione dei grafici a pag. 309.](#)

Nodo Distribuzione

Un grafico o tabella di distribuzione mostra l'occorrenza di valori simbolici (non numerici), per esempio tipo di ipoteca o sesso, in un insieme di dati. Un nodo Distribuzione viene in genere utilizzato per mostrare gli sbilanciamenti nei dati che possono essere corretti se si aggiunge un nodo Bilanciamento prima di creare un modello. Per generare automaticamente un nodo Bilanciamento, utilizzare il menu Genera nella finestra di un grafico o tabella di distribuzione.

Per generare grafici con barre dei conteggi è anche possibile utilizzare il nodo Lavagna grafica. Tuttavia, il nodo qui illustrato dispone di un numero maggiore di opzioni da cui scegliere.

Nota: per visualizzare l'occorrenza di valori numerici, è necessario utilizzare un nodo Istogramma.

Scheda Plot del nodo Distribuzione

Figura 5-23
Impostazioni della scheda Plot per un nodo Distribuzione

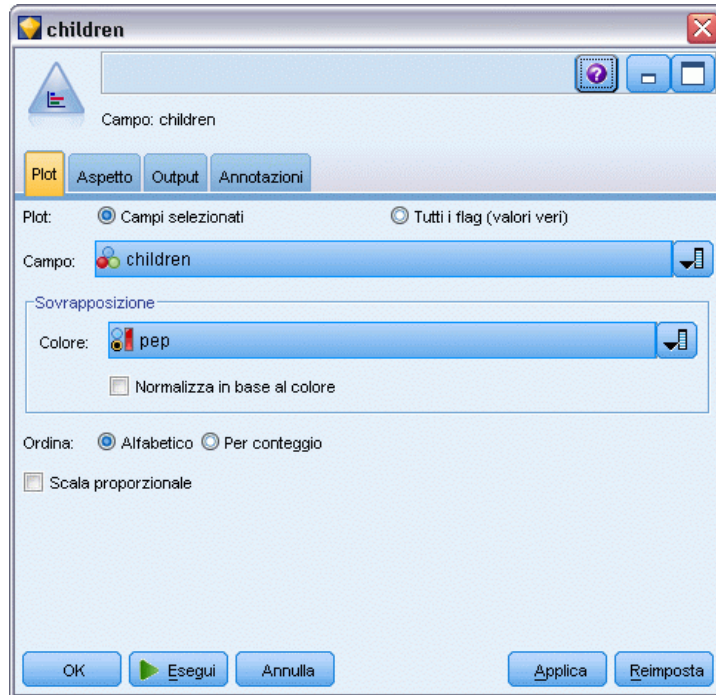


Grafico. Selezionare il tipo di distribuzione. Selezionare Campi selezionati per visualizzare la distribuzione del campo selezionato. Selezionare Tutti i flag (valori veri) per visualizzare la distribuzione dei valori veri per i campi flag nell'insieme di dati.

Campo. Selezionare un campo nominale o flag per il quale visualizzare la distribuzione dei valori. Nell'elenco sono inclusi solo i campi non esplicitamente impostati come numerici.

Sovrapposizione. Selezionare un campo nominale o flag da utilizzare come sovrapposizione di colore che rappresenti la distribuzione dei valori del campo all'interno di ciascun valore del campo specificato. Ad esempio, è possibile utilizzare la risposta a una campagna di marketing (*pep*) come sovrapposizione per numero di figli (*figli*) per rappresentare la risposta in base alla dimensione del nucleo familiare. [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Normalizza in base al colore. Selezionare questa opzione per graduare le barre, in modo che tutte occupino l'intera larghezza del grafico. I valori di sovrapposizione corrispondono a una proporzione di ogni barra, pertanto il confronto tra le categorie risulta più semplice.

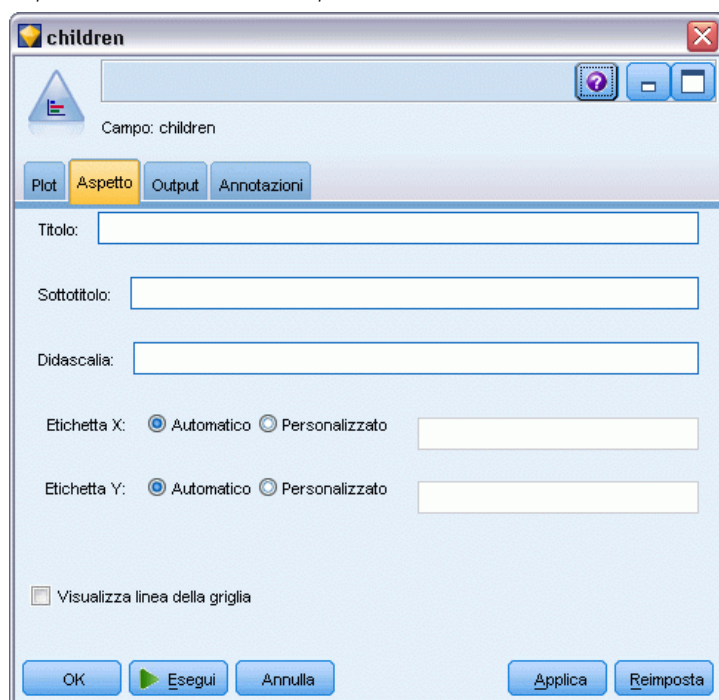
Ordina. Selezionare il metodo utilizzato per visualizzare i valori nel grafico di distribuzione. Selezionare Alfabetico per utilizzare l'ordine alfabetico, oppure Per conteggio per elencare i valori in ordine di occorrenza decrescente.

Scala proporzionale. Selezionare questa opzione per graduare la distribuzione dei valori in modo che il valore con il conteggio maggiore occupi l'intera larghezza del grafico. Tutte le altre barre vengono graduate rispetto a questo valore. Se questa opzione è deselezionata, le barre vengono graduate in base al conteggio totale per ogni valore.

Scheda Aspetto del nodo Distribuzione

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-24
Impostazioni della scheda Aspetto



Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

Utilizzo di un nodo Distribuzione

I nodi Distribuzione vengono utilizzati per rappresentare la distribuzione dei valori simbolici in un insieme di dati. Vengono spesso aggiunti prima dei nodi di manipolazione, in modo da analizzare i dati e correggere gli eventuali sbilanciamenti. Ad esempio, se nell'ambito di dati relativi a clienti che hanno risposto favorevolmente a campagne di marketing le istanze di famiglie con figli sono più frequenti rispetto ad altre tipologie di clienti, è possibile che si desideri ridurre tali istanze in modo da generare una regola più utile nelle successive operazioni di data mining. Un nodo Distribuzione consente di analizzare questo tipo di sbilanciamenti e prendere decisioni appropriate.

Il nodo Distribuzione è insolito in quanto produce sia un grafico sia una tabella per analizzare i dati.

Figura 5-25

Grafico di distribuzione che mostra il numero di individui con o senza figli che hanno risposto favorevolmente a una campagna di marketing

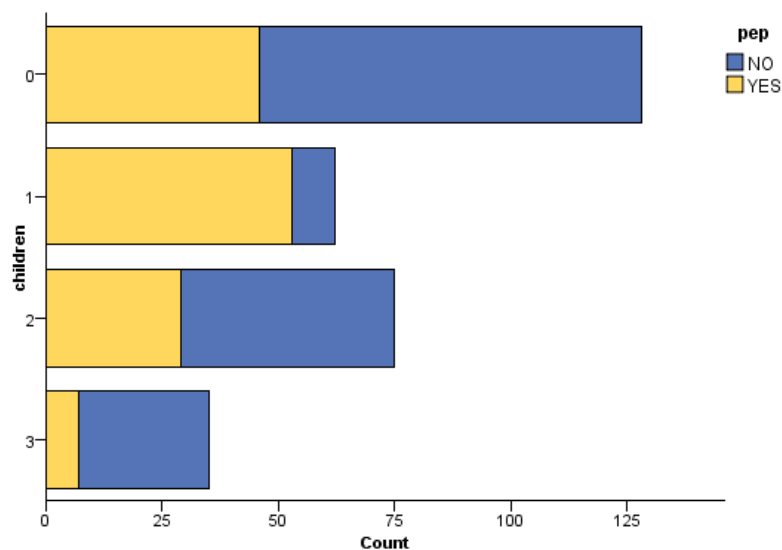
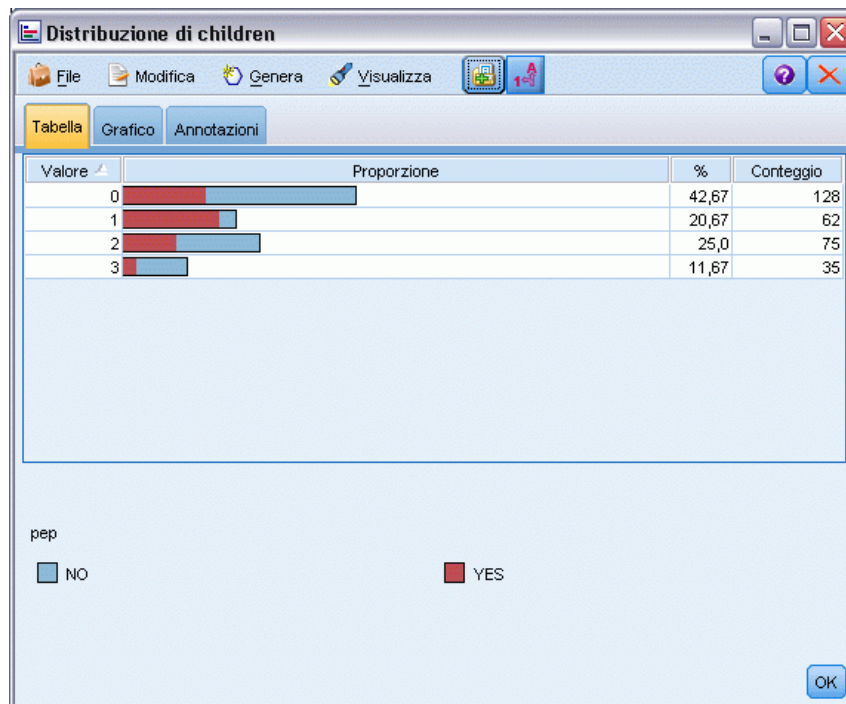


Figura 5-26

Tabella di distribuzione che mostra la proporzione degli individui con o senza figli che hanno risposto favorevolmente a una campagna di marketing



Dopo aver creato un grafico e una tabella di distribuzione e averne analizzato i risultati, è possibile utilizzare le opzioni dei menu per raggruppare i valori, copiarli e generare una serie di nodi per preparare i dati. Inoltre, è possibile copiare o esportare le informazioni della tabella o del grafico per utilizzarle in altre applicazioni, come MS Word o MS PowerPoint. [Per ulteriori informazioni, vedere l'argomento Stampa, salvataggio, copia ed esportazione di grafici a pag. 344.](#)

Per selezionare e copiare valori da una tabella di distribuzione

- ▶ Fare clic e tenere premuto il pulsante del mouse mentre si trascina il cursore sulle righe per selezionare un insieme di valori. È anche possibile utilizzare il menu Modifica e scegliere Seleziona tutto per selezionare tutti i valori.
- ▶ Dal menu Modifica scegliere Copia tabella o Copia tabella (inc. nomi di campo).
- ▶ Incollare negli Appunti o nell'applicazione desiderata.

Nota: le barre non vengono copiate direttamente. I valori della tabella, invece, vengono copiati. Pertanto, i valori sovrapposti non verranno visualizzati nella tabella copiata.

Per raggruppare valori da una tabella di distribuzione

- ▶ Selezionare i valori da raggruppare utilizzando il metodo Ctrl+clic.
- ▶ Dal menu Modifica scegliere Raggruppa.

Nota: Quando si raggruppano e si dividono i valori, il grafico nella scheda Grafico viene automaticamente ridisegnato per visualizzare le modifiche.

È anche possibile:

- Dividere valori selezionando il nome del gruppo nell'elenco di distribuzione e scegliendo *Separa* dal menu *Modifica*.
- Modificare gruppi selezionando il nome del gruppo nell'elenco di distribuzione e scegliendo *Modifica gruppo* dal menu *Modifica*. In questo modo viene aperta una finestra di dialogo in cui è possibile modificare i valori del gruppo.

Figura 5-27

Finestra di dialogo Modifica gruppo



Opzioni del menu *Genera*

È possibile utilizzare le opzioni del menu *Genera* per selezionare un sottoinsieme di dati, derivare un campo flag, raggruppare nuovamente valori, riclassificare valori o bilanciare i dati di un grafico o di una tabella. Tali opzioni generano un nodo *Data Preparation* e lo posizionano nell'area di disegno dello stream. Per utilizzare il nodo generato è necessario connetterlo a uno stream esistente. [Per ulteriori informazioni, vedere l'argomento *Generazioni di nodi da grafici* a pag. 319.](#)

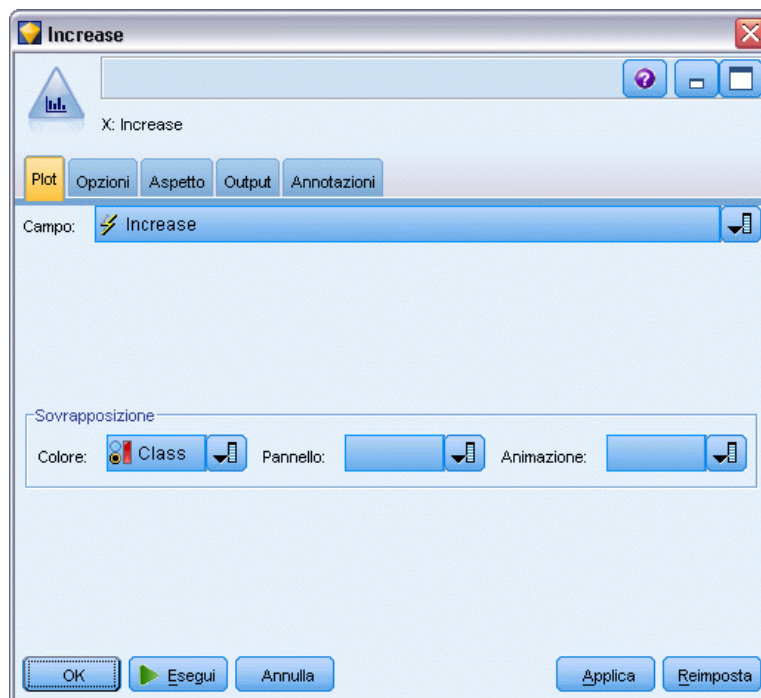
Nodo Istogramma

I nodi *Istogramma* mostrano l'occorrenza dei valori per i campi numerici. Vengono spesso utilizzati per analizzare i dati prima delle manipolazioni e della generazione del modello. Come il nodo *Distribuzione*, anche i nodi *Istogramma* vengono frequentemente utilizzati per rivelare sbilanciamenti nei dati. Benché per generare un istogramma sia possibile utilizzare anche il nodo *Lavagna grafica*, il nodo qui illustrato dispone di un numero maggiore di opzioni tra cui scegliere.

Nota: per mostrare l'occorrenza di valori per campi simbolici, è necessario utilizzare un nodo *Distribuzione*.

Scheda Plot del nodo Istogramma

Figura 5-28
Impostazioni della scheda Plot per un nodo Istogramma

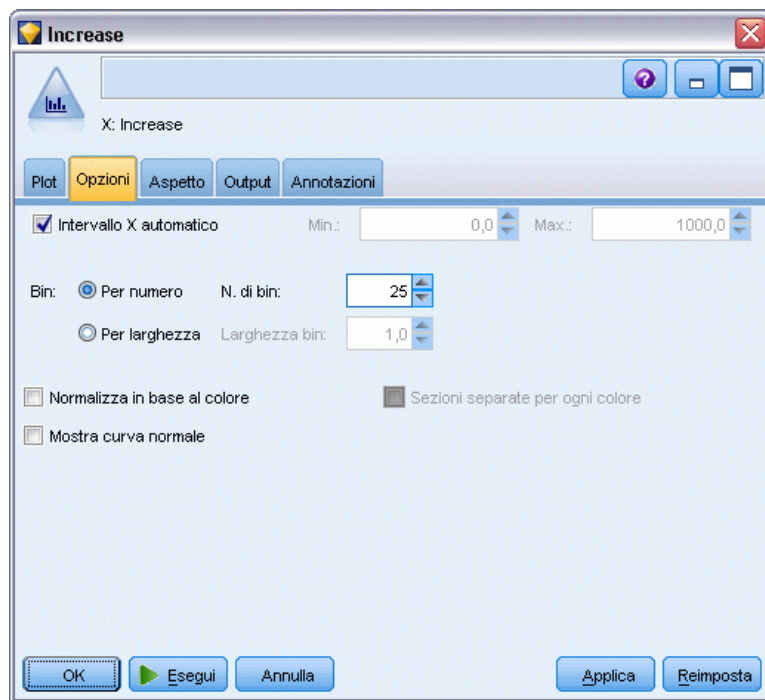


Campo. Selezionare un campo numerico per il quale mostrare la distribuzione dei valori. Verranno elencati solo i campi non esplicitamente definiti come simbolici (categoriali).

Sovrapposizione. Selezionare un campo simbolico per visualizzare le categorie dei valori per il campo specificato. La selezione di un campo di sovrapposizione converte l'istogramma in un grafico sovrapposto in cui le varie categorie del campo di sovrapposizione sono rappresentate da colori diversi. Nel nodo Istogramma sono disponibili tre tipi di sovrapposizioni: colori, pannelli e animazione. [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Scheda Opzioni del nodo Istogramma

Figura 5-29
Impostazioni della scheda Opzioni per un nodo Istogramma



Intervallo X automatico. Selezionare questa opzione per utilizzare l'intero intervallo di valori nei dati lungo l'asse. Deselezionare questa opzione per utilizzare un sottoinsieme esplicito di valori compresi tra i valori Min e Max specificati. Digitare i valori direttamente oppure utilizzare le frecce. Per default, gli intervalli vengono selezionati automaticamente in modo da consentire una più rapida creazione dei grafici.

Bin. Selezionare Per numero oppure Per larghezza.

- Selezionare Per numero per visualizzare un numero fisso di barre la cui larghezza dipende dall'intervallo e dal numero di bin specificati. Indicare il numero di bin da utilizzare nel grafico nell'opzione N. di bin. Utilizzare le frecce per modificare il numero.
- Selezionare Per larghezza per creare un grafico con barre di larghezza fissa. Il numero di bin dipende dalla larghezza e dall'intervallo di valori specificati. Indicare la larghezza delle barre nell'opzione Larghezza bin.

Normalizza in base al colore. Selezionare questa opzione per impostare tutte le barre sulla stessa altezza, visualizzando i valori sovrapposti come percentuale dei casi totali in ogni barra.

Mostra curva normale. Selezionare questa opzione per aggiungere al grafico una curva normale che mostri la media e la varianza dei dati.

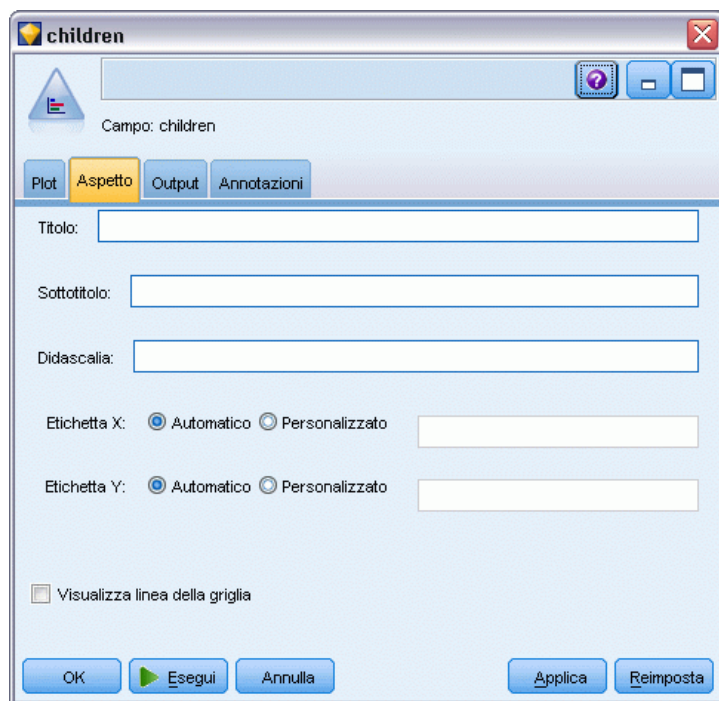
Sezioni separate per ogni colore. Selezionare questa opzione per visualizzare ogni valore sovrapposto come una sezione separata nel grafico.

Scheda Aspetto del nodo Istogramma

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-30

Impostazioni della scheda Aspetto per la maggior parte dei nodi Grafico



Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

Utilizzo degli istogrammi

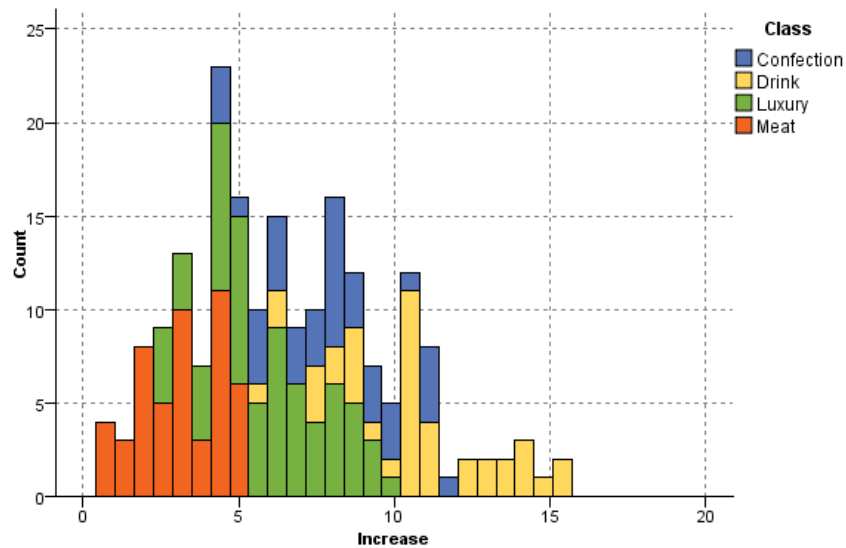
Gli istogrammi mostrano la distribuzione dei valori in un campo numerico i cui valori sono disposti lungo l'asse x . Gli istogrammi funzionano in modo simile ai grafici raccolta. Le raccolte mostrano la distribuzione dei valori di un campo numerico *rispetto ai valori di un altro campo* anziché l'occorrenza dei valori per un singolo campo.

Dopo avere creato un grafico è possibile esaminare i risultati e definire delle sezioni per suddividere i valori sull'asse x o definire delle aree. È possibile anche contrassegnare degli elementi del grafico. [Per ulteriori informazioni, vedere l'argomento Esplorazione dei grafici a pag. 309.](#)

È possibile utilizzare le opzioni del menu Genera per creare nodi Bilanciamento, Seleziona o Nuovo campo con i dati del grafico o più specificamente all'interno di sezioni, aree o elementi contrassegnati. Questo tipo di grafico viene spesso utilizzato prima dei nodi di manipolazione per analizzare i dati e correggere eventuali sbilanciamenti mediante la generazione di un nodo Bilanciamento da utilizzare nello stream a partire dal grafico. È anche possibile generare un nodo Nuovo campo flag per aggiungere un campo che indichi la sezione in cui ogni record è incluso, oppure un nodo Seleziona per selezionare tutti i record entro un particolare insieme o intervallo di valori. Queste operazioni consentono di delimitare in modo preciso un determinato sottoinsieme di dati per eseguire ulteriori analisi. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Figura 5-31

Istogramma che mostra la distribuzione della crescita degli acquisti per categoria a seguito di una promozione

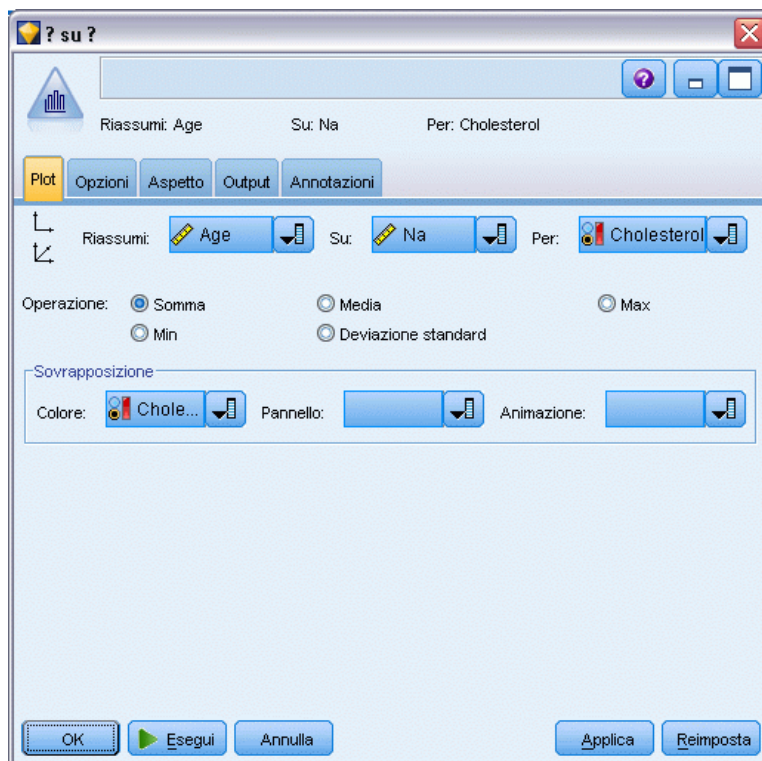


Nodo Raccolta

Le raccolte sono simili agli istogrammi tranne per il fatto che le prime mostrano la distribuzione dei valori di un campo numerico rispetto ai valori di un altro campo anziché l'occorrenza dei valori per un singolo campo. Una raccolta è utile per illustrare una variabile o un campo i cui valori vengono modificati nel tempo. La grafica 3-D consente inoltre di includere un asse simbolico che visualizza le distribuzioni per categoria. Le raccolte bidimensionali sono visualizzate come grafici a barre sovrapposte con sovrapposizioni (se utilizzate). [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Scheda Plot del nodo Raccolta

Figura 5-32
Impostazioni della scheda Plot per un nodo Raccolta



Riassumi. Selezionare un campo i cui valori verranno raccolti e visualizzati in base all'intervallo di valori del campo specificato in Su. Vengono elencati solo i campi che non sono stati definiti come simbolici.

Su. Selezionare un campo i cui valori verranno utilizzati per visualizzare il campo specificato in Riassumi.

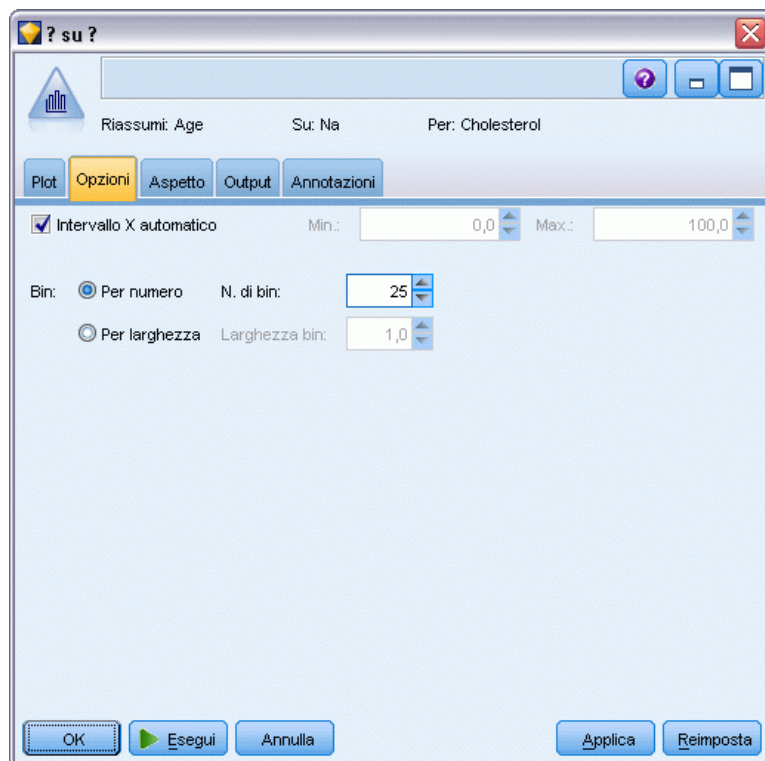
Per. Questa opzione, attivata quando si crea un grafico 3-D, consente di selezionare un campo nominale o flag che viene utilizzato per visualizzare il campo di raccolta per categorie.

Operazione. Selezionare il tipo di operazione che verrà rappresentata da ogni barra nel grafico raccolta. Le opzioni includono Somma, Media, Max, Min e Deviazione standard.

Sovrapposizione. Selezionare un campo simbolico per visualizzare le categorie dei valori per il campo selezionato. Se viene selezionato un campo di sovrapposizione, il grafico raccolta viene convertito e vengono aggiunte più barre di colori diversi per ogni categoria. Questo nodo dispone di tre tipi di sovrapposizioni: colori, pannelli e animazione. [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Scheda Opzioni del nodo Raccolta

Figura 5-33
Impostazioni della scheda Opzioni per un nodo Raccolta



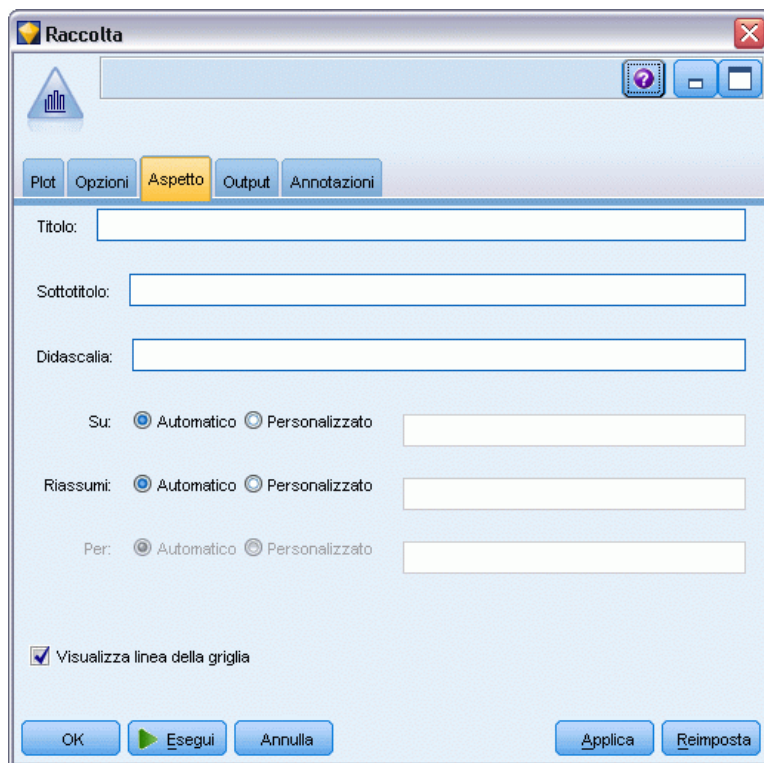
Intervallo X automatico. Selezionare questa opzione per utilizzare l'intero intervallo di valori nei dati lungo l'asse. Deselezionare questa opzione per utilizzare un sottoinsieme esplicito di valori compresi tra i valori Min e Max specificati. Digitare i valori direttamente oppure utilizzare le frecce. Per default, gli intervalli vengono selezionati automaticamente in modo da consentire una più rapida creazione dei grafici.

Bin. Selezionare Per numero oppure Per larghezza.

- Selezionare Per numero per visualizzare un numero fisso di barre la cui larghezza dipende dall'intervallo e dal numero di bin specificati. Indicare il numero di bin da utilizzare nel grafico nell'opzione N. di bin. Utilizzare le frecce per modificare il numero.
- Selezionare Per larghezza per creare un grafico con barre di larghezza fissa. Il numero di bin dipende dalla larghezza e dall'intervallo di valori specificati. Indicare la larghezza delle barre nell'opzione Larghezza bin.

Scheda Aspetto del nodo Raccolta

Figura 5-34
Impostazioni della scheda Aspetto per un nodo Raccolta



Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Su. Accettare l'etichetta automaticamente generata oppure selezionare Personalizzato per specificare un'etichetta.

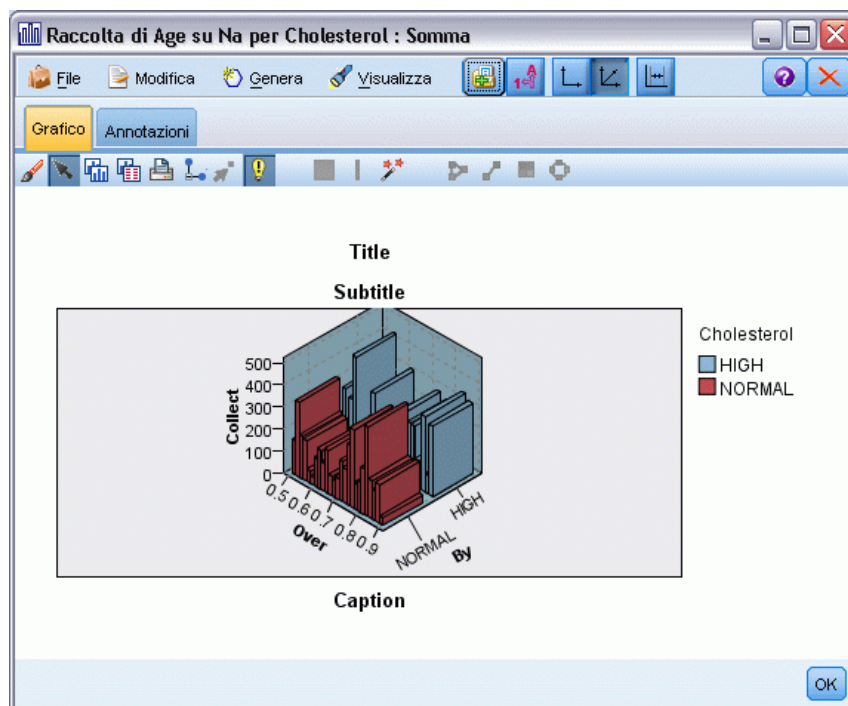
Riassumi. Accettare l'etichetta automaticamente generata oppure selezionare Personalizzato per specificare un'etichetta.

Per. Accettare l'etichetta automaticamente generata oppure selezionare Personalizzato per specificare un'etichetta.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

Il seguente esempio mostra dove vengono posizionate le opzioni della scheda Aspetto su una versione 3-D del grafico.

Figura 5-35
Posizione delle opzioni della scheda Aspetto in un grafico raccolta 3-D



Utilizzo del grafico di un nodo Raccolta

Le raccolte mostrano la distribuzione dei valori di un campo numerico *rispetto ai valori di un altro campo* anziché l'occorrenza dei valori per un singolo campo. Gli istogrammi funzionano in modo simile ai grafici raccolta. Gli istogrammi mostrano la distribuzione dei valori in un campo numerico i cui valori sono disposti lungo l'asse x .

Dopo avere creato un grafico è possibile esaminare i risultati e definire delle sezioni per suddividere i valori sull'asse x o definire delle aree. È possibile anche contrassegnare degli elementi del grafico. [Per ulteriori informazioni, vedere l'argomento Esplorazione dei grafici a pag. 309.](#)

È possibile utilizzare le opzioni del menu Genera per creare nodi Bilanciamento, Seleziona o Nuovo campo con i dati del grafico o più specificamente all'interno di sezioni, aree o elementi contrassegnati. Questo tipo di grafico viene spesso utilizzato prima dei nodi di manipolazione per analizzare i dati e correggere eventuali sbilanciamenti mediante la generazione di un nodo Bilanciamento da utilizzare nello stream a partire dal grafico. È anche possibile generare un nodo Nuovo campo flag per aggiungere un campo che indichi la sezione in cui ogni record è incluso, oppure un nodo Seleziona per selezionare tutti i record entro un particolare insieme o intervallo di valori. Queste operazioni consentono di delimitare in modo preciso un determinato sottoinsieme di dati per eseguire ulteriori analisi. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Figura 5-36

Grafico raccolta 3-D che mostra la somma di Na_su_K rispetto a Età per livelli di colesterolo alti e nella norma

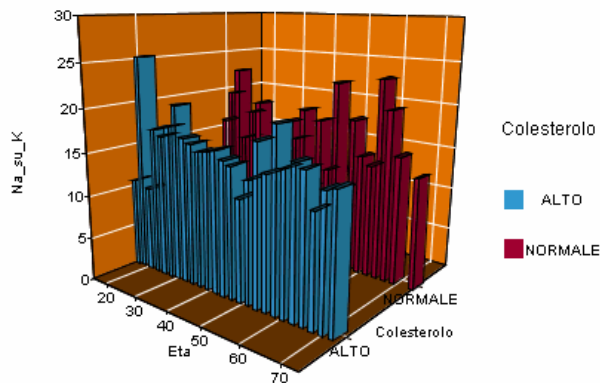
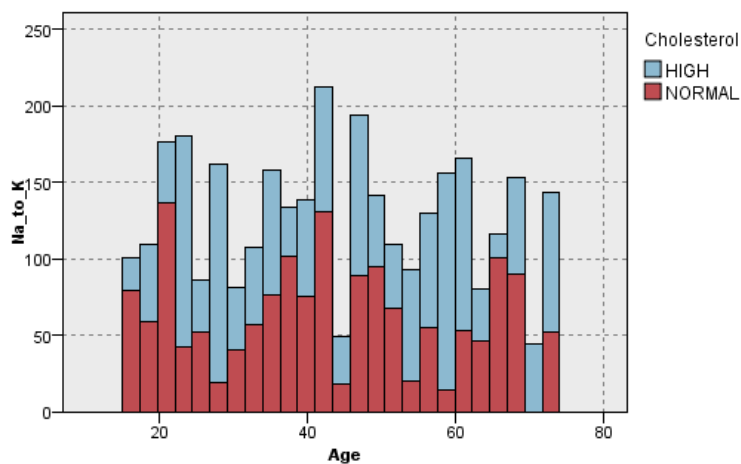


Figura 5-37

Grafico raccolta senza la visualizzazione dell'asse z ma con il campo Colesterolo come sovrapposizione di colore

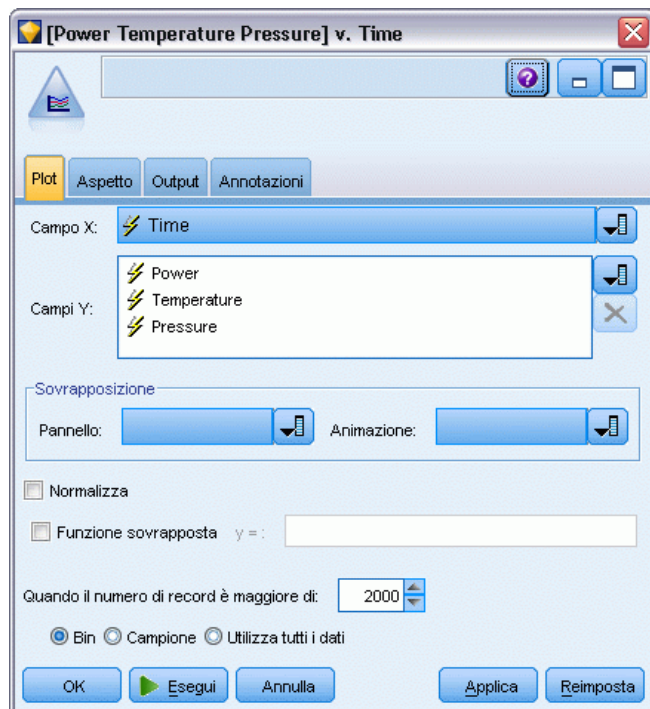


Nodo Plot multiplo

Un plot multiplo è un tipo speciale di grafico che consente di visualizzare più campi Y in un singolo campo X . I campi Y sono rappresentati come linee colorate e ognuno di essi equivale a un nodo Plot con lo Stile impostato su Linea e la Modalità X impostata su Ordina. I plot multipli sono utili quando si utilizzano dati di serie storica e si desidera analizzare la fluttuazione di più variabili nel tempo.

Scheda Plot di un nodo Plot multiplo

Figura 5-38
Impostazioni della scheda Plot per un nodo Plot multiplo



Campo X. Selezionare dall'elenco il campo da visualizzare lungo l'asse x orizzontale.

Campi Y. Selezionare dall'elenco uno o più campi da visualizzare sull'intervallo dei valori del campo X. Utilizzare il pulsante di selezione dei campi per selezionare più campi. Fare clic sul pulsante contrassegnato da una X per rimuovere i campi dall'elenco.

Sovrapposizione. Esistono diverse modalità di rappresentazione delle categorie per i valori dei dati. Ad esempio, è possibile utilizzare una sovrapposizione di animazione per visualizzare plot multipli per ogni valore nei dati. Questo è utile per gli insiemi che contengono più di 10 categorie. Nel caso di insiemi con più di 15 categorie si può notare una diminuzione delle prestazioni. [Per ulteriori informazioni, vedere l'argomento Effetti estetici, sovrapposizioni, riquadri e animazioni a pag. 232.](#)

Normalizza. Selezionare questa opzione se si desidera graduare tutti i valori Y nell'intervallo 0–1 per la visualizzazione nel grafico. La normalizzazione consente di visualizzare in modo chiaro la relazione tra linee che potrebbero risultare non completamente visibili a causa di differenze nell'intervallo dei valori di ogni serie ed è consigliata quando si rappresentano più linee nello stesso grafico o quando si confrontano plot in riquadri affiancati (la normalizzazione non è necessaria quando tutti i valori di dati rientrano in un intervallo simile).

Figura 5-39

Plot multiplo standard che mostra la fluttuazione di una centrale elettrica nel tempo (si noti che senza la normalizzazione è impossibile vedere il grafico relativo alla pressione)

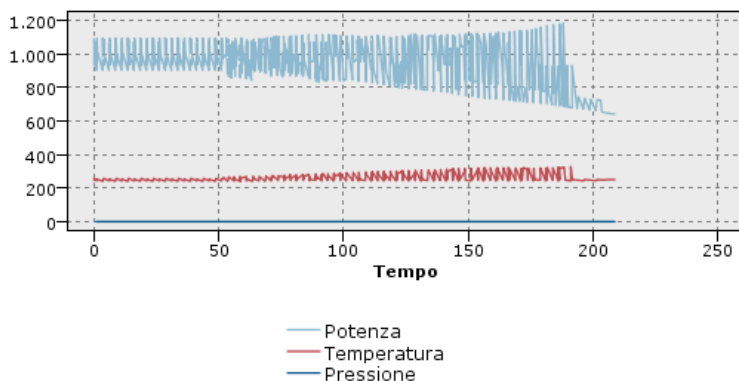
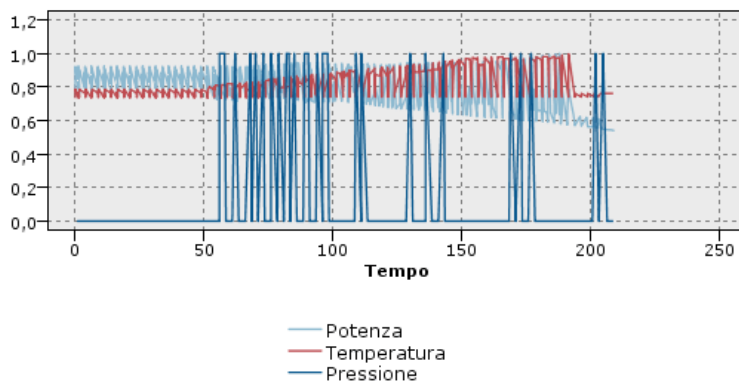


Figura 5-40

Plot multiplo normalizzato che mostra un grafico relativo a valori di pressione



Funzione sovrapposta. Selezionare questa opzione per specificare una funzione conosciuta da confrontare con i valori effettivi. Ad esempio, per confrontare valori effettivi con valori previsti, è possibile rappresentare la funzione $y = x$ come una sovrapposizione. Specificare una funzione nella casella di testo $y =$. La funzione di default è $y = x$, tuttavia è possibile specificare una funzione qualsiasi, ad esempio una funzione quadratica o un'espressione arbitraria, al posto di x .

Nota: le funzioni sovrapposte non sono disponibili per un grafico contenente pannelli o animazione.

Quando il numero di record è maggiore di. Specifica un metodo per rappresentare insiemi di dati di notevoli dimensioni. È possibile specificare una dimensione massima per l'insieme di dati oppure utilizzare l'impostazione di default di 2.000 punti. Per gli insiemi di dati di grandi dimensioni le performance risultano migliori se si seleziona l'opzione Bin o Campione. In alternativa, è possibile rappresentare tutti i punti dei dati selezionando Utilizza tutti i dati, tuttavia occorre tenere presente che in questo modo le prestazioni del software potrebbero risultare notevolmente ridotte.

Nota: quando la Modalità X viene impostata su Sovrapposizione o su Ordine di lettura, queste opzioni non sono disponibili e vengono utilizzati solo i primi n record.

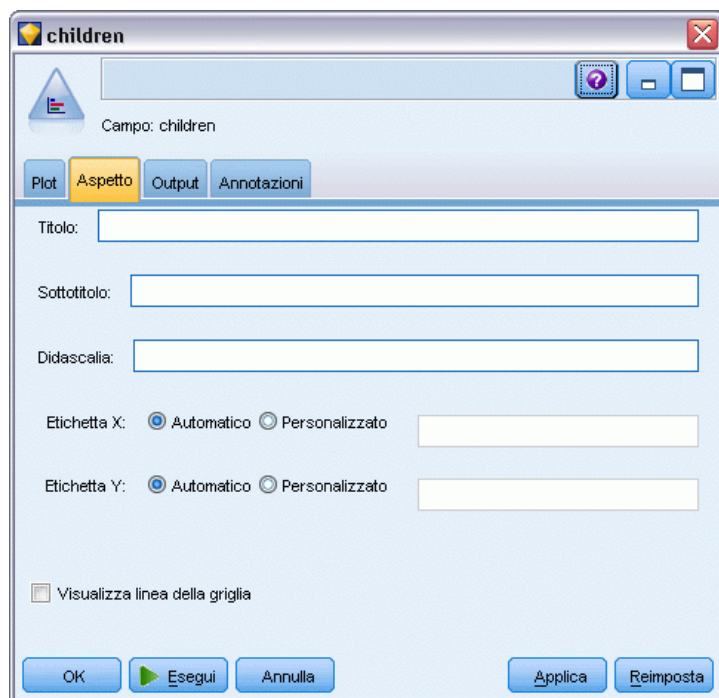
- **Bin.** Selezionare questa opzione per attivare gli intervalli quando l'insieme di dati contiene un numero di record maggiore di quello specificato. L'attivazione degli intervalli determina la divisione del grafico in griglie fini prima della rappresentazione effettiva e il conteggio del numero di connessioni che verrebbero visualizzate in ogni cella della griglia. Nel grafico finale viene rappresentata un'unica connessione per cella sul baricentro del bin (media di tutti i punti di connessione nel bin).
- **Esempio.** Selezionare questa opzione per eseguire un campionamento casuale dei dati sul numero di record specificato.

Scheda Aspetto del nodo Plot multiplo

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-41

Impostazioni della scheda Aspetto per la maggior parte dei nodi Grafico



Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

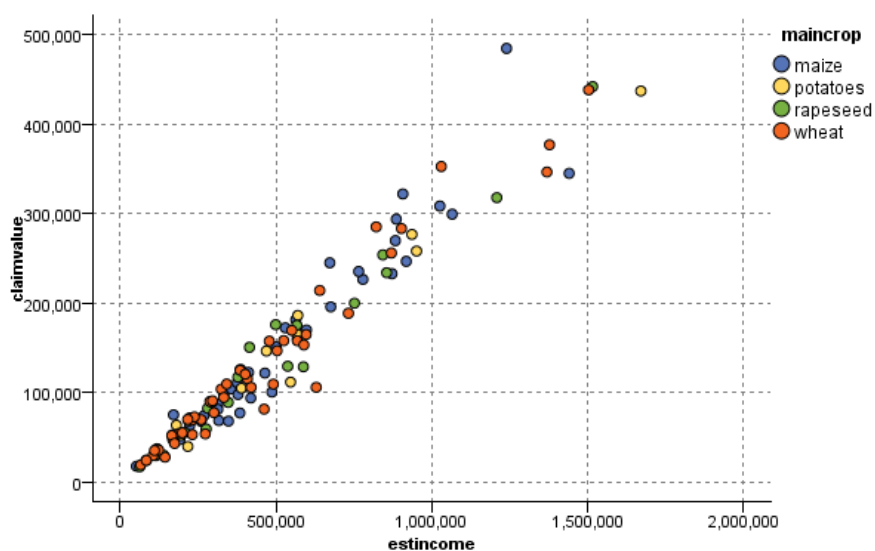
Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

Utilizzo del grafico di un nodo Plot multiplo

I nodi Plot e Plot multiplo creano essenzialmente grafici in cui X viene rappresentato rispetto a Y . Per esempio, se si sta analizzando una potenziale frode nelle richieste di concessioni agricole (come illustrato in *fraud.str* nella cartella *Demos* dell'installazione di IBM® SPSS® Modeler), si potrebbe voler rappresentare il reddito dichiarato nella domanda a fronte di quello stimato da una rete neurale. Una sovrapposizione, ad esempio il tipo di raccolto, consentirà di vedere se esiste una relazione tra domande (valore o numero) e tipo di raccolto.

Figura 5-42

Grafico della relazione tra reddito stimato e valore della domanda con il tipo raccolto principale come sovrapposizione

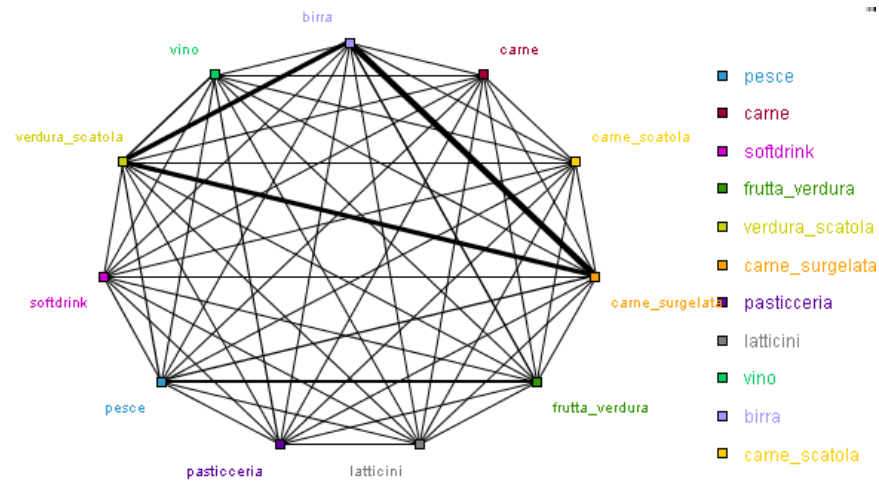


Dal momento che i grafici prodotti dai nodi Plot, Plot multiplo e Valutazione sono visualizzazioni bidimensionali di Y rispetto a X , è facile interagire con essi selezionando le aree, contrassegnando elementi o persino disegnando sezioni mediante il mouse. È inoltre possibile generare nodi per i dati rappresentati da queste aree, sezioni o elementi. [Per ulteriori informazioni, vedere l'argomento Esplorazione dei grafici a pag. 309.](#)

Nodo Web

I nodi Web mostrano l'intensità delle relazioni tra valori di due o più campi simbolici. Il grafico mostra le connessioni utilizzando tipi di linea diversi a seconda dell'intensità delle connessioni. Un nodo Web può essere utilizzato, ad esempio, per analizzare la relazione tra l'acquisto di vari oggetti in un sito di e-commerce o in un tradizionale punto vendita al dettaglio.

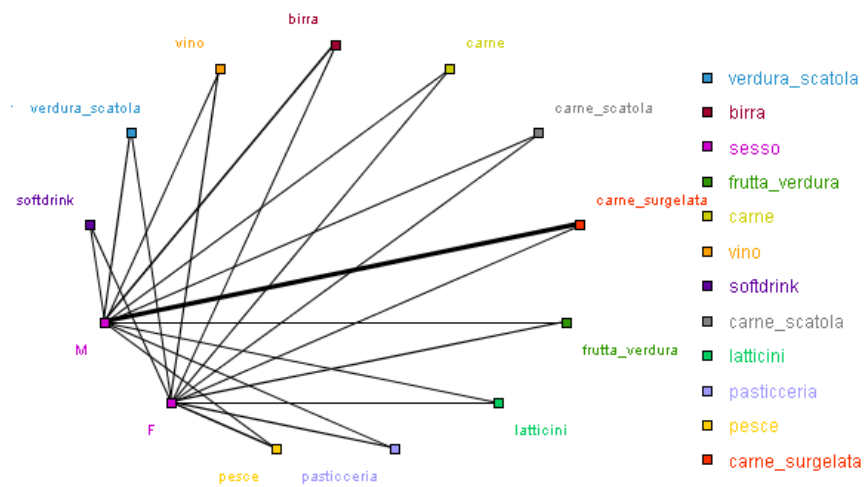
Figura 5-43
Grafico Web che mostra le relazioni tra l'acquisto di generi alimentari



Grafici Web diretti

I nodi Web diretti mostrano, come i nodi Web, l'intensità delle relazioni tra campi simbolici. Tuttavia, i grafici Web diretti mostrano esclusivamente le connessioni da uno o più campi Da a un singolo campo A. Le connessioni sono unidirezionali, ovvero procedono solo in un senso.

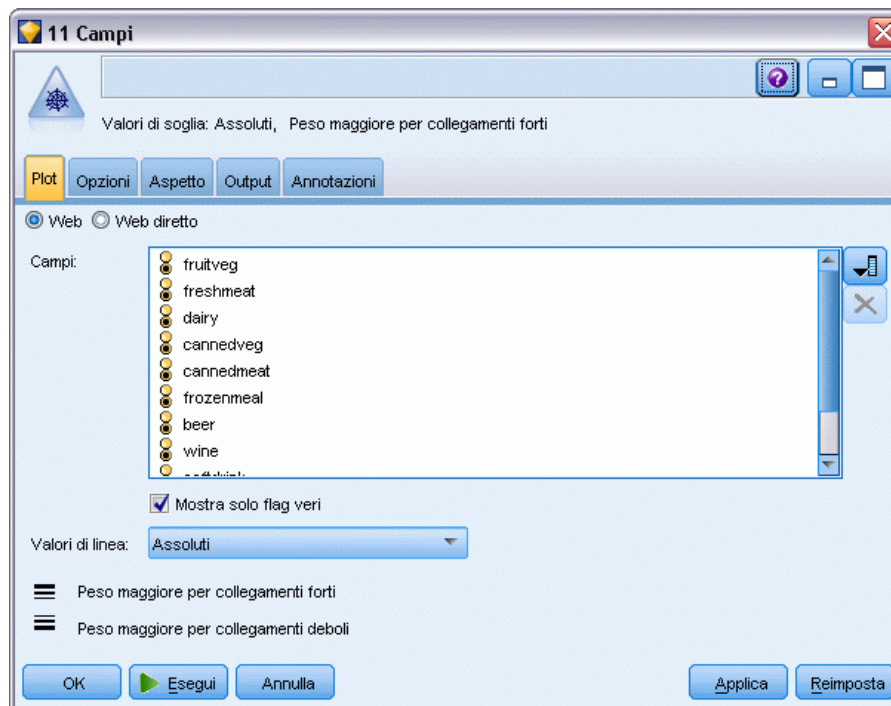
Figura 5-44
Grafico Web diretto che mostra la relazione tra l'acquisto di generi alimentari e il sesso dei clienti



Analogamente ai nodi Web, il grafico visualizza le connessioni utilizzando tipi di linea diversi a seconda dell'intensità della connessione. Un nodo Web diretto può essere utilizzato, ad esempio, per analizzare la relazione tra il sesso dei clienti e la propensione per alcuni prodotti.

Scheda Plot del nodo Web

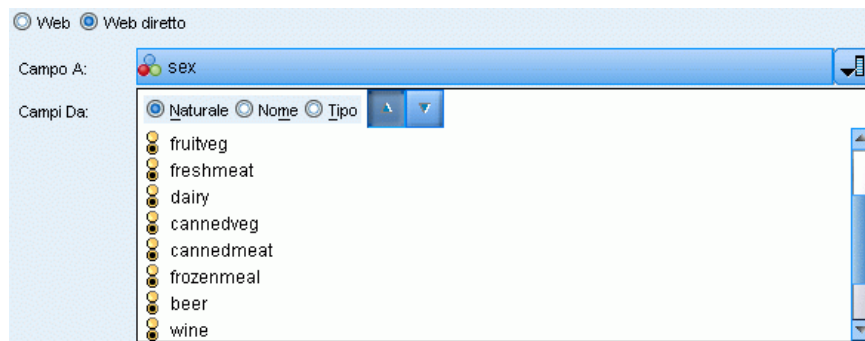
Figura 5-45
Impostazioni della scheda Plot per un nodo Web



Web Selezionare questa opzione per creare un grafico Web che illustri l'intensità delle relazioni tra tutti i campi specificati.

Web diretto. Selezionare questa opzione per creare un grafico Web direzionale che illustri l'intensità delle relazioni tra più campi e i valori di un campo, ad esempio il sesso o la religione. Quando questa opzione è selezionata, viene attivato un Campo A e, per maggiore chiarezza, il controllo Campi viene rinominato Campi Da.

Figura 5-46
Opzioni relative a Web diretto



Campo A (solo Web diretto). Selezionare un campo flag o nominale da utilizzare per un Web diretto. Verranno elencati solo i campi non esplicitamente impostati come numerici.

Campi/Campi Da. Selezionare i campi per la creazione di un grafico Web. Verranno elencati solo i campi non esplicitamente impostati come numerici. Utilizzare il pulsante Selezione campi per selezionare più campi o per selezionarli in base al tipo.

Nota: per un Web diretto, questo controllo è utilizzato per selezionare i campi Da.

Mostra solo flag veri. Selezionare questa opzione per visualizzare solo i flag veri per un campo flag. Questa opzione semplifica la visualizzazione Web e spesso viene utilizzata per i dati in cui l'occorrenza di valori positivi è molto importante.

Valori di linea. Selezionare un tipo di soglia dall'elenco a discesa.

- Assoluti imposta le soglie in base al numero di record che includono ogni coppia di valori.
- Percentuali globali mostra il numero assoluto di casi rappresentati dal collegamento come percentuale di tutte le occorrenze di ogni coppia di valori rappresentata nel grafico Web.
- Percentuali del campo/valore più piccolo e Percentuali del campo/valore più grande indicano il campo/valore da utilizzare per la valutazione delle percentuali. Ad esempio, si supponga che 100 record includano il valore *curaY* per il campo *Cura* e solo 10 abbiano il valore *BASSO* per il campo *Pressione*. Se sette record includono entrambi i valori *curaY* e *BASSO*, questa percentuale è 70% o 7% a seconda del campo a cui si fa riferimento, più piccolo (*Pressione*) o più grande (*Cura*).

Nota: per i grafici Web diretti non sono disponibili la terza e la quarta opzione sopra riportate. Sono invece disponibili le opzioni Percentuali di campo/valore "A" e Percentuali di campo/valore "Da".

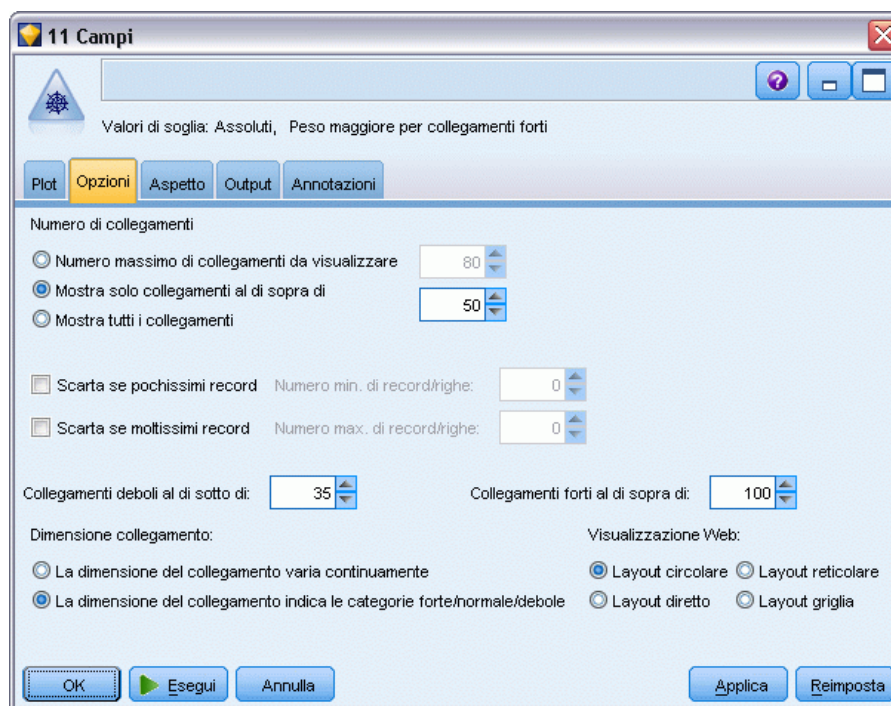
Peso maggiore per collegamenti forti. Questa opzione, selezionata per default, rappresenta la modalità comunemente utilizzata per visualizzare i collegamenti tra i campi.

Peso maggiore per collegamenti deboli. Selezionare questa opzione per invertire il significato dei collegamenti rappresentati da linee in grassetto. Questa opzione viene in genere utilizzata per il rilevamento di comportamenti fraudolenti o l'analisi di valori anomali.

Scheda Opzioni del nodo Web

La scheda Opzioni per i nodi Web contiene alcune opzioni aggiuntive per la personalizzazione del grafico di output.

Figura 5-47
Impostazioni della scheda Opzioni per un nodo Web



Numero di collegamenti. Le seguenti opzioni sono utilizzate per impostare il numero di collegamenti visualizzati nel grafico di output. Alcune di queste opzioni, ad esempio Collegamenti deboli al di sotto di e Collegamenti forti al di sopra di, sono disponibili anche nella finestra del grafico di output. Nel grafico finale è inoltre possibile utilizzare un dispositivo di scorrimento per modificare il numero di collegamenti visualizzati.

- **Numero massimo di collegamenti da visualizzare.** Specificare un numero che indichi il numero massimo di collegamenti da visualizzare nel grafico di output. Utilizzare le frecce per modificare il valore.
- **Mostra solo collegamenti al di sopra di.** Specificare un numero che indichi il valore minimo per il quale mostrare una connessione nel grafico Web. Utilizzare le frecce per modificare il valore.
- **Mostra tutti i collegamenti.** Specificare questa opzione per visualizzare tutti i collegamenti a prescindere dai valori minimi o massimi. Se i campi sono numerosi, la selezione di questa opzione potrebbe determinare un aumento nel tempo di elaborazione.

Scarta se pochissimi record. Selezionare questa opzione per ignorare le connessioni supportate da un numero limitato di record. Impostare la soglia per questa opzione immettendo un numero in Numero min. di record/righe.

Scarta se moltissimi record. Selezionare questa opzione per ignorare le connessioni supportate da un numero di record eccessivo. Immettere un numero in Numero max. di record/righe.

Collegamenti deboli al di sotto di. Specificare un numero che indichi la soglia di separazione tra le connessioni deboli (linee tratteggiate) e le connessioni normali (linee normali). Tutte le connessioni inferiori a questo valore verranno considerate deboli.

Collegamenti forti al di sopra di. Specificare una soglia di separazione tra le connessioni forti (linee forti) e le connessioni normali (linee normali). Tutte le connessioni superiori a questo valore verranno considerate forti.

Dimensione collegamento. Specificare le opzioni per controllare la dimensione dei collegamenti:

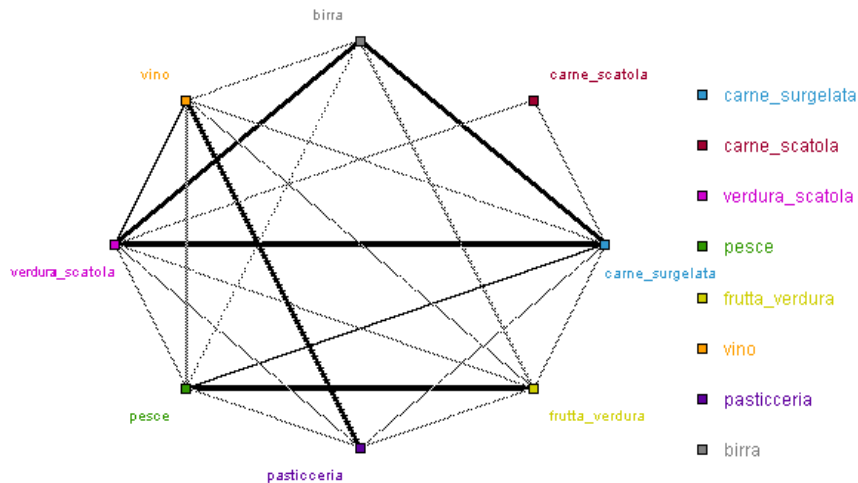
- **La dimensione del collegamento varia continuamente.** Selezionare questa opzione per visualizzare un intervallo delle dimensioni dei collegamenti che rifletta la variazione nell'intensità delle connessioni in base agli effettivi valori dei dati.
- **La dimensione del collegamento indica le categorie forte/normale/debole.** Selezionare questa opzione per visualizzare tre tipi di connessioni, ovvero forte, normale e debole. I punti di interruzione per queste categorie possono essere specificati sia in questa finestra sia nel grafico finale.

Visualizzazione Web. Selezionare un tipo di visualizzazione Web:

- **Layout circolare.** Selezionare questa opzione per utilizzare la visualizzazione Web standard.
- **Layout reticolare.** Selezionare questa opzione per utilizzare un algoritmo in modo da raggruppare i collegamenti più forti. Questa opzione è stata progettata per evidenziare i collegamenti forti utilizzando sia la differenziazione spaziale sia le linee ponderate.
- **Layout diretto.** Selezionare questa opzione per creare una visualizzazione Web diretta che utilizza la selezione Campo A della scheda Plot come direzione.
- **Layout griglia.** Selezionare questa opzione per creare una visualizzazione Web strutturata in uno schema a griglia con spaziature uniformi.

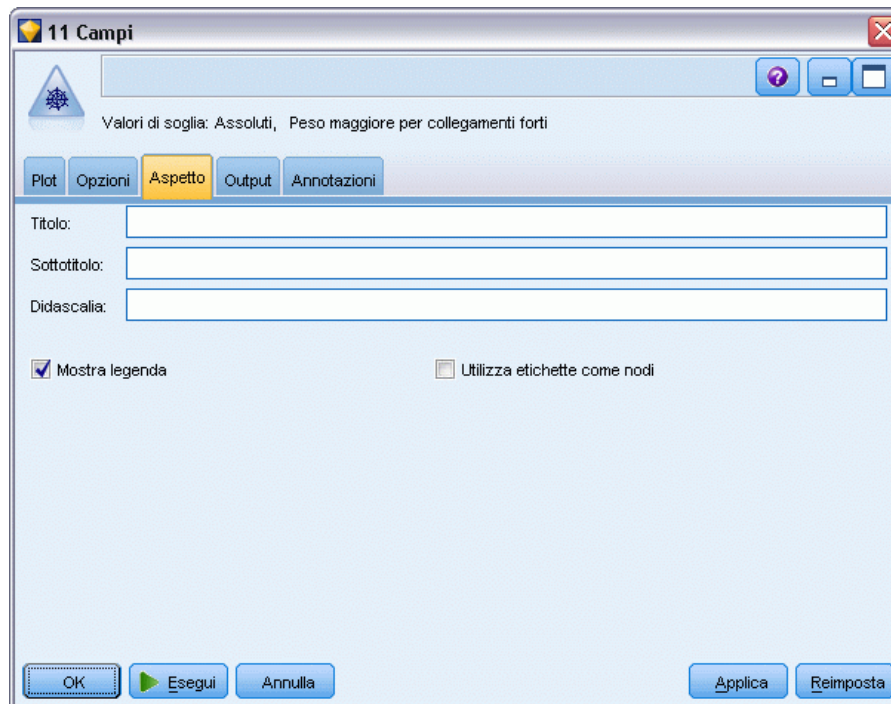
Figura 5-48

Grafico Web che mostra le connessioni forti tra diversi generi alimentari



Scheda Aspetto del nodo Web

Figura 5-49
Impostazioni della scheda Aspetto per un nodo Web



Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Titolo. Immettere il testo da utilizzare come titolo del grafico.

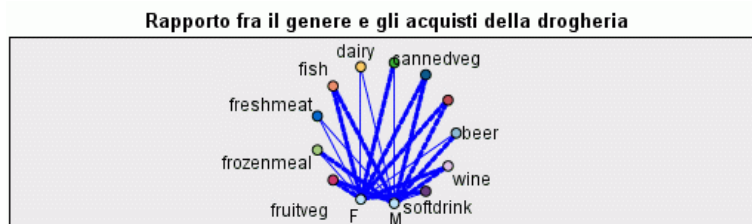
Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Mostra legenda. Specifica se la legenda viene visualizzata. Per i grafici Plot con un numero elevato di campi, può essere consigliabile nascondere la legenda per migliorare l'aspetto del grafico.

Utilizza etichette come nodi. Include il testo delle etichette in ogni nodo anziché visualizzare le etichette adiacenti. Nei grafici con un numero limitato di campi, questa opzione può aumentare la leggibilità del grafico.

Figura 5-50
Grafico Web che visualizza le etichette come nodi



Utilizzo di un grafico Web

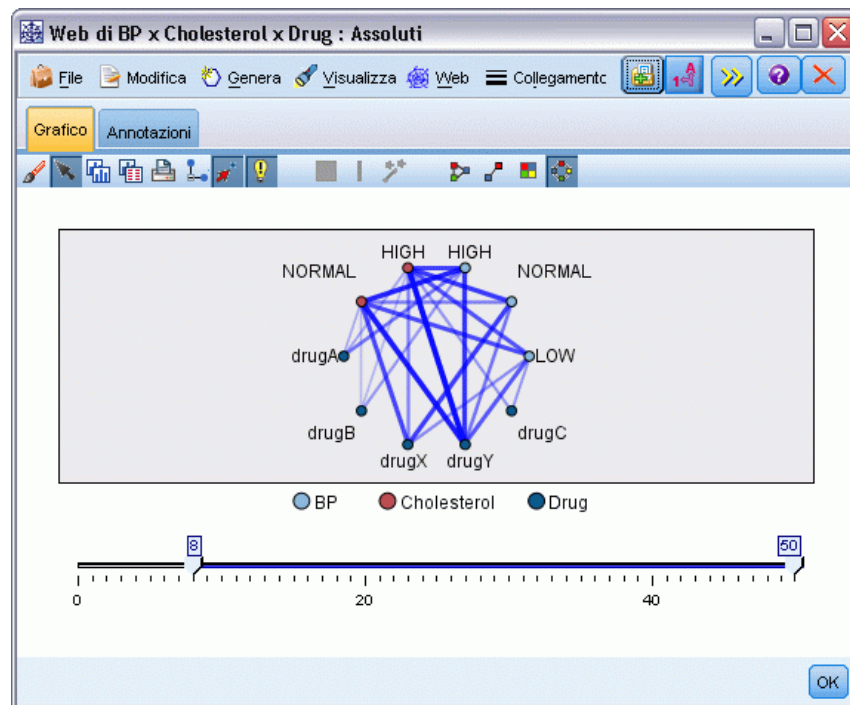
I nodi Web sono utilizzati per mostrare l'intensità delle relazioni tra i valori di due o più campi simbolici. Le connessioni sono visualizzate in un grafico con tipi di linea diversi a seconda dell'intensità. Un nodo Web può essere utilizzato, ad esempio, per analizzare la relazione tra i livelli di colesterolo, la pressione sanguigna e la cura che si è rivelata efficace nel trattamento della malattia del paziente.

- Le connessioni forti sono rappresentate da una linea spessa. Questo tipo di linea indica che i due valori sono fortemente correlati e devono essere sottoposti a un'analisi più approfondita.
- Le connessioni medie sono rappresentate da una linea di spessore normale.
- Le connessioni deboli sono rappresentate da una linea tratteggiata.
- L'assenza di una linea di collegamento tra due valori indica che i due valori non ricorrono mai nello stesso record oppure sono contemporaneamente presenti in un numero di record inferiore alla soglia specificata nella finestra di dialogo del nodo Web.

Dopo aver creato un nodo Web, è possibile utilizzare diverse opzioni per modificare la visualizzazione del grafico e generare nodi per ulteriori analisi.

Figura 5-51

Grafico Web che indica un numero di relazioni forti, ad esempio una pressione sanguigna normale con curaX e un livello di colesterolo elevato con curaY



Sia per i nodi Web che per i nodi Web diretto è possibile:

- Modificare il layout della visualizzazione Web.
- Nascondere punti per semplificare la visualizzazione.
- Modificare le soglie che controllano gli stili di linea.

- Evidenziare le linee tra i valori per indicare una relazione “selezionata”.
- Generare un nodo Seleziona per uno o più record “selezionati” o un nodo Nuovo campo flag associato a una o più relazioni nel grafico Web.

Per modificare i punti

- **Spostare** i punti facendo clic con il pulsante del mouse su un punto e trascinandolo in una nuova posizione. Il grafico Web verrà ridisegnato per riflettere la nuova posizione.
- **Nascondere** i punti facendo clic con il pulsante destro del mouse su un punto nel grafico Web e scegliendo Nascondi o Nascondi e ripianifica dal menu di scelta rapida. Nascondi nasconde semplicemente il punto selezionato e tutte le linee ad esso associate. Nascondi e ripianifica ridisegna il grafico Web in base alle modifiche apportate. Qualsiasi spostamento manuale viene annullato.
- **Mostrare** tutti i punti nascosti scegliendo Mostra tutto o Mostra tutto e ripianifica dal menu Web nella finestra del grafico. Se si seleziona Mostra tutto e ripianifica, il grafico Web viene ridisegnato per includere tutti i punti nascosti in precedenza e le relative connessioni.

Per selezionare o evidenziare le linee

Le linee selezionate sono evidenziate in rosso.

- ▶ Per selezionare una sola linea, fare clic su di essa con il pulsante sinistro del mouse.
- ▶ Per selezionare più linee, effettuare una delle seguenti operazioni:
 - Utilizzando il cursore, disegnare un cerchio attorno ai punti di cui si desidera selezionare le linee.
 - Tenere premuto il tasto Ctrl e fare clic con il pulsante sinistro del mouse sulle singole linee da selezionare.

Per deselezionare tutte le linee selezionate, fare clic sullo sfondo del grafico oppure scegliere Cancella selezione dal menu Web nella finestra del grafico.

Per visualizzare il grafico Web utilizzando un layout diverso

- ▶ Dal menu Web, scegliere Layout circolare, Layout reticolare, Layout diretto o Layout griglia per modificare il layout del grafico.

Per attivare e disattivare il dispositivo di scorrimento dei collegamenti

- ▶ Dal menu Visualizza, scegliere Dispositivo di scorrimento dei collegamenti.

Per selezionare o generare un flag per i record di una singola relazione

- ▶ Fare clic con il pulsante destro del mouse sulla linea che rappresenta la relazione desiderata.
- ▶ Dal menu di scelta rapida scegliere Genera nodo Seleziona per il collegamento o Genera nodo Nuovo campo per il collegamento.

Nell'area di disegno dello stream viene aggiunto automaticamente un nodo Seleziona o un nodo Nuovo campo con le opzioni appropriate e le condizioni specificate:

- Il nodo Seleziona seleziona tutti i record nella relazione selezionata.
- Il nodo Nuovo campo genera un flag che indica se la relazione selezionata è vera per i record dell'intero insieme di dati. Il nome del campo flag è composto dai due valori presenti nella relazione uniti da un carattere di sottolineatura, ad esempio *BASSO_curaC* o *curaC_BASSO*.

Per selezionare o generare un flag per i record di un gruppo di relazioni

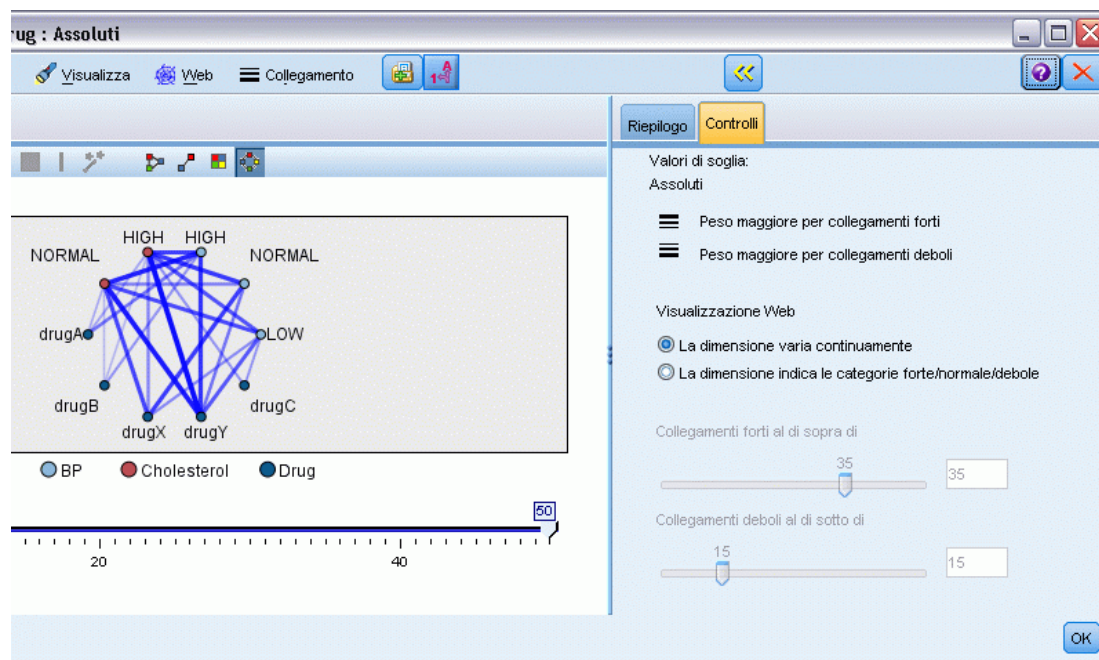
- ▶ Selezionare la linea o le linee nella visualizzazione Web che rappresentano le relazioni selezionate.
- ▶ Dal menu Genera nella finestra del grafico, scegliere Nodo Seleziona ("E"), Nodo Seleziona ("O"), Nodo Nuovo campo ("E") o Nodo Nuovo campo ("O").
 - I nodi "O" sono basati sulla separazione delle condizioni. Ciò significa che il nodo verrà applicato ai record per i quali vale una qualsiasi delle relazioni selezionate.
 - I nodi "E" sono basati sull'unione delle condizioni. Ciò significa che il nodo verrà applicato solo ai record per i quali valgono tutte le relazioni selezionate. Se le relazioni selezionate si escludono a vicenda, viene generato un errore.

Al termine della selezione, nell'area di disegno dello stream viene automaticamente aggiunto un nodo Seleziona o Nuovo campo, con le opzioni appropriate e le condizioni specificate.

Modifica delle soglie di un grafico Web

Al termine della creazione di un grafico Web, è possibile modificare le soglie che controllano gli stili della linea utilizzando il dispositivo di scorrimento della barra degli strumenti per modificare la linea minima visibile. È anche possibile visualizzare opzioni aggiuntive di soglia facendo clic sul pulsante con due frecce gialle disponibile nella barra degli strumenti per espandere la finestra del grafico Web. Quindi fare clic sulla scheda Controlli per visualizzare le opzioni aggiuntive.

Figura 5-52
Finestra espansa con le opzioni di visualizzazione e di soglia



Valori di soglia. Mostra il tipo di soglia selezionato durante la creazione nella finestra di dialogo del nodo Web.

Peso maggiore per collegamenti forti. Questa opzione, selezionata per default, rappresenta la modalità comunemente utilizzata per visualizzare i collegamenti tra i campi.

Peso maggiore per collegamenti deboli. Selezionare questa opzione per invertire il significato dei collegamenti rappresentati da linee in grassetto. Questa opzione viene in genere utilizzata per il rilevamento di comportamenti fraudolenti o l'analisi di valori anomali.

Visualizzazione Web. Specificare le opzioni per controllare la dimensione dei collegamenti nel grafico di output:

- **La dimensione varia continuamente.** Selezionare questa opzione per visualizzare un intervallo delle dimensioni dei collegamenti che rifletta la variazione nell'intensità delle connessioni in base agli effettivi valori dei dati.
- **La dimensione indica le categorie forte/normale/debole.** Selezionare questa opzione per visualizzare tre tipi di connessioni, ovvero forte, normale e debole. I punti di interruzione per queste categorie possono essere specificati sia in questa finestra sia nel grafico finale.

Collegamenti forti al di sopra di. Specificare una soglia di separazione tra le connessioni forti (linee forti) e le connessioni normali (linee normali). Tutte le connessioni superiori a questo valore verranno considerate forti. Utilizzare il dispositivo di scorrimento per modificare il valore o immettere un numero nel campo.

Collegamenti deboli al di sotto di. Specificare un numero che indichi la soglia di separazione tra le connessioni deboli (linee tratteggiate) e le connessioni normali (linee normali). Tutte le connessioni inferiori a questo valore verranno considerate deboli. Utilizzare il dispositivo di scorrimento per modificare il valore o immettere un numero nel campo.

Dopo aver modificato le soglie per un grafico Web, è possibile ripianificare o ridisegnare la visualizzazione Web con i nuovi valori di soglia tramite il menu Web posto nella barra degli strumenti del grafico Web. Dopo aver individuato le impostazioni che rivelano gli schemi più significativi, è possibile aggiornare le impostazioni originali nel nodo Web (denominato anche nodo Web genitore) scegliendo **Aggiorna nodo genitore** dal menu Web nella finestra del grafico.

Creazione di un riepilogo Web

È possibile creare un documento di riepilogo Web in cui siano elencati i collegamenti forti, medi e deboli, facendo clic sul pulsante con due frecce gialle nella barra degli strumenti per espandere la finestra del grafico Web. Quindi fare clic sulla scheda Riepilogo per visualizzare le tabelle per ogni tipo di collegamento. Le tabelle possono essere espanse e compresse utilizzando i relativi pulsanti con le frecce.

Figura 5-53

Riepilogo Web contenente l'elenco delle connessioni tra la pressione, il colesterolo e il tipo di cura

Riepilogo		Controlli	
- Collegamenti forti			
Collegamenti	Campo 1	Campo 2	
47	Cholesterol = "HIGH"	Drug = "drugY"	
44	Cholesterol = "NORMAL"	Drug = "drugY"	
42	BP = "HIGH"	Cholesterol = "NORMAL"	
38	BP = "HIGH"	Drug = "drugY"	
37	BP = "NORMAL"	Cholesterol = "HIGH"	
36	BP = "NORMAL"	Drug = "drugX"	
- Collegamenti medi			
Collegamenti	Campo 1	Campo 2	
35	BP = "HIGH"	Cholesterol = "HIGH"	
34	Cholesterol = "NORMAL"	Drug = "drugX"	
33	BP = "LOW"	Cholesterol = "NORMAL"	
31	BP = "LOW"	Cholesterol = "HIGH"	
30	BP = "LOW"	Drug = "drugY"	
23	BP = "NORMAL"	Drug = "drugY"	
23	BP = "HIGH"	Drug = "drugA"	
22	BP = "NORMAL"	Cholesterol = "NORMAL"	
20	Cholesterol = "HIGH"	Drug = "drugX"	
18	BP = "LOW"	Drug = "drugX"	
16	BP = "LOW"	Drug = "drugC"	
16	Cholesterol = "HIGH"	Drug = "drugC"	
16	BP = "HIGH"	Drug = "drugB"	
- Collegamenti deboli			
Collegamenti	Campo 1	Campo 2	
12	Cholesterol = "HIGH"	Drug = "drugA"	
11	Cholesterol = "NORMAL"	Drug = "drugA"	
8	Cholesterol = "HIGH"	Drug = "drugB"	
8	Cholesterol = "NORMAL"	Drug = "drugB"	

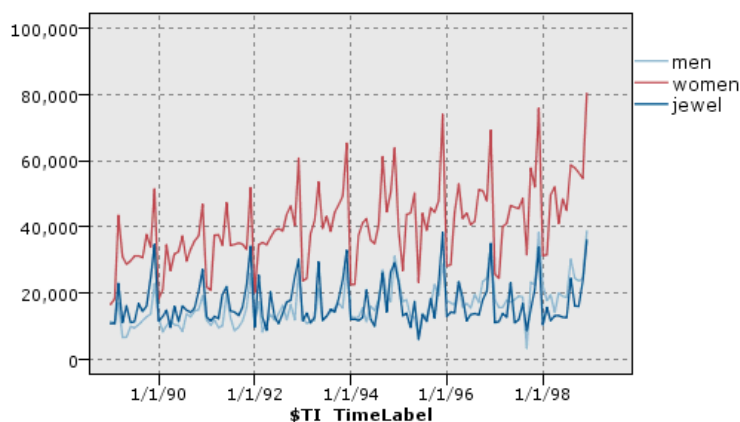
Per stampare i riepilogo, scegliere le seguenti opzioni dal menu nella finestra del grafico Web:
File > Stampa riepilogo

Nodo Plot tempo

I nodi Plot tempo consentono di visualizzare una o più serie storiche rappresentate nel tempo. Le serie rappresentate devono contenere valori numerici e si presuppone che ricorrano in un intervallo di tempo in cui i periodi sono uniformi. Il nodo Intervalli di tempo viene utilizzato in genere prima del nodo Plot tempo per creare un campo *EtichettaTempo*, utilizzato per default per identificare l'asse delle x nei grafici. [Per ulteriori informazioni, vedere l'argomento Nodo Intervalli di tempo in il capitolo 4 a pag. 204.](#)

Figura 5-54

Rappresentazione grafica delle vendite di abbigliamento e gioielli maschili e femminili nel tempo

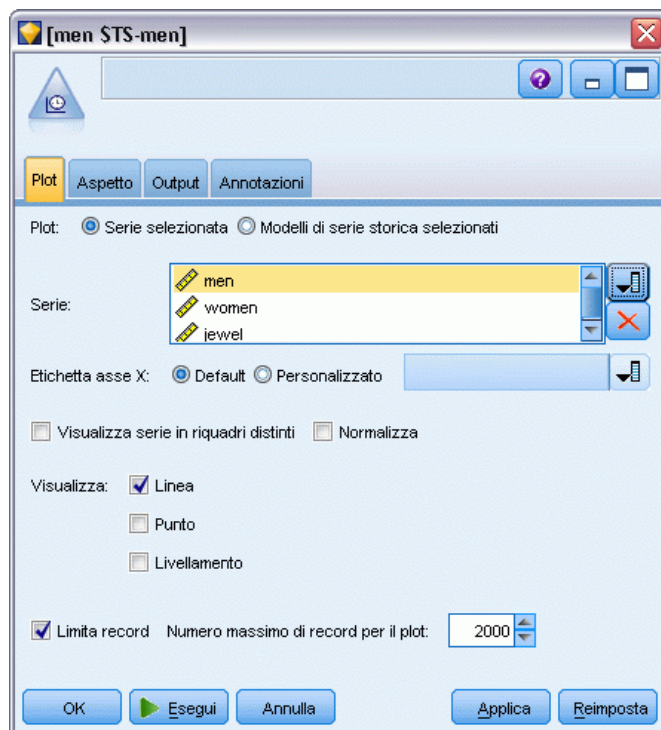


Creazione di interventi ed eventi

È possibile creare campi di evento e di intervento dal Plot tempo generando un nodo Nuovo campo (flag o nominale) dai menu di scelta rapida. Per esempio, si può creare un campo di evento per uno sciopero ferroviario, dove lo stato guida è Vero se l'evento si verifica e Falso in caso contrario. Nel caso di un campo di Intervento, per esempio per un aumento di prezzo, si può invece utilizzare un nodo Crea variabile conteggio per identificare la data dell'aumento, con il valore 0 per il prezzo vecchio e il valore 1 per quello nuovo. [Per ulteriori informazioni, vedere l'argomento Nuovo campo in il capitolo 4 a pag. 154.](#)

Scheda Plot del nodo Plot tempo

Figura 5-55
Impostazioni della scheda Plot per un nodo Plot tempo



Plot. Consente di scegliere il modo in cui rappresentare graficamente i dati di serie storica.

- **Serie selezionata.** Rappresenta i valori della serie storica selezionata. Se si seleziona questa opzione quando si rappresentano gli intervalli di confidenza, deselezionare la casella di controllo Normalizza.
- **Modelli di serie storica selezionati.** Utilizzata con un modello di serie storica, questa opzione rappresenta tutti i campi correlati (valori effettivi e previsti e intervalli di confidenza) per una o più serie storiche selezionate. La selezione di questa opzione comporta la disattivazione di alcune altre opzioni nella finestra di dialogo. Si tratta dell'opzione ideale per la rappresentazione degli intervalli di confidenza.

Serie. Selezionare uno o più campi contenenti i dati di serie storica da rappresentare nel grafico. I dati devono essere numerici.

Etichetta asse X. Scegliere l'etichetta di default o un singolo campo da utilizzare come etichetta dell'asse x nelle rappresentazioni grafiche. Se si seleziona l'opzione Default, il sistema utilizza il campo EtichettaTempo creato a partire da un nodo Intervalli di tempo situato a monte o, in assenza del nodo Intervalli di tempo, da numeri interi sequenziali. [Per ulteriori informazioni, vedere l'argomento Nodo Intervalli di tempo in il capitolo 4 a pag. 204.](#)

Visualizza serie in riquadri distinti. Consente di specificare se le singole serie devono essere visualizzate in riquadri distinti. In alternativa, se si decide di non utilizzare i riquadri, tutte le serie storiche vengono rappresentate sullo stesso grafico e le funzioni di livellamento non sono

disponibili. Quando si rappresentano tutte le serie storiche sullo stesso grafico, ogni serie è contraddistinta da un colore diverso.

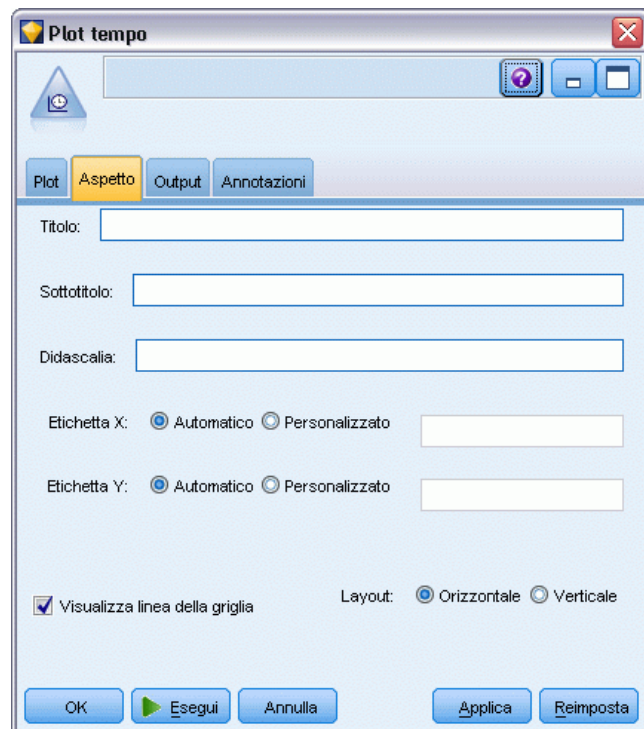
Normalizza. Selezionare questa opzione se si desidera graduare tutti i valori Y nell'intervallo 0–1 per la visualizzazione nel grafico. La normalizzazione consente di visualizzare in modo chiaro la relazione tra linee che potrebbero risultare non completamente visibili a causa di differenze nell'intervallo dei valori di ogni serie ed è consigliata quando si rappresentano più linee nello stesso grafico o quando si confrontano plot in riquadri affiancati (la normalizzazione non è necessaria quando tutti i valori di dati rientrano in un intervallo simile).

Visualizzazione. Seleziona uno o più elementi da visualizzare nella rappresentazione grafica. È possibile scegliere tra linee, punti e livellamenti (LOESS). I livellamenti sono disponibili solo se le serie vengono visualizzate in riquadri distinti. Per default è selezionato l'elemento linea. Verificare di avere selezionato almeno un elemento da rappresentare prima di eseguire il nodo Grafico. In caso contrario, il sistema restituirà un errore segnalando che non sono stati selezionati elementi da rappresentare.

Limita record. Selezionare questa opzione per limitare il numero dei record da rappresentare nel grafico. Specificare il numero di record (letti dall'inizio del file di dati) che dovranno essere rappresentati nell'opzione Numero massimo di record per il plot. Per default, questo numero è impostato su 2.000. Per rappresentare gli ultimi n record del file di dati, è possibile utilizzare un nodo Ordina prima di questo nodo per ordinare i record in ordine decrescente in base all'ora.

Scheda Aspetto del nodo Plot tempo

Figura 5-56
Impostazioni della scheda Aspetto per un nodo Plot tempo



Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Didascalia. Immettere il testo da utilizzare come didascalia del grafico.

Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

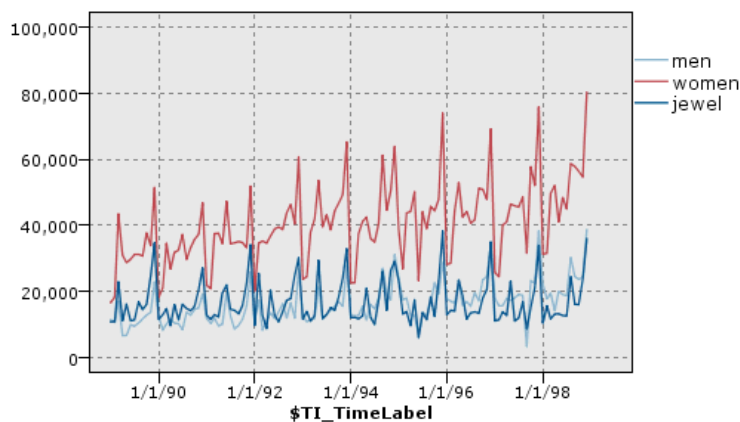
Layout. Solo per i Plot tempo, è possibile specificare se i valori temporali devono essere rappresentati su un asse orizzontale o verticale.

Utilizzo del grafico Plot tempo

Dopo aver creato il grafico Plot tempo è possibile utilizzare diverse opzioni per modificare la visualizzazione del grafico e generare nodi per ulteriori analisi. [Per ulteriori informazioni, vedere l'argomento Esplorazione dei grafici a pag. 309.](#)

Figura 5-57

Rappresentazione grafica delle vendite di abbigliamento e gioielli maschili e femminili nel tempo



Dopo aver creato un Plot tempo, definito le sezioni ed esaminato i risultati, è possibile utilizzare le opzioni del menu Genera e il menu di scelta rapida per creare nodi Seleziona o Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Nodo Valutazione

Il nodo Valutazione consente di valutare e mettere a confronto modelli predittivi per scegliere il miglior modello per l'applicazione da utilizzare. I grafici prodotti dal nodo Valutazione (o grafici di valutazione) mostrano il comportamento dei modelli nella previsione di particolari risultati. Questi grafici ordinano i record in base al valore previsto e alla confidenza della previsione, quindi li suddividono in gruppi di uguale dimensione (**quantili**) e infine rappresentano il valore del criterio commerciale per ciascun quantile, dal più alto al più basso. I modelli multipli sono mostrati nel grafico come linee separate.

I risultati vengono gestiti definendo un valore specifico o un intervallo di valori come **risultato positivo**. I risultati positivi in genere indicano un qualche successo (ad esempio una vendita a un cliente) o un evento significativo (ad esempio una diagnosi medica specifica). Per definire i criteri dei risultati positivi, utilizzare la scheda Opzioni della finestra di dialogo oppure utilizzare i criteri di default, come segue:

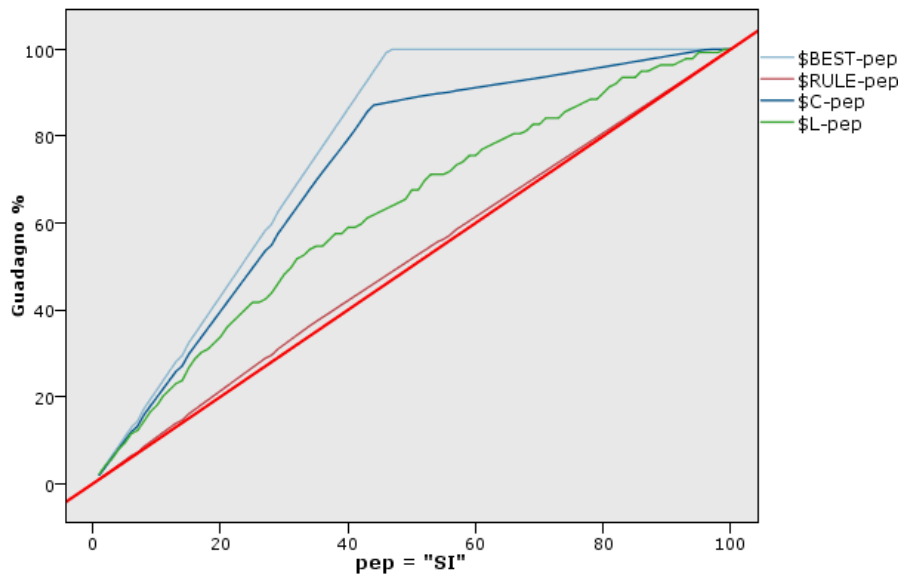
- I campi di output **flag** sono di semplice interpretazione; infatti, i risultati positivi corrispondono a valori *veri*.
- Nei campi di output **nominali** il primo valore dell'insieme definisce un risultato positivo.
- Nei campi di output **continui** i risultati positivi equivalgono ai valori superiori al valore intermedio dell'intervallo del campo.

Sono disponibili cinque tipi di grafici di valutazione, ognuno dei quali enfatizza un diverso criterio di valutazione.

Grafici dei guadagni

I guadagni sono rappresentati dalla percentuale di risultati positivi complessivi in ogni quantile. I guadagni vengono calcolati dividendo il numero di risultati positivi nel quantile per il numero totale di risultati positivi e quindi moltiplicando il risultato ottenuto per 100%.

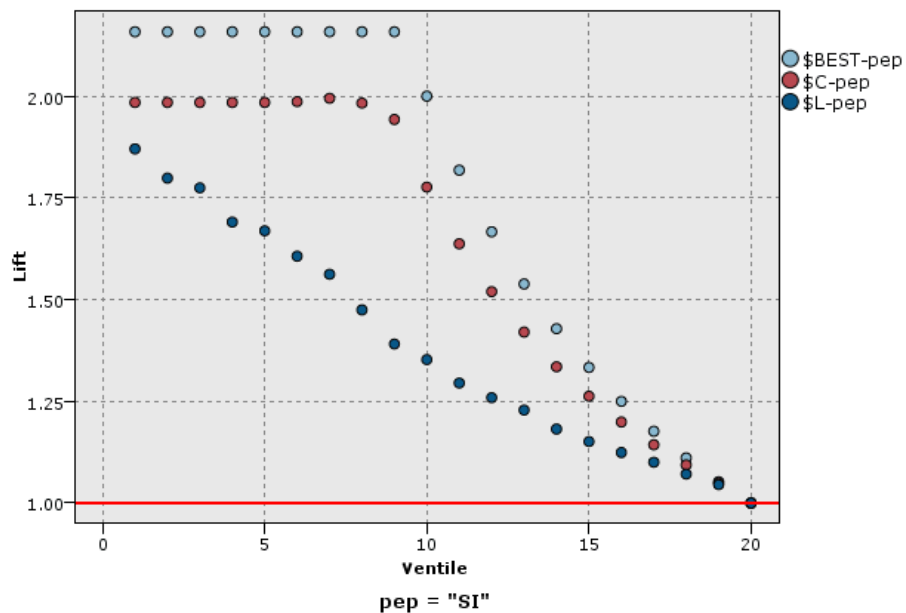
Figura 5-58
Grafico dei guadagni (cumulato) con linea base, linea migliore e regola aziendale



Grafici lift

Il grafico lift confronta la percentuale di record che in ogni quantile corrispondono a risultati positivi con la percentuale globale di risultati positivi nei dati di addestramento. La risposta viene calcolata dividendo i risultati positivi nel quantile per i record nel quantile e quindi dividendo il risultato ottenuto per il risultato della divisione dei risultati positivi totali per i record totali.

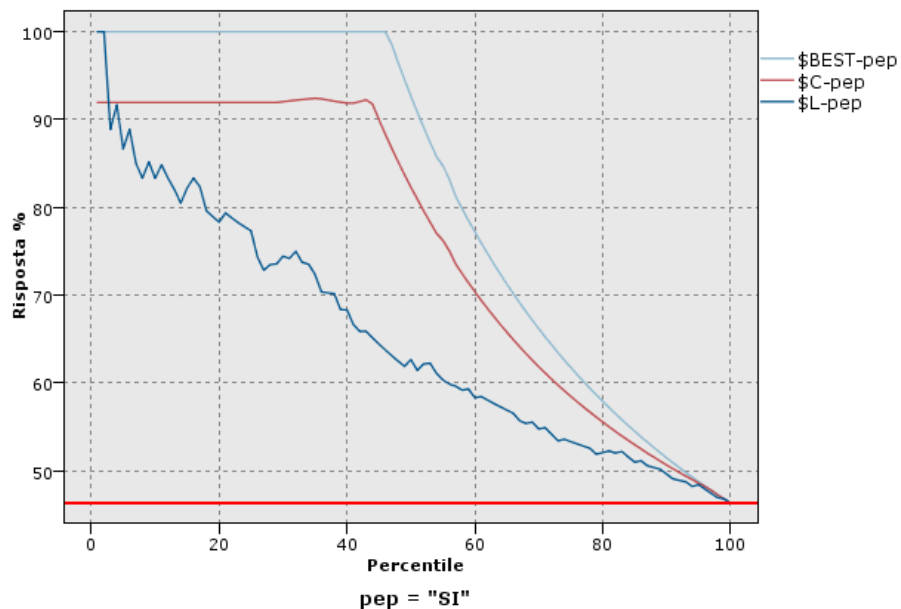
Figura 5-59
Grafico lift (cumulato) con l'utilizzo di punti e linea migliore



Grafici di risposta

La risposta è rappresentata dalla percentuale di record nel quantile corrispondenti a valori positivi. La risposta viene calcolata dividendo i valori positivi nel quantile per i record nel quantile e quindi moltiplicando il risultato ottenuto per 100%.

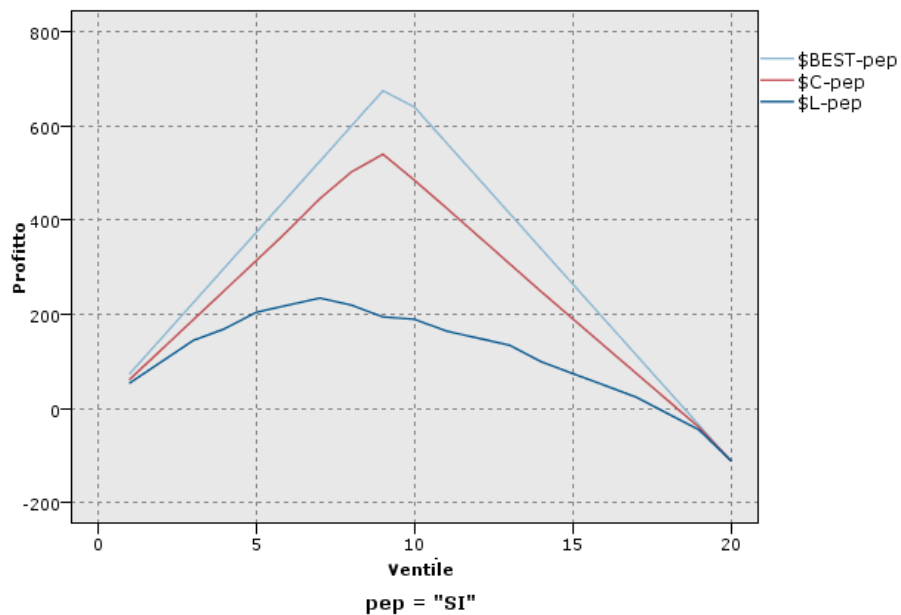
Figura 5-60
Grafico delle risposte (cumulato) con linea migliore



Grafici di profitto

Il profitto equivale alle **entrate** relative a ogni record meno il **costo** per il record. I profitti di un quantile sono semplicemente la somma dei profitti di tutti i record nel quantile. Si presuppone che le entrate vengano applicate solo ai risultati positivi, mentre i costi vengono applicati a tutti i record. Profitti e costi possono essere fissi oppure definiti da campi nei dati. I profitti vengono calcolati sottraendo alla somma delle entrate dei record nel quantile la somma dei costi dei record nel quantile.

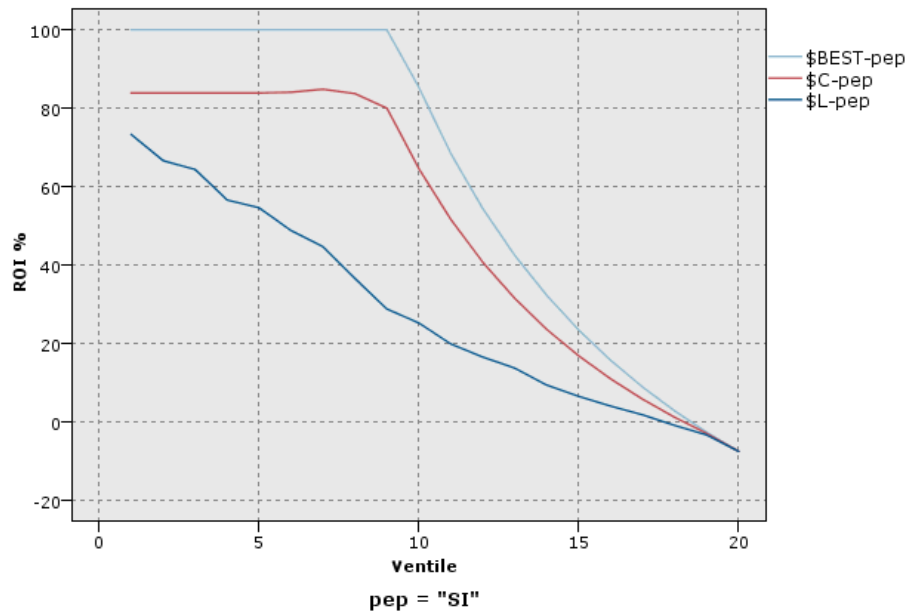
Figura 5-61
Grafico dei profitti (cumulato) con linea migliore



Grafici ROI

Il valore ROI (Return On Investment) è simile al profitto in quanto implica la definizione di entrate e costi. Il ROI confronta i profitti con i costi per il quantile. Il ROI viene calcolato dividendo i profitti del quantile per i costi del quantile e quindi moltiplicando il risultato ottenuto per 100%.

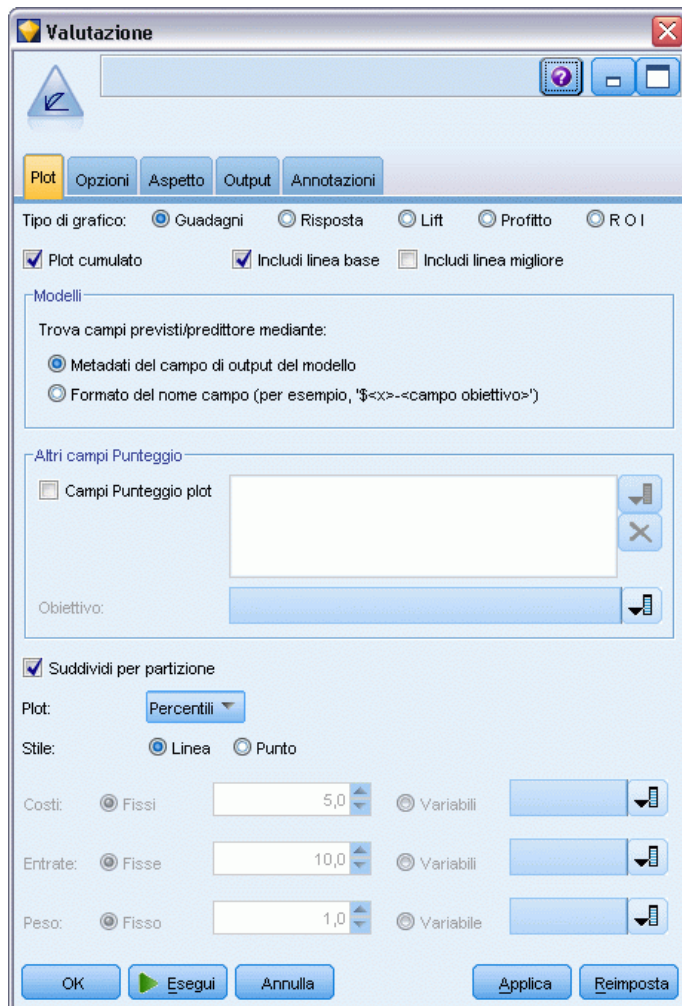
Figura 5-62
Grafico ROI (cumulato) con linea miglior



I grafici di valutazione possono essere anche cumulati, per cui ogni punto equivale al valore del quantile corrispondente più tutti i quantili superiori. I grafici cumulati in genere offrono una rappresentazione migliore delle performance globali dei modelli, mentre i grafici non cumulati risultano spesso più efficaci per mettere in evidenza particolari aree problematiche dei modelli.

Scheda Plot del nodo Valutazione

Figura 5-63
Impostazioni della scheda Plot per un nodo Valutazione



Tipo di grafico. Selezionare uno dei seguenti tipi: Guadagni, Risposta, Lift, Profitto o ROI (Return On Investment).

Plot cumulato. Selezionare questa opzione per creare un grafico cumulato. I valori presenti nei grafici cumulati vengono rappresentati per ogni quantile più tutti i quantili superiori.

Includi linea base. Selezionare questa opzione per includere una linea base nel grafico, che indica una distribuzione casuale dei risultati positivi quando la confidenza diventa irrilevante (Includi linea base non è disponibile per i grafici Profitto e ROI).

Includi linea migliore. Selezionare questa opzione per includere una linea migliore nel grafico, che indica una confidenza perfetta (i risultati sono pari al 100% dei casi).

Trova campi previsti/predittore mediante. Selezionare Metadati del campo di output del modello per cercare i campi previsti nel grafico utilizzando i relativi metadati o selezionare Formato del nome campo per eseguire la ricerca in base ai nomi.

Campi Punteggio plot. Selezionare questa casella di controllo per attivare la funzione di selezione dei campi punteggio. Selezionare quindi uno o più campi punteggio intervallo o continui, ovvero campi che non sono modelli predittivi in senso stretto ma che potrebbero essere utili per classificare i record a seconda della loro propensione a essere risultati positivi. Il nodo Valutazione può confrontare qualunque combinazione di uno o più campi punteggio con uno o più modelli predittivi. Un esempio tipico potrebbe essere il confronto di vari campi RFM con il modello predittivo migliore.

Obiettivo. Selezionare il campo obiettivo mediante la funzione di selezione dei campi. Scegliere tutti i campi flag o nominali istanziati con due o più valori.

Nota: questo campo obiettivo è valido solo per i campi punteggio (i modelli predittivi definiscono da soli i propri obiettivi) e viene ignorato se nella scheda Opzioni viene impostato un criterio personalizzato per i risultati positivi.

Suddividi per partizione. Se si utilizza un campo di partizione per suddividere i record in campioni di addestramento, di test e di validazione, selezionare questa opzione per visualizzare un grafico di valutazione distinto per ogni partizione. [Per ulteriori informazioni, vedere l'argomento Nodo Partizione in il capitolo 4 a pag. 193.](#)

Nota: Se la suddivisione è in base alla partizione, i record con valori nulli nel campo di partizione saranno esclusi dalla valutazione. Ciò non rappresenta un problema se si utilizza un nodo Partizione, poiché i nodi Partizione non generano valori nulli.

Grafico. Selezionare dall'elenco a discesa la dimensione dei quantili da rappresentare nel grafico. Le opzioni includono Quartili, Quintili, Decili, Ventili, Percentili e 1000-tili.

Stile. Selezionare Linea o Punto.

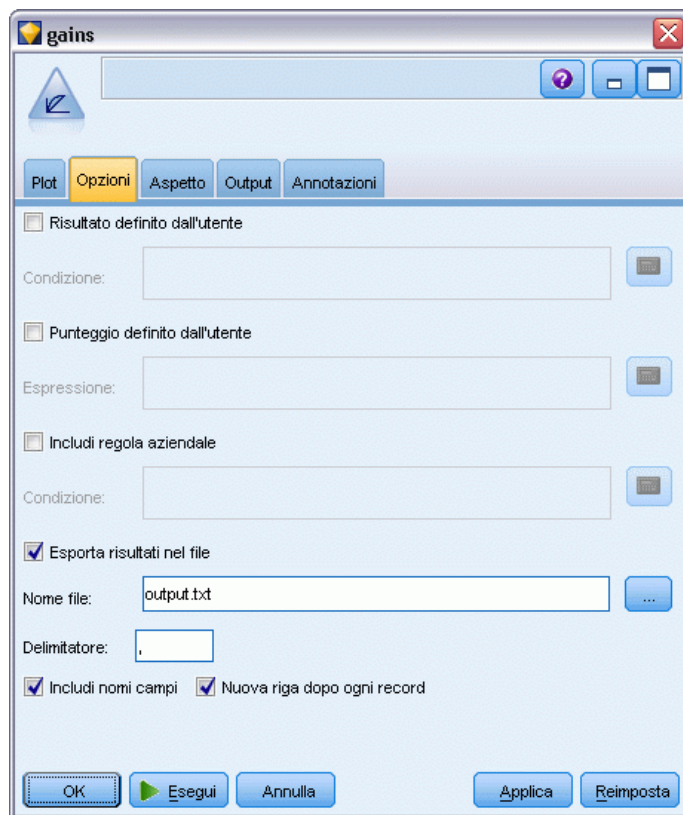
Grafici Profitto e ROI. Per i grafici Profitto e ROI sono disponibili controlli aggiuntivi che consentono di specificare costi, entrate e pesi.

- **Costi.** Specificare il costo associato a ciascun record. È possibile selezionare un costo Fisso o Variabile. Per i costi fissi, specificare il valore del costo. Per i costi variabili, fare clic su Selezione campi per selezionare un campo come campo dei costi.
- **Entrate.** Specificare le entrate associate a ogni record che rappresenta un risultato positivo. È possibile selezionare un costo Fisso o Variabile. Per entrate fisse, specificare il valore delle entrate. Per entrate variabili, fare clic su Selezione campi per selezionare un campo come campo delle entrate.
- **Peso.** Se i record nei dati rappresentano più di una unità, è possibile utilizzare i pesi per modificare i risultati. Specificare il peso associato a ciascun record, utilizzando i pesi Fisso o Variabile. Per pesi fissi, specificare il valore del peso (numero di unità per record). Per pesi variabili, fare clic sul pulsante di selezione dei campi per selezionare un campo come campo dei pesi.

Scheda Opzioni del nodo Valutazione

La scheda Opzioni per i grafici di valutazione garantisce la flessibilità nella definizione dei risultati, dei criteri di calcolo del punteggio e delle regole aziendali visualizzati nel grafico. È anche possibile impostare opzioni aggiuntive per l'esportazione dei risultati della valutazione del modello.

Figura 5-64
Impostazioni della scheda Opzioni per un nodo Valutazione



Risultato definito dall'utente. Selezionare questa opzione per specificare una condizione personalizzata da utilizzare per indicare un risultato. Questa opzione è utile per definire il risultato desiderato anziché dedurlo dal tipo di campo obiettivo e dall'ordine dei valori.

- **Condizione.** Quando è stato selezionato Risultato definito dall'utente, è necessario specificare un'espressione CLEM per una condizione di risultato. Per esempio, @TARGET = "Sì" è una condizione valida che indica che un valore Sì per il campo obiettivo verrà considerato come un risultato positivo nella valutazione. La condizione specificata verrà utilizzata per tutti i campi obiettivo. Per creare una condizione, digitare direttamente nel campo o utilizzare il Generatore di espressioni per generare un'espressione di condizione. Se i dati sono istanziati, è possibile immettere valori direttamente dal Generatore espressioni.

Punteggio definito dall'utente. Selezionare questa opzione per specificare una condizione utilizzata per calcolare il punteggio dei casi prima di assegnarli ai quantili. Il punteggio di default viene calcolato in base al valore previsto e alla confidenza. Utilizzare il campo Espressione per creare un'espressione di punteggio personalizzata.

- **Espressione.** Specificare un'espressione CLEM utilizzata per il calcolo del punteggio. Per esempio, se un output numerico nell'intervallo 0–1 viene ordinato in modo che i valori più bassi abbiano la priorità su quelli più alti, è possibile definire un risultato come @TARGET < 0.5 e il punteggio associato come 1 • @PREDICTED. L'espressione del punteggio deve generare un valore numerico. Per creare una condizione, digitare direttamente nel campo o utilizzare il Generatore di espressioni per generare un'espressione di condizione.

Includi regola aziendale. Selezionare questa opzione per specificare una condizione di regola che rifletta i criteri desiderati. Per esempio, è possibile visualizzare una regola per tutti i casi in cui mutuo = "Y" e reddito \geq 33000. Le regole aziendali vengono tracciate sul grafico ed etichettate nella chiave come *Regola*.

- **Condizione.** Specificare un'espressione CLEM utilizzata per definire una regola aziendale nel grafico di output. Digitare direttamente nel campo o utilizzare il Generatore di espressioni per generare un'espressione di condizione. Se i dati sono istanziati, è possibile immettere valori direttamente dal Generatore espressioni.

Esporta risultati nel file. Selezionare questa opzione per esportare i risultati della valutazione del modello in un file di testo delimitato. È possibile leggere questo file per eseguire analisi specializzate sui valori calcolati. Impostare le seguenti opzioni per l'esportazione:

- **Nome file.** Immettere il nome del file di output. Utilizzare il pulsante con i puntini di sospensione (...) per passare alla cartella desiderata.
- **Delimitatore.** Immettere un carattere, ad esempio una virgola o uno spazio, da utilizzare come separatore di campo.

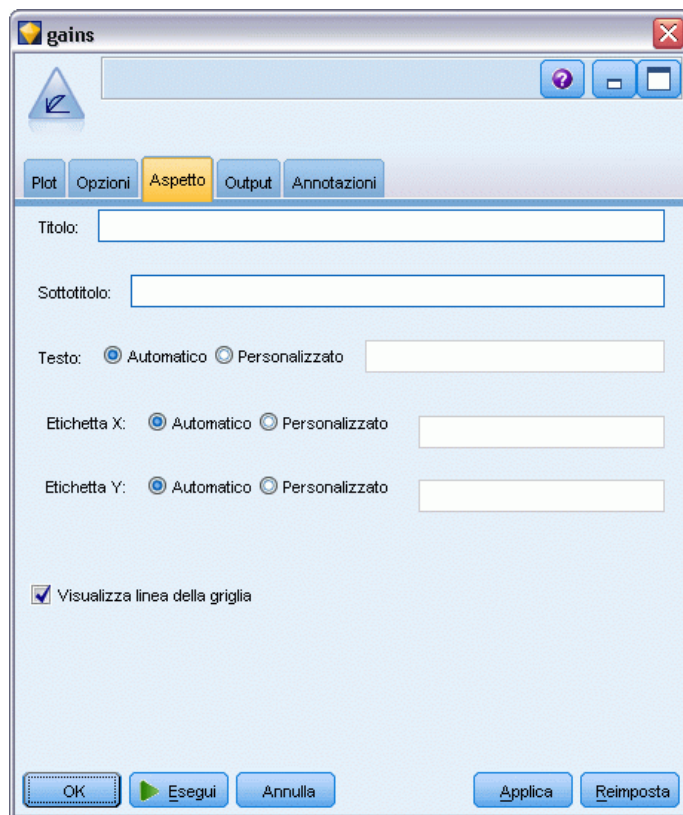
Includi nomi campi. Selezionare questa opzione per includere i nomi dei campi come prima riga del file di output.

Nuova riga dopo ogni record. Selezionare questa opzione se si desidera che ogni record inizi in una nuova riga.

Scheda Aspetto del nodo Valutazione

Prima di creare un grafico è possibile specificarne le opzioni dell'aspetto.

Figura 5-65
Impostazioni della scheda Aspetto per un nodo Valutazione



Titolo. Immettere il testo da utilizzare come titolo del grafico.

Sottotitolo. Immettere il testo da utilizzare come sottotitolo del grafico.

Testo. Accettare l'etichetta di testo generata automaticamente o selezionare Personalizzato per specificare un'etichetta.

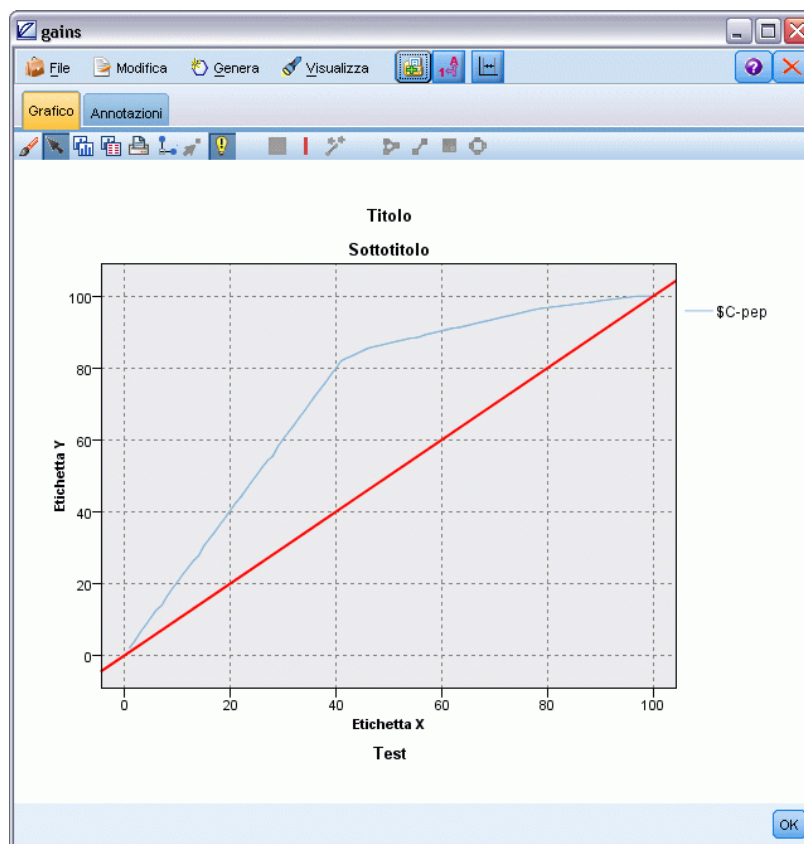
Etichetta X. Accettare l'etichetta dell'asse x (orizzontale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Etichetta Y. Accettare l'etichetta dell'asse y (verticale) generata automaticamente oppure selezionare Personalizzato per specificare un'etichetta personalizzata.

Visualizza griglia. Questa opzione è selezionata per default e consente di visualizzare una griglia dietro il grafico o il grafico plot per stabilire più facilmente i punti di interruzione di aree e sezioni. Le griglie sono sempre visualizzate in bianco. Sono di colore grigio solo se anche lo sfondo del grafico è bianco.

Il seguente esempio mostra dove vengono posizionate le opzioni della scheda Aspetto sul grafico.

Figura 5-66
Posizione delle opzioni della scheda Aspetto in un grafico di valutazione



Letture dei risultati di un grafico di valutazione relativo a un modello

L'interpretazione di un grafico di valutazione dipende entro certi limiti dal tipo di grafico, tuttavia alcune caratteristiche sono comuni a tutti i grafici di valutazione. Nel caso dei grafici cumulati, le linee superiori indicano i modelli migliori, in particolare nella parte sinistra del grafico. In molti casi, quando si confrontano più modelli le linee si incrociano, pertanto un modello risulterà più alto in una parte del grafico e un altro risulterà più alto in un'altra parte del grafico. In questi casi è necessario definire la parte del campione desiderata (la parte che individua un punto sull'asse x) quando si sceglie il modello da utilizzare.

La maggior parte dei grafici non cumulati sarà molto simile. Per modelli efficaci, i grafici non cumulati devono essere alti nella parte sinistra del grafico e bassi nella parte destra (se un grafico non cumulato presenta un andamento a denti di sega, è possibile livellarlo riducendo il numero di quantili da rappresentare ed eseguendo nuovamente il grafico). La presenza di depressioni sulla sinistra del grafico o di picchi sulla destra può indicare aree in cui il modello ha una scarsa efficacia predittiva. Una linea piatta che attraversa tutto il grafico indica un modello che non fornisce informazioni utili.

Grafici dei guadagni. I grafici dei guadagni cumulati iniziano sempre nella parte sinistra allo 0% e terminano nella parte destra al 100%. Per un modello attendibile, il grafico dei guadagni aumenterà ripidamente verso il 100% per poi scendere. Un modello che non fornisce informazioni seguirà la diagonale che va dalla parte inferiore sinistra alla parte superiore destra (tale diagonale viene visualizzata nel grafico se l'opzione *Includi linea base* è stata selezionata).

Grafici lift. I grafici lift cumulati tendono a iniziare nella parte sinistra al di sopra di 1,0 e quindi decrescono gradualmente verso destra fino a tornare a 1,0. Il bordo destro del grafico rappresenta l'intero insieme dei dati, pertanto il rapporto dei risultati positivi nei quantili cumulati rispetto ai risultati positivi nei dati è 1,0. Il lift di un modello efficace deve iniziare nella parte sinistra all'altezza di un valore molto più alto di 1,0, quindi rimanere a un livello elevato e infine decrescere rapidamente nella parte destra del grafico fino a 1,0. Per un modello che non fornisce informazioni, la linea si manterrà attorno a 1,0 per l'intero grafico (se è selezionata l'opzione *Includi linea base*, nel grafico viene mostrata come riferimento una linea orizzontale posizionata all'altezza di 1,0).

Grafici delle risposte. I grafici delle risposte cumulati tendono a essere molto simili ai grafici lift, tranne per la graduazione. I grafici delle risposte in genere iniziano a un livello vicino al 100% e scendono gradualmente nella parte destra del grafico fino a raggiungere il tasso di risposta globale (ovvero il risultato della divisione dei risultati positivi totali per i record totali). La linea di un modello efficace inizierà nella parte sinistra a un livello vicino o coincidente con 100%, si manterrà a un livello elevato e quindi decrescerà rapidamente nella parte destra del grafico fino a raggiungere il tasso di risposta globale. Per un modello che non fornisce informazioni, la linea si manterrà attorno al tasso di risposta globale per l'intero grafico (se è selezionata l'opzione *Includi linea base*, nel grafico viene mostrata come riferimento una linea orizzontale in corrispondenza del tasso di risposta globale).

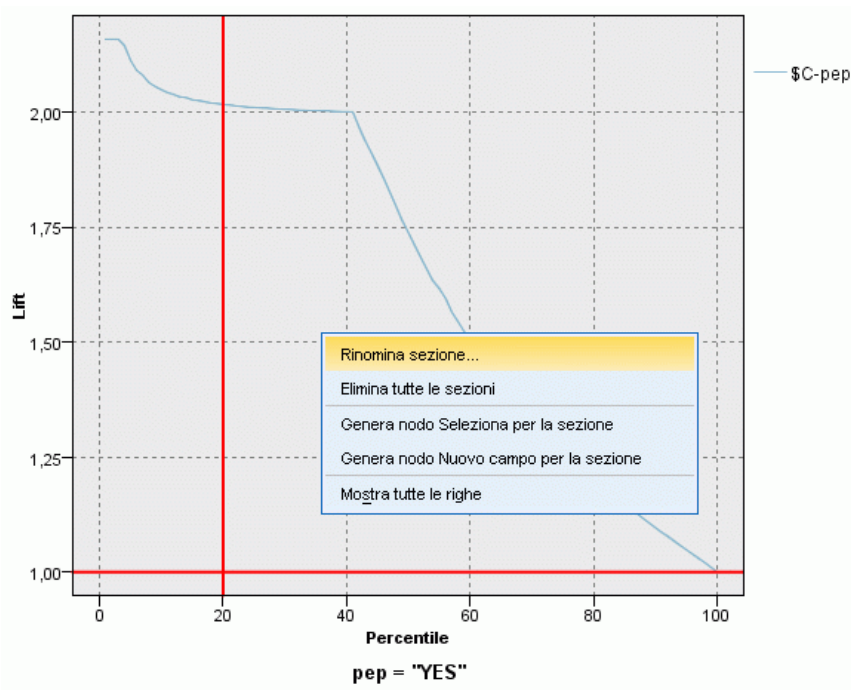
Grafici dei profitti. I grafici dei profitti cumulati mostrano la somma dei profitti man mano che viene aumentata la dimensione del campione selezionato, procedendo da sinistra verso destra. I grafici dei profitti iniziano in genere vicino allo 0, aumentano costantemente man mano che ci si sposta verso destra fino a raggiungere un picco o un livello elevato costante al centro e poi scendono verso il bordo destro del grafico. I profitti di un modello efficace mostreranno un picco ben definito nella parte centrale del grafico. Per un modello che non fornisce informazioni, la linea sarà relativamente diritta e potrà seguire un andamento crescente, decrescente o piatto a seconda della struttura costo/entrate pertinente.

Grafici ROI. I grafici ROI (Return On Investment) cumulati tendono a essere simili ai grafici delle risposte e ai grafici lift, tranne per la graduazione. I grafici ROI in genere iniziano a un livello superiore allo 0% e decrescono gradualmente fino a raggiungere il ROI globale per l'intero insieme di dati (che può essere negativo). La linea di un modello efficace deve iniziare a un livello molto più alto dello 0%, mantenersi a un livello elevato e quindi decrescere abbastanza rapidamente nella parte destra del grafico verso il ROI globale. Per un modello che non fornisce informazioni, la linea si manterrà attorno al valore del ROI globale.

Utilizzo di un grafico di valutazione

L'utilizzo del mouse per esplorare un grafico di valutazione è simile all'utilizzo di un istogramma o di un grafico raccolta. L'asse x rappresenta i punteggi del modello dei quantili specificati, ad esempio ventili o decili.

Figura 5-67
Utilizzo di un grafico di valutazione

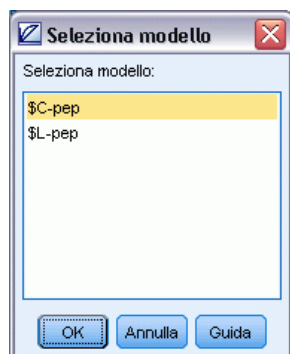


È possibile suddividere l'asse x in sezioni in modo analogo a un istogramma utilizzando l'icona di divisione che fornisce opzioni per la suddivisione automatica dell'asse in sezioni uguali. Per ulteriori informazioni, vedere l'argomento [Esplorazione dei grafici](#) a pag. 309. Per modificare manualmente i limiti delle sezioni, selezionare Sezioni grafico dal menu Modifica.

Dopo aver creato un grafico di valutazione, definito sezioni ed esaminato i risultati, è possibile utilizzare le opzioni del menu Genera e il menu di scelta rapida per creare automaticamente nodi basati sulle selezioni nel grafico. Per ulteriori informazioni, vedere l'argomento [Generazioni di nodi da grafici](#) a pag. 319.

Quando si generano nodi da un grafico di valutazione, viene richiesto di selezionare un singolo modello tra tutti quelli disponibili nel grafico.

Figura 5-68
Selezione di un modello per la generazione del nodo



Selezionare un modello e fare clic su OK per generare il nuovo nodo nell'area di disegno dello stream.

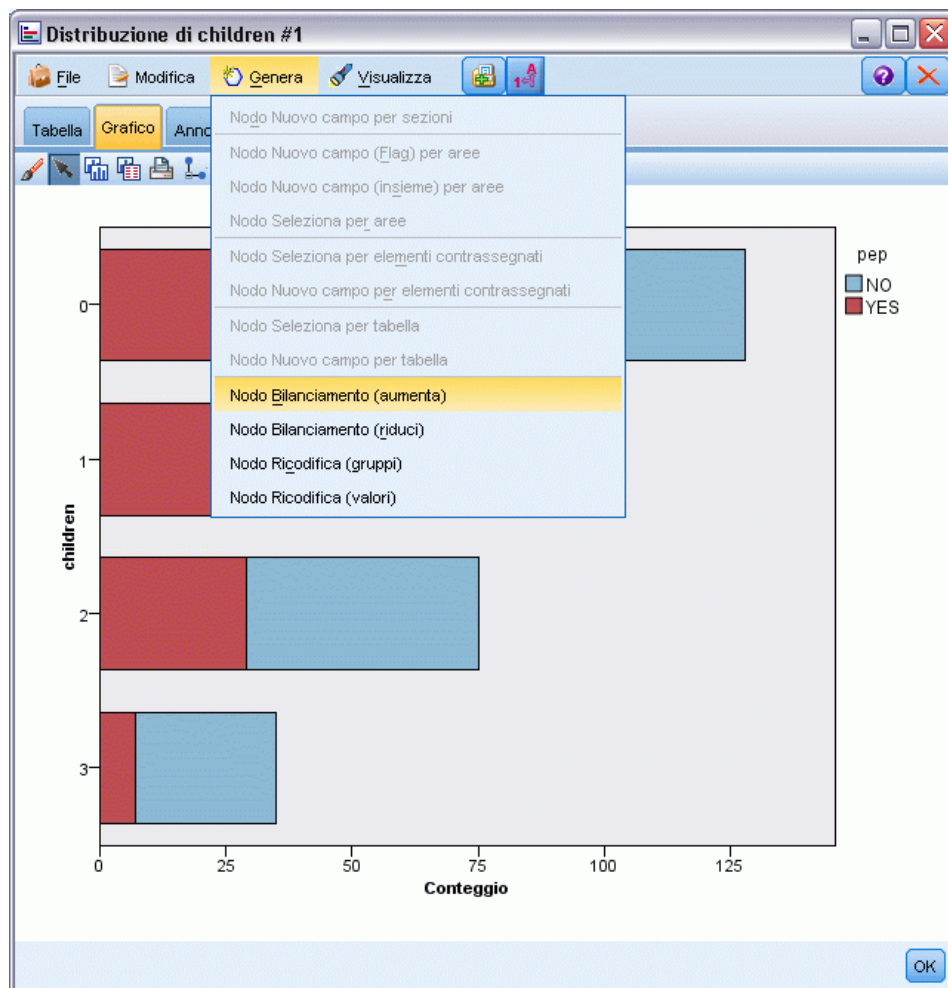
Esplorazione dei grafici

Mentre la modalità Modifica consente di modificare layout e aspetto del grafico, la modalità Esplora consente di esplorare analiticamente dati e valori rappresentati dal grafico. Lo scopo principale dall'esplorazione è l'analisi dei dati e la successiva identificazione di valori tramite l'utilizzo di sezioni, aree ed elementi contrassegnati per generare nodi Seleziona, Nuovo campo o Bilanciamento. Per selezionare questa modalità, scegliere Visualizza > Modalità Esplora dai menu (oppure fare clic sull'icona nella barra degli strumenti).

Mentre alcuni grafici supportano tutti gli strumenti di esplorazione, altri ne accettano uno solo. La modalità Esplora consente di:

- Definire e modificare sezioni, che vengono utilizzate per suddividere i valori lungo un asse x scala. [Per ulteriori informazioni, vedere l'argomento Utilizzo delle sezioni a pag. 310.](#)
- Definire e modificare aree, che vengono utilizzate per identificare un gruppo di valori all'interno di un'area rettangolare. [Per ulteriori informazioni, vedere l'argomento Utilizzo delle aree a pag. 314.](#)
- Aggiungere ed eliminare contrassegni sugli elementi per selezionare manualmente valori che potrebbero essere utilizzati per generare un nodo Seleziona o Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Utilizzo degli elementi contrassegnati a pag. 317.](#)
- Generare nodi utilizzando i valori identificati da sezioni, aree, elementi contrassegnati e collegamenti Web da utilizzare nello stream. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Figura 5-69
Grafico con il menu *Genera* visualizzato



Utilizzo delle sezioni

In tutti i grafici con un campo scala sull'asse x è possibile disegnare righe di sezione verticali per suddividere l'intervallo di valori sull'asse x . Se un grafico ha più pannelli, una riga di sezione tracciata in un pannello viene rappresentata anche negli altri pannelli.

Non tutti i grafici supportano le sezioni. Di seguito sono elencati alcuni di quelli che le supportano: istogrammi, grafici a barre e di distribuzione, grafici plot (a linee, a dispersione, tempo e così via), grafici di raccolta e di valutazione. Nei grafici con pannelli, le sezioni vengono visualizzate in tutti i pannelli. Inoltre, nei grafici SPLOM è visibile in alcuni casi una riga di sezione orizzontale, poiché l'asse su cui è stata tracciata la sezione di campo/variabile è stato capovolto.

Figura 5-70
Grafico con tre sezioni

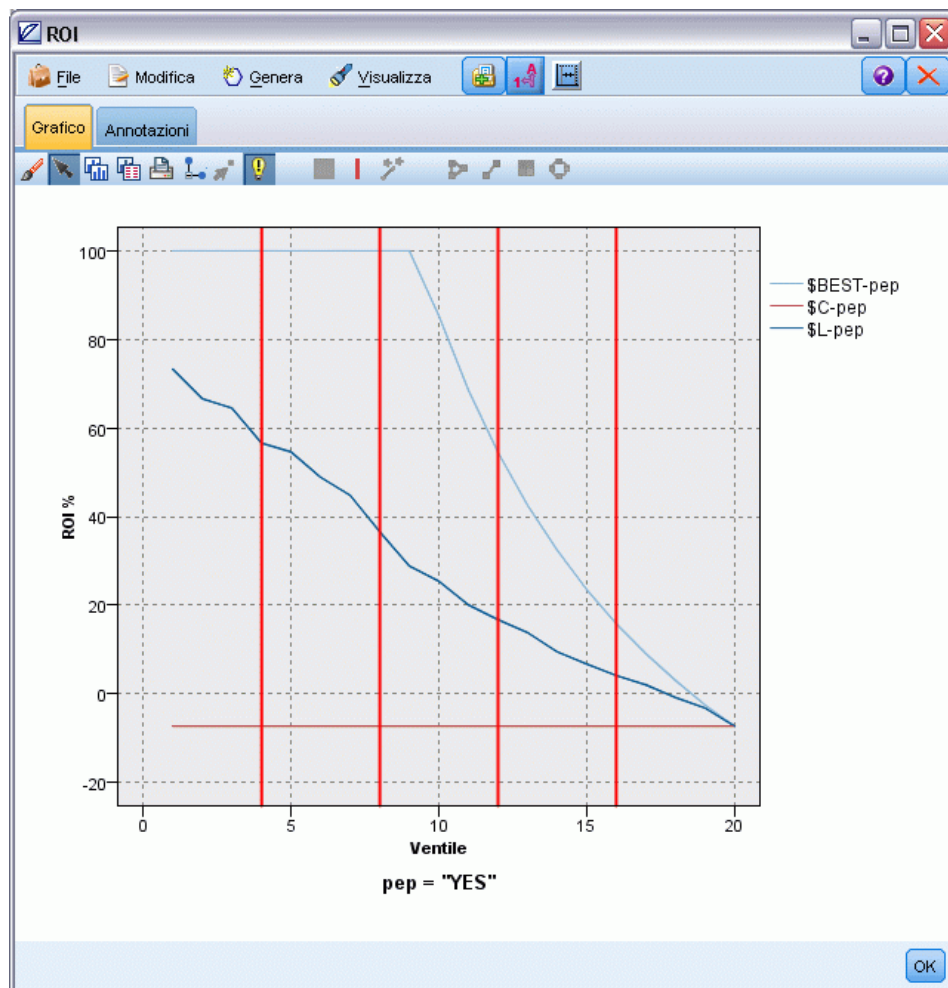
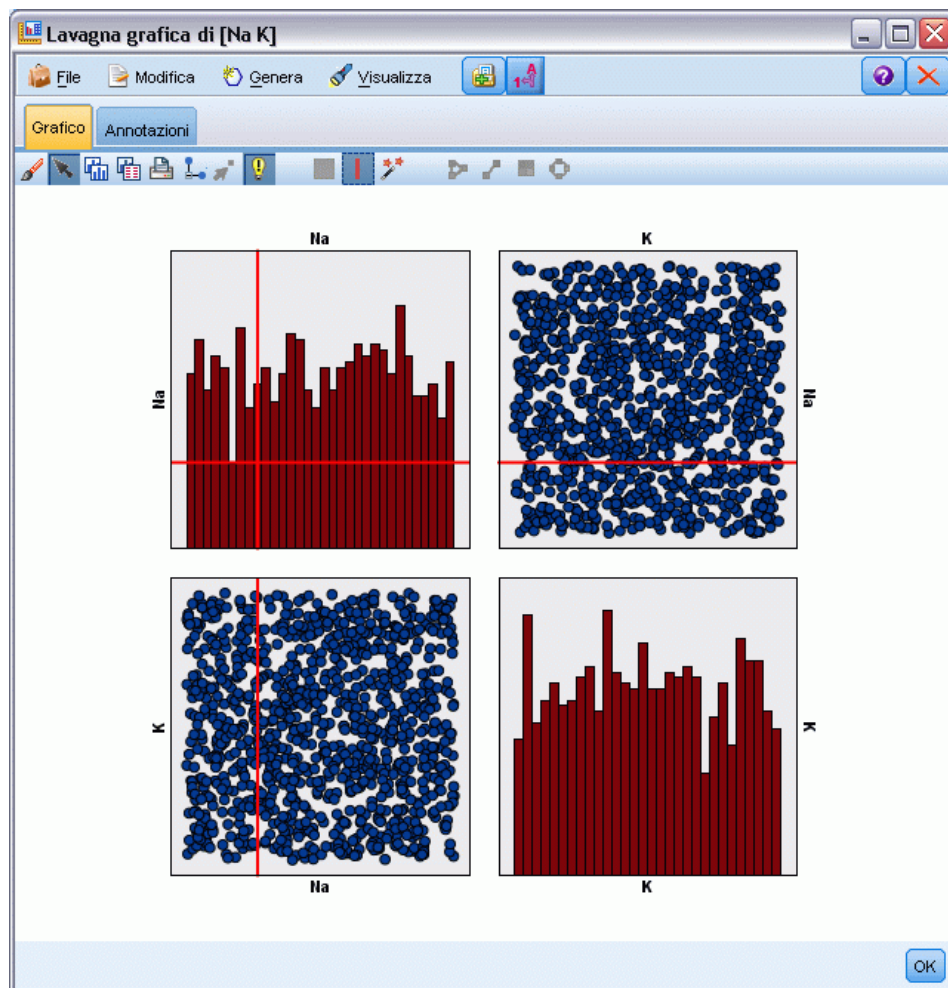


Figura 5-71
SPLOM con sezioni



Definizione delle sezioni

In un grafico privo di sezioni, aggiungendo una riga di sezione si suddivide il grafico in due sezioni. Il valore della riga di sezione rappresenta il punto di partenza, denominato anche limite inferiore, della seconda sezione leggendo il grafico da sinistra verso destra. Analogamente, in un grafico con due sezioni, l'aggiunta di una riga di sezione suddivide una sezione in due, generandone tre. Per default, le sezioni sono denominate *sezione N* , dove N corrisponde al numero di sezioni sull'asse x calcolate da sinistra verso destra.

Dopo aver definito una sezione, è possibile selezionarla e trascinarla per riposizionarla sull'asse x . Facendo clic con il pulsante destro del mouse all'interno di una sezione sono disponibili ulteriori scelte rapide che consentono di rinominare, eliminare e generare nodi per tale sezione.

Per definire sezioni:

- Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.

- Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Crea sezione.

Figura 5-72

Pulsante della barra degli strumenti Crea sezione



- In un grafico che supporta le sezioni, fare clic sul punto del valore dell'asse x in corrispondenza del quale si desidera definire una riga di sezione.

Nota: in alternativa, fare clic sull'icona della barra degli strumenti Suddividi il grafico in sezioni, immettere il numero di sezioni uguali desiderate e fare clic su Suddividi.

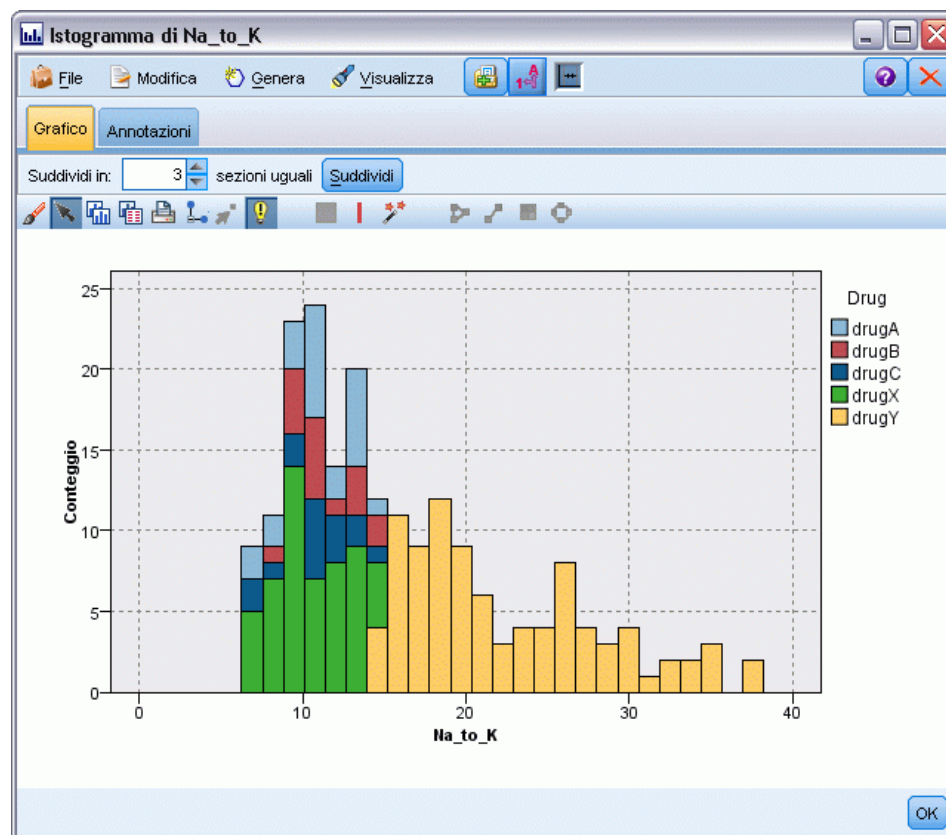
Figura 5-73

L'icona di divisione utilizzata per espandere la barra degli strumenti con opzioni relative alla suddivisione in sezioni



Figura 5-74

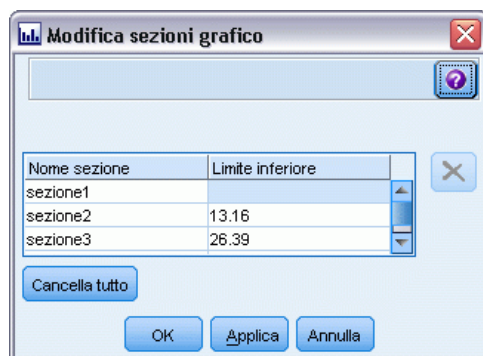
Creazione di sezioni uguali con le sezioni attivate



Modifica, ridenominazione ed eliminazione delle sezioni

È possibile modificare le proprietà di sezioni esistenti nella finestra di dialogo Modifica sezioni grafico oppure tramite menu di scelta rapida nel grafico stesso.

Figura 5-75
Finestra di dialogo Modifica sezioni grafico



Per modificare sezioni:

- ▶ Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.
- ▶ Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Crea sezione.
- ▶ Dai menu, scegliere Modifica > Sezioni grafico. Viene visualizzata la finestra di dialogo Modifica sezioni grafico.
- ▶ Se il grafico include più campi (per esempio i grafici SPLOM), è possibile selezionare il campo desiderato dall'elenco a discesa.
- ▶ Per aggiungere una nuova sezione, digitarne il nome e il limite inferiore. Premere quindi INVIO per iniziare una nuova riga.
- ▶ Per modificare il limite di una sezione, regolarne il Limite inferiore.
- ▶ Per rinominare una sezione, immettere un nuovo nome di sezione.
- ▶ Per eliminare una sezione, selezionare la riga corrispondente nella tabella e fare clic sul pulsante Elimina.
- ▶ Fare clic su OK per applicare le modifiche e chiudere la finestra di dialogo.

Nota: in alternativa, è possibile eliminare e rinominare sezioni direttamente nel grafico facendo clic con il pulsante destro del mouse sulla riga della sezione e scegliendo l'opzione desiderata dai menu di scelta rapida.

Utilizzo delle aree

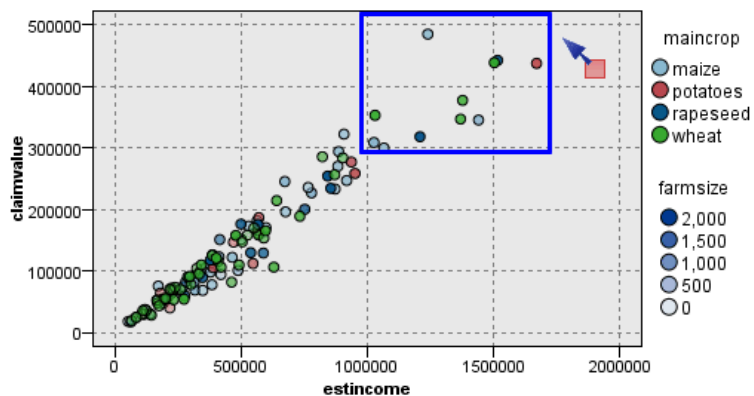
In tutti i grafici con due assi scala (o intervallo) è possibile disegnare aree per raggruppare valori all'interno di un'area rettangolare denominata appunto "area". Per **area** si intende una porzione del grafico delimitata da valori X e Y minimi e massimi. Se un grafico ha più pannelli, un'area tracciata in un pannello viene rappresentata anche negli altri pannelli.

Non tutti i grafici supportano le aree. Di seguito sono elencati alcuni di quelli che le supportano: grafici plot (a linee, a dispersione, a bolle, tempo e così via), SPLOM e grafici raccolta. Tali aree vengono disegnate nello spazio X, Y e non possono pertanto essere definite in grafici 1-D, 3-D o animati. Nei grafici con pannelli, le aree vengono visualizzate in tutti i pannelli. In un

grafico SPLOM (scatterplot matrix), un'area corrispondente verrà visualizzata nei plot superiori corrispondenti ma non in quelli diagonali poiché essi mostrano solo un campo scala.

Figura 5-76

Definizione di un'area di valori di domanda elevati



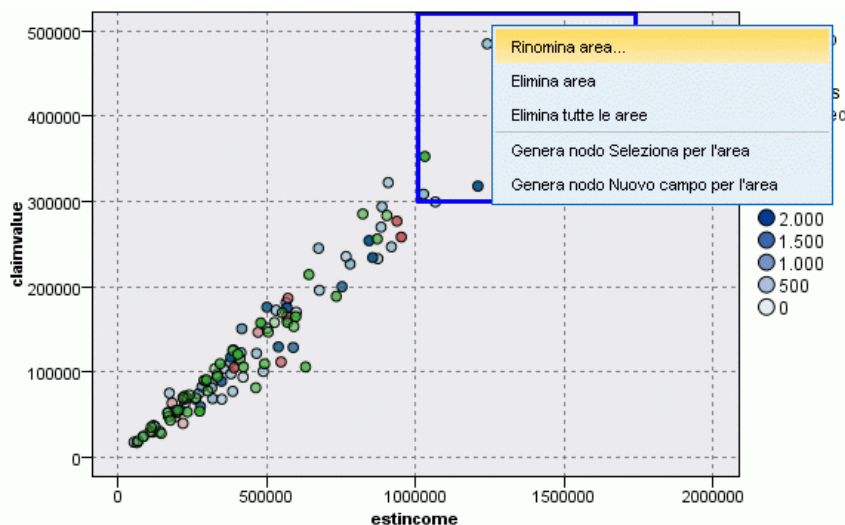
Definizione delle aree

Ogni volta che si definisce un'area, si crea un raggruppamento di valori. Per default, ogni nuova area viene denominata *Area<N>*, dove *N* corrisponde al numero di aree già create.

Dopo aver definito un'area, è possibile fare clic con il pulsante destro del mouse sulla riga dell'area per visualizzare alcune scelte rapide di base. Tuttavia, facendo clic con il pulsante destro del mouse all'interno dell'area (anziché sulla riga) sono disponibili ulteriori scelte rapide che consentono di rinominare, eliminare e generare nodi Seleziona e Nuovo campo per tale area.

È possibile selezionare sottoinsiemi di record in base alla loro inclusione in una particolare area o in un'area formata da più aree. È inoltre possibile incorporare le informazioni dell'area per un record creando un nodo Nuovo campo per generare un flag per i record in base all'inclusione in un'area. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Figura 5-77
Analisi dell'area contenente valori di domanda elevati



Per definire aree:

- ▶ Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.
- ▶ Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Crea area.

Figura 5-78
Pulsante della barra degli strumenti Crea area

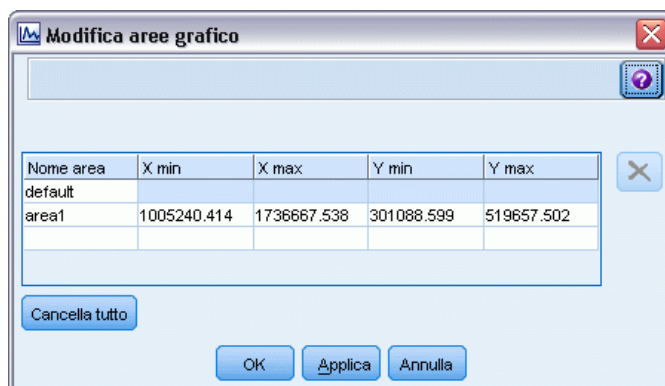


- ▶ In un grafico che supporta le aree, trascinare tenendo premuto il pulsante del mouse per disegnare un'area rettangolare.

Modifica, ridenominazione ed eliminazione delle aree

È possibile modificare proprietà di aree esistenti nella finestra di dialogo Modifica aree grafico oppure tramite menu di scelta rapida nel grafico stesso.

Figura 5-79
Impostazione delle proprietà relative alle aree definite



Per modificare aree:

- ▶ Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.
- ▶ Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Crea area.
- ▶ Dai menu, scegliere Modifica > Aree grafico. Viene visualizzata la finestra di dialogo Modifica aree grafico.
- ▶ Se il grafico include più campi (per esempio, grafici SPLOM), è necessario definire il campo per l'area nelle colonne *Campo A* e *Campo B*.
- ▶ Per aggiungere una nuova area su una nuova riga, digitarne il nome, selezionare eventuali nomi di campo e definire i limiti massimo e minimo per ciascun campo. Premere quindi INVIO per iniziare una nuova riga.
- ▶ Per modificare limiti di area esistenti, regolare i valori Min e Max per *A* e *B*.
- ▶ Per rinominare un'area, modificarne il nome nella tabella.
- ▶ Per eliminare un'area, selezionare la riga corrispondente nella tabella e fare clic sul pulsante Elimina.
- ▶ Fare clic su OK per applicare le modifiche e chiudere la finestra di dialogo.

Nota: in alternativa, è possibile eliminare e rinominare aree direttamente nel grafico facendo clic con il pulsante destro del mouse sulla riga dell'area e scegliendo l'opzione desiderata dai menu di scelta rapida.

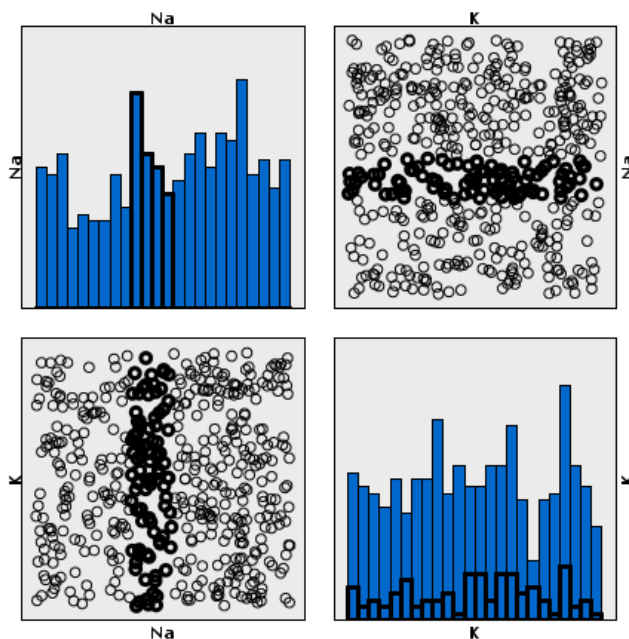
Utilizzo degli elementi contrassegnati

È possibile contrassegnare elementi, quali barre, sezioni e punti di un grafico. Non è possibile contrassegnare linee, aree e superfici di grafici diversi da plot tempo, plot multipli e grafici di valutazione, poiché in questo caso le linee fanno riferimento a campi. Ogni volta che si contrassegna un elemento, si evidenziano essenzialmente tutti i dati rappresentati da tale elemento. In un grafico dove lo stesso caso è rappresentato in più posizioni (per esempio, un grafico SPLOM), contrassegnare equivale a "pennellare". È possibile contrassegnare elementi di grafici

anche all'interno di sezioni e aree. Ogni volta che si contrassegna un elemento e si torna in modalità Modifica, gli elementi contrassegnati rimangono visibili.

Figura 5-80

Contrassegno di elementi in un grafico SPLOM



È possibile contrassegnare elementi ed eliminare i contrassegni in un grafico facendo clic sugli elementi desiderati. Al primo clic, l'elemento viene contrassegnato e appare con un bordo spesso colorato a indicare che è stato contrassegnato. Facendo nuovamente clic sull'elemento, il bordo scompare e l'elemento non risulta più contrassegnato. Per contrassegnare più elementi, è possibile tenere premuto il tasto Ctrl mentre si fa clic sugli elementi desiderati oppure trascinare il mouse intorno a ciascun elemento da contrassegnare utilizzando la "bacchetta magica". Ricordare che, se si fa clic su un'altra area o un altro elemento senza tenere premuto Ctrl, tutti gli elementi precedentemente contrassegnati perdono il contrassegno.

Dagli elementi contrassegnati di un grafico è possibile generare nodi Seleziona e Nuovo campo. [Per ulteriori informazioni, vedere l'argomento Generazioni di nodi da grafici a pag. 319.](#)

Per contrassegnare elementi:

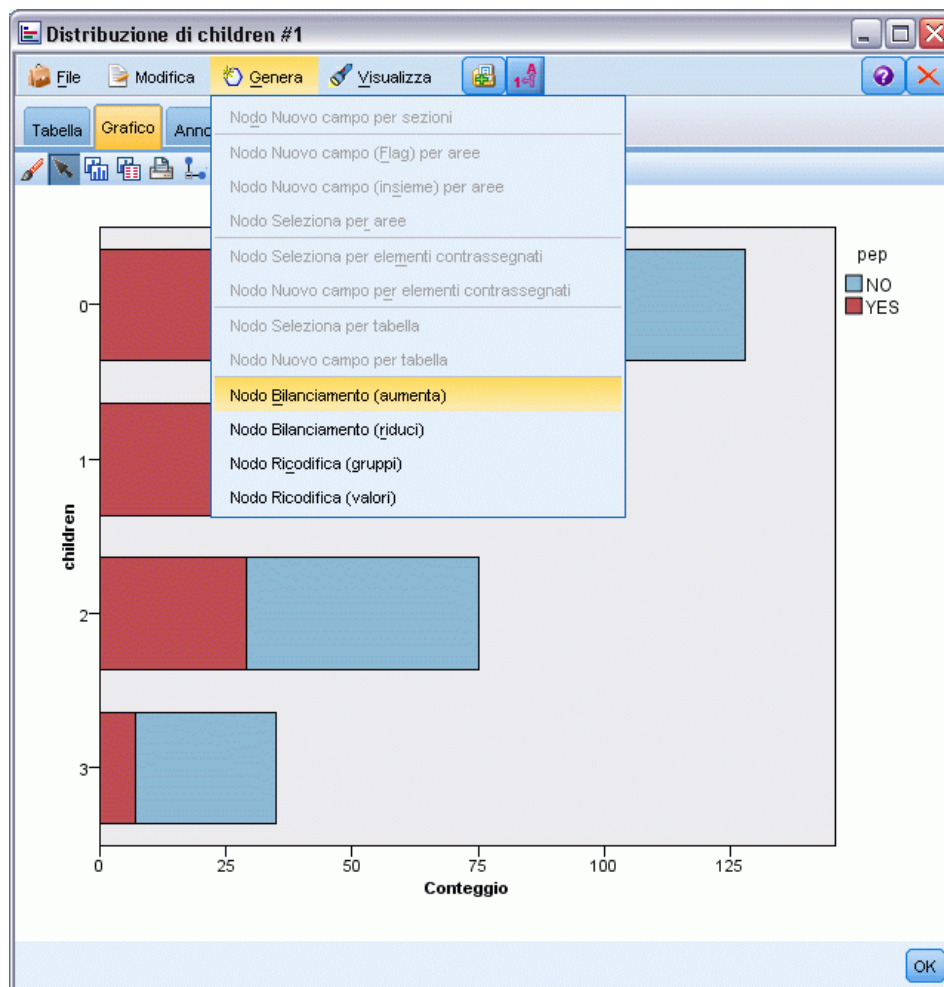
- ▶ Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.
- ▶ Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Contrassegna elementi.
- ▶ Fare clic sull'elemento desiderato oppure fare clic e trascinare tenendo premuto il pulsante del mouse per delimitare l'area che contiene più elementi.

Generazioni di nodi da grafici

Una delle funzionalità più potenti offerte dai grafici di IBM® SPSS® Modeler è la possibilità di generare nodi da un grafico o da una selezione all'interno di un grafico. Per esempio, da un grafico plot tempo è possibile generare nodi Seleziona e Nuovo campo in base a una selezione o a un'area di dati, utilizzando quindi un sottoinsieme di dati. Questa potente funzione può essere utilizzata, per esempio, per individuare ed escludere i valori anomali.

Ogni volta che si disegna una sezione, è inoltre possibile generare un nodo Nuovo campo. Nei grafici con due assi scala è possibile generare nodi Seleziona o Nuovo campo dalle aree disegnate nel grafico. Nei grafici con elementi contrassegnati è possibile generare nodi Nuovo campo, nodi Seleziona e in taluni casi nodi Filtro da tali elementi. È inoltre possibile generare nodi Bilanciamento per tutti quei grafici che mostrano una distribuzione di conteggi.

Figura 5-81
Grafico con il menu Genera visualizzato



Ogni volta che si genera un nodo, questo viene posizionato direttamente nell'area di disegno dello stream in modo che possa essere collegato a uno stream esistente. Dai grafici è possibile generare i seguenti nodi: Seleziona, Nuovo campo, Bilanciamento, Filtro e Ricodifica.

Nodi Selezione

È possibile generare nodi Selezione per verificare l'inclusione dei record all'interno di una determinata area e l'esclusione di tutti gli altri record al di fuori di tale area o viceversa allo scopo di accelerare l'elaborazione.

- **Per sezioni.** È possibile generare un nodo Selezione che includa o escluda i record all'interno di una determinata sezione. L'opzione Nodo Selezione solo per sezioni è disponibile unicamente tramite i menu di scelta rapida, poiché è necessario selezionare quale sezione utilizzare nel nodo Selezione.
- **Per aree.** È possibile generare un nodo Selezione che includa o escluda i record all'interno di una determinata area.
- **Per elementi contrassegnati.** È possibile generare nodi Selezione per acquisire i record corrispondenti agli elementi contrassegnati o ai collegamenti Web di un grafico.

Nodi Nuovo campo

È possibile generare nodi Nuovo campo da aree, sezioni ed elementi contrassegnati. Tutti i grafici possono produrre nodi Nuovo campo. Nel caso di grafici di valutazione, viene visualizzata una finestra di dialogo per la selezione del modello. Nel caso di grafici Web, è possibile scegliere Nodo Nuovo campo ("E") e Nodo Nuovo campo ("O").

- **Per sezioni.** È possibile generare un nodo Nuovo campo che produce una categoria per ogni intervallo contrassegnato sull'asse, utilizzando come nomi di categoria i nomi di sezione elencati nella finestra Modifica sezioni.
- **Per aree.** È possibile generare un nodo Nuovo campo (Deriva come flag) che crea un campo flag denominato *in_area* con i flag impostati su *V* per i record inclusi in un'area e su *F* per i record esterni a tutte le aree. È inoltre possibile generare un nodo Nuovo campo (Deriva come insieme) che produce un insieme con un valore per ogni area con un nuovo campo denominato *area* per ogni record, che assume come proprio valore il nome dell'area contenente i record. Ai record esterni a qualsiasi area viene assegnato il nome dell'area di default I nomi dei valori diventano i nomi delle aree elencati nella finestra di dialogo Modifica aree.
- **Per elementi contrassegnati.** È possibile generare un nodo Nuovo campo che calcola un flag che è *Vero* per tutti gli elementi contrassegnati e *Falso* per tutti gli altri record.

Nodi Bilanciamento

È possibile generare nodi Bilanciamento per correggere eventuali sbilanciamenti nei dati, per esempio per ridurre la frequenza di valori comuni [utilizzare la voce di menu Nodo Bilanciamento (riduci)] o per aumentare l'occorrenza di valori non frequenti [utilizzare la voce di menu Nodo Bilanciamento (aumenta)]. È possibile generare nodi Bilanciamento per tutti quei grafici che mostrano una distribuzione di conteggi, quali istogrammi, grafici a punti, grafici raccolta, grafici con barre dei conteggi, grafici con torta dei conteggi e plot multipli.

Nodi Filtro

È possibile generare nodi Filtro per rinominare o filtrare campi in base alle linee o ai nodi contrassegnati nel grafico. Nel caso dei grafici di valutazione, la linea di migliore approssimazione non genera un nodo Filtro.

Nodi Ricodifica

È possibile generare nodi Ricodifica per ricodificare valori. Questa opzione è utilizzata per i grafici di distribuzione. È possibile generare un nodo Ricodifica per **gruppi** per ricodificare valori specifici di un campo visualizzato in base alla loro inclusione in un gruppo (selezionare i gruppi con Ctrl+clic nella scheda Tabelle). È inoltre possibile generare un nodo Ricodifica per **valori** per ricodificare i dati in un insieme esistente di numerosi valori, per esempio per ricodificare i dati in un insieme standard di valori al fine di unire i dati finanziari di varie aziende per l'analisi.

Nota: se i valori sono predefiniti, è possibile leggerli in SPSS Modeler come un file piatto e utilizzare un nodo Distribuzione per visualizzare tutti i valori. Generare quindi un nodo Ricodifica (valori) per questo campo direttamente dal grafico. In questo modo tutti i valori obiettivo verranno inseriti nella colonna *Nuovi valori* del nodo Ricodifica (elenco a discesa).

Generazioni di nodi da grafici

È possibile utilizzare il menu Genera della finestra di output del grafico per generare nodi. Il nodo generato verrà posizionato nell'area di disegno dello stream. Per utilizzare il nodo è necessario connetterlo a uno stream esistente.

Per generare un nodo da un grafico:

- ▶ Controllare che sia attiva la modalità Esplora. Dai menu, scegliere Visualizza > Modalità Esplora.
- ▶ Sulla barra degli strumenti della modalità Esplora, fare clic sul pulsante Area.
- ▶ Definire eventuali sezioni, aree o elementi contrassegnati necessari per generare il nodo.
- ▶ Dal menu Genera, scegliere il tipo di nodo che si desidera generare. Sono attivati solo quelli che è possibile generare.

Nota: in alternativa, è inoltre possibile generare nodi direttamente dal grafico facendo clic con il pulsante destro del mouse e scegliendo l'opzione Genera desiderata dai menu di scelta rapida.

Modifica delle visualizzazioni

Mentre la modalità Esplora consente di esaminare in modo analitico i dati e i valori presenti nella visualizzazione, con la modalità Modifica è possibile intervenire sul layout e sull'aspetto della visualizzazione. Per esempio, è possibile modificare caratteri e colori in base alla guida di stile della propria organizzazione. Per selezionare questa modalità, scegliere Visualizza > Modalità Modifica dai menu (oppure facendo clic sull'icona nella barra degli strumenti).

In modalità Modifica sono disponibili numerose barre degli strumenti che riguardano diversi aspetti del layout della visualizzazione. Se lo si desidera, è possibile nascondere le barre degli strumenti che non vengono utilizzate in modo da incrementare lo spazio della finestra di dialogo in cui il grafico viene visualizzato. Per selezionare o deselezionare le barre degli strumenti, fare clic sui relativi nomi nel menu Visualizza.

Nota: per aumentare ulteriormente il livello di dettagli delle visualizzazioni, è possibile applicare etichette dei titoli, delle annotazioni e degli assi. [Per ulteriori informazioni, vedere l'argomento Aggiunta di titoli e annotazioni a pag. 339.](#)

In **modalità Modifica** sono disponibili svariate opzioni per modificare le visualizzazioni. È possibile:

- Modificare il testo e formattarlo.
- Modificare il colore di riempimento, la trasparenza e il motivo delle cornici e degli elementi grafici.
- Modificare il colore e il tratteggio dei bordi e delle linee.
- Ruotare e modificare la forma e le proporzioni degli elementi puntiformi.
- Modificare le dimensioni degli elementi grafici (quali barre e punti).
- Regolare lo spazio attorno agli elementi utilizzando margini e spaziatura.
- Specificare la formattazione dei numeri.
- Modificare le impostazioni degli assi e delle scale.
- Ordinare, escludere e comprimere categorie su un asse di categorie.
- Impostare l'orientamento dei riquadri.
- Applicare trasformazioni a un sistema di coordinate.
- Modificare le statistiche, i tipi di elemento grafico e i modificatori di collisione.
- Modificare la posizione della legenda.
- Applicare fogli di stile di visualizzazione.

Le sezioni seguenti illustrano come eseguire le operazioni elencate sopra. Si consiglia inoltre di leggere le regole generali per la modifica dei grafici.

Come passare alla modalità Modifica

- ▶ Dai menu, scegliere:
Visualizza > Modalità Modifica

Regole generali per la modifica delle visualizzazioni

Modalità Modifica

Tutte le modifiche vengono apportate in modalità Modifica, attivabile nei menu selezionando:
Visualizza > Modalità Modifica

Selezione

Le opzioni di modifica disponibili dipendono dalla selezione. A seconda dell'elemento selezionato, nella barra degli strumenti e nella palette delle proprietà vengono attivate opzioni diverse. Per la selezione corrente sono valide solo le opzioni attivate. Per esempio, se si seleziona un asse, nella palette delle proprietà saranno disponibili le schede Scala, Segni di graduazione maggiori e Segni di graduazione minori.

Di seguito vengono riportati suggerimenti per la selezione degli elementi nella visualizzazione:

- Fare clic su un elemento per selezionarlo.

- Selezionare un elemento grafico (quale i punti di un grafico a dispersione o le barre di un grafico a barre) con un singolo clic. Dopo questa prima selezione, fare nuovamente clic per restringere la selezione a gruppi di elementi grafici o a un singolo elemento grafico.
- Premere Esc per deselezionare tutto.

Tavolozze

Quando nella visualizzazione è selezionato un oggetto, le diverse tavolozze vengono aggiornate per riflettere tale selezione. Le tavolozze contengono i controlli per apportare modifiche alla selezione. Possono essere barre degli strumenti o un riquadro con più comandi e più schede. Poiché le tavolozze possono essere nascoste, è opportuno verificare che siano visibili quelle necessarie per apportare le modifiche. Tale verifica è possibile utilizzando il menu *Visualizza*.

Le tavolozze possono essere spostate facendo clic e trascinando lo spazio vuoto all'interno della tavolozza di una barra degli strumenti o il bordo sinistro di altre tavolozze. Indicazioni visive mostrano il punto in cui è possibile agganciare la tavolozza. Nel caso di tavolozze non relative a una barra degli strumenti, è possibile fare clic sul pulsante di chiusura per nascondere la tavolozza e quindi su quello di sgancio per visualizzarla in un'altra finestra. Il pulsante *Aiuto* consente di visualizzare informazioni sulla tavolozza in questione.

Impostazioni automatiche

Alcune impostazioni sono dotate di un'opzione *-automatico-* che indica l'applicazione di valori automatici. Le impostazioni automatiche utilizzate dipendono dalla specifica visualizzazione e dai valori dei dati. È possibile inserire un valore e ignorare l'impostazione automatica. Per ripristinare l'impostazione automatica, eliminare il valore corrente e premere *Invio*. Per l'impostazione sarà visualizzata nuovamente la dicitura *-automatico-*.

Rimuovere o nascondere elementi

È possibile rimuovere/nascondere vari elementi nella visualizzazione. Per esempio, è possibile nascondere la legenda o l'etichetta di un asse. Per eliminare un elemento, selezionarlo e premere *Canc*. Se l'eliminazione dell'elemento non è consentita, l'operazione non avrà alcun effetto. Se si elimina accidentalmente un elemento, premere *Ctrl+Z* per annullare l'eliminazione.

State

Alcune barre degli strumenti cambiano a seconda dello stato della selezione corrente, mentre altre no. La palette delle proprietà rispecchia sempre lo stato. Se una barra degli strumenti *non* rispecchia lo stato della selezione, questo è segnalato nell'argomento che descrive la barra degli strumenti.

Modifica e formattazione del testo

È possibile modificare il testo direttamente nel grafico e cambiare la formattazione di un intero blocco di testo. Si noti che non è possibile modificare il testo collegato direttamente ai valori dei dati. Per esempio, non è possibile modificare l'etichetta di un segno di graduazione poiché

il suo contenuto deriva dai dati sottostanti. È comunque possibile formattare tutto il testo della visualizzazione.

Come modificare il testo direttamente nel grafico

- ▶ Fare doppio clic sul blocco di testo, per selezionare tutto il testo. Tutte le barre degli strumenti sono disattivate, dal momento che non si può intervenire su altre parti della visualizzazione durante la modifica del testo.
- ▶ Digitare il testo con cui sostituire quello selezionato. È possibile anche fare nuovamente clic sul testo per visualizzare un cursore, posizionarlo nel punto desiderato e inserire il testo aggiuntivo.

Come formattare il testo

- ▶ Selezionare la cornice contenente il testo, senza fare doppio clic sul testo.
- ▶ Formattare il testo mediante la barra degli strumenti dei caratteri. Se la barra degli strumenti non è attivata, verificare che sia selezionata solo la *cornice* contenente il testo. Se è selezionato il testo, la barra degli strumenti sarà disattivata.

Figura 5-82
Barra degli strumenti dei caratteri



Per i caratteri è possibile modificare:

- Colore
- Famiglia (per esempio Arial o Verdana)
- Dimensioni (l'unità è pt a meno che non ne venga indicata una diversa, per esempio pc)
- Peso
- Allineamento rispetto alla cornice del testo

La formattazione viene applicata a tutto il testo all'interno della cornice. Non è possibile modificare la formattazione delle singole lettere o parole all'interno di un determinato blocco di testo.

Modifica dei colori, dei motivi, dei tratteggi e della trasparenza

Molti degli elementi di una visualizzazione sono caratterizzati da un bordo e da un riempimento. L'esempio più ovvio sono le barre all'interno dei grafici a barre. Il colore delle barre è il colore di riempimento. Attorno alle barre può anche essere presente un bordo nero continuo.

Nella visualizzazione sono presenti elementi meno ovvi dotati di colori di riempimento. Se il colore di riempimento è trasparente è possibile che non ci si accorga della sua presenza. Si pensi per esempio al testo dell'etichetta di un asse: il testo sembra "mobile", ma in realtà è contenuto in una cornice con colore di riempimento trasparente. La cornice diventa visibile se si seleziona l'etichetta dell'asse.

Tutte le cornici della visualizzazione possono avere uno stile di riempimento e di bordo, compresa quella che racchiude l'intera rappresentazione. Inoltre, a tutti i riempimenti è associato un livello di opacità/trasparenza che può essere regolato.

Come modificare i colori, i motivi, il tratteggio e la trasparenza

- ▶ Selezionare l'elemento che si desidera formattare. Per esempio, selezionare le barre di un grafico a barre o una cornice contenente del testo. Se la visualizzazione è suddivisa in base a una variabile o a un campo categoriale, è anche possibile selezionare il gruppo che corrisponde a una singola categoria. Ciò consente di modificare l'aspetto di default assegnato a quel gruppo. Per esempio, è possibile modificare il colore di uno dei gruppi sovrapposti in un grafico a barre sovrapposte.
- ▶ Per modificare il colore di riempimento o del bordo o il motivo di riempimento, utilizzare la barra degli strumenti dei colori.

Figura 5-83
Barra degli strumenti dei colori

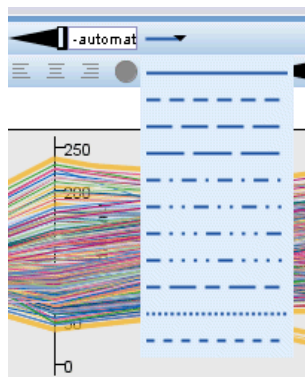


Nota: questa barra degli strumenti non rispecchia lo stato della selezione corrente.

Per modificare un colore o un riempimento, è possibile fare clic sul pulsante per selezionare l'opzione visualizzata o sulla freccia a discesa per sceglierne un'altra. Per i colori si noterà la presenza di un colore che sembra bianco attraversato da una riga diagonale rossa. Questo è il colore trasparente, che si può utilizzare, per esempio, per nascondere i bordi delle barre di un istogramma.

- Il primo pulsante controlla il colore di riempimento.
 - Il secondo controlla il colore del bordo.
 - Il terzo controlla il motivo di riempimento. Il motivo di riempimento utilizza il colore del bordo e quindi è visibile solo se è presente un colore del bordo visibile.
 - Il quarto controllo è un cursore e una casella di testo che controllano l'opacità, nonché il colore e il motivo di riempimento. Minore è la percentuale, minore risulta l'opacità e maggiore la trasparenza. Una percentuale pari a 100% conferisce un'opacità totale (senza trasparenza).
- ▶ Per modificare il tratteggio di un bordo o di una linea, utilizzare la barra degli strumenti delle linee.

Figura 5-84
Barra degli strumenti delle linee



Nota: questa barra degli strumenti non rispecchia lo stato della selezione corrente.

Come avviene per l'altra barra degli strumenti, è possibile fare clic sul pulsante per selezionare l'opzione visualizzata oppure fare clic sulla freccia a discesa per scegliere un'altra opzione.

Rotazione e modifica della forma e delle proporzioni degli elementi puntiformi

È possibile ruotare gli elementi puntiformi, assegnare loro una forma predefinita diversa o modificarne le proporzioni (il rapporto tra larghezza e altezza).

Come modificare gli elementi puntiformi

- ▶ Selezionare gli elementi puntiformi. Non è possibile ruotare o modificare la forma e le proporzioni di un singolo elemento puntiforme.
- ▶ Per modificare i punti, utilizzare la barra degli strumenti dei simboli.

Figura 5-85
Barra degli strumenti dei simboli



- Il primo pulsante consente di modificare la forma dei punti. Fare clic sulla freccia a discesa e selezionare una forma predefinita.
- Il secondo pulsante consente di ruotare i punti secondo un orientamento specifico una bussola simbolica. Fare clic sulla freccia a discesa e trascinare l'ago nella posizione desiderata.
- Il terzo pulsante consente di modificare le proporzioni. Fare clic prima sulla freccia a discesa e poi sul rettangolo così visualizzato, quindi trascinarlo per modificarne le proporzioni. La forma del rettangolo rappresenta le proporzioni dei punti.

Modifica delle dimensioni degli elementi grafici

È possibile modificare le dimensioni degli elementi grafici della visualizzazione, (barre, linee, punti e simili). Se le dimensioni di un elemento grafico sono determinate da una variabile o da un campo, le dimensioni specificate sono quelle *minime*.

Come modificare le dimensioni degli elementi grafici

- ▶ Selezionare gli elementi grafici da ridimensionare.
- ▶ Utilizzare il dispositivo di scorrimento oppure inserire dimensioni specifiche per l'opzione disponibile nella barra degli strumenti dei simboli. L'unità è il pixel a meno che non ne venga indicata una diversa (vedere di seguito per l'elenco completo delle abbreviazioni di unità). È possibile anche specificare una percentuale (per esempio il 30%) per indicare che un elemento grafico utilizza la percentuale impostata dello spazio disponibile. Lo spazio disponibile dipende dal tipo di elemento grafico e dalla specifica visualizzazione.

Tabella 5-2

Abbreviazioni di unità valide

Abbreviazione	Unità
cm	centimetro
in	pollice
mm	millimetro
pc	pica
pt	punto
px	pixel

Figura 5-86

Controllo delle dimensioni nella barra degli strumenti dei simboli



Impostazione dei margini e della spaziatura

Se c'è troppo (o troppo poco) spazio intorno o all'interno di una cornice di una visualizzazione, è possibile modificarne le impostazioni dei margini e di riempimento. Il **margin** è la quantità di spazio compresa tra la cornice e gli elementi che la circondano. La **spaziatura** è la quantità di spazio compresa tra il bordo della cornice e il suo *contenuto*.

Come impostare i margini e la spaziatura

- ▶ Selezionare la cornice di cui si desidera impostare i margini e la spaziatura. Può trattarsi di una cornice di testo, della cornice attorno alla legenda o anche della cornice dei dati che visualizza gli elementi grafici (quali barre e punti).
- ▶ Per specificare le impostazioni, utilizzare la scheda Margini nella palette delle proprietà. Le dimensioni sono espresse in pixel a meno che non venga indicata un'unità diversa, per esempio cm o in.

Figura 5-87

Scheda Margini



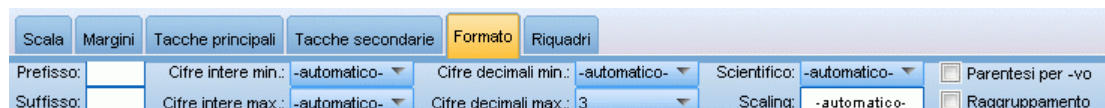
Formattazione dei numeri

È possibile specificare il formato dei numeri nelle etichette delle tacche su un asse continuo o nelle etichette dei valori dei dati caratterizzate da un numero. Ad esempio, è possibile specificare che i numeri visualizzati nelle etichette delle tacche siano rappresentati in migliaia.

Come specificare i formati dei numeri

- ▶ Selezionare le etichette delle tacche dell'asse continuo o le etichette dei valori dei dati, qualora contengano numeri.
- ▶ Fare clic sulla scheda Formato della tavolozza delle proprietà.

Figura 5-88
Scheda Formato



- ▶ Selezionare le opzioni desiderate per la formattazione dei numeri:

Prefisso. Un carattere da visualizzare all'inizio del numero. Immettere ad esempio il simbolo di euro (€) nel caso in cui i numeri si riferiscano ai salari in questa valuta.

Suffisso. Un carattere da visualizzare alla fine del numero. Ad esempio, immettere il simbolo di percentuale (%) nel caso in cui i numeri si riferiscano a percentuali.

Cifre intere min.. Il numero minimo di cifre da visualizzare nella parte intera di una rappresentazione decimale. Se il valore effettivo non contiene il numero minimo di cifre, la parte intera del valore verrà riempita di zeri.

Cifre intere max.. Il numero massimo di cifre da visualizzare nella parte intera di una rappresentazione decimale. Se il valore effettivo supera il numero minimo di cifre, la parte intera del valore verrà sostituita da asterischi.

Cifre decimali min.. Il numero minimo di cifre da visualizzare nella parte decimale di una rappresentazione decimale o scientifica. Se il valore effettivo non contiene il numero minimo di cifre, la parte decimale del valore verrà riempita di zeri.

Cifre decimali max.. Il numero massimo di cifre da visualizzare nella parte decimale di una rappresentazione decimale o scientifica. Se il valore effettivo supera il numero minimo di cifre, il decimale viene arrotondato al numero di cifre appropriato.

Scientifica. Consente di scegliere se visualizzare i numeri in notazione scientifica. La notazione scientifica è utile per numeri molto grandi o molto piccoli. Grazie all'impostazione -automatico-, l'applicazione è in grado di stabilire quando è il caso di utilizzare notazioni scientifiche.

Scaling. Un fattore di scala, ossia un numero per il quale viene diviso il valore originario. Utilizzare un fattore di scala quando i numeri sono grandi ma non si desidera che l'etichetta venga eccessivamente ampliata per adattarsi al numero. Se si modifica il formato del numero delle etichette delle tacche, accertarsi di modificare il titolo dell'asse per indicare come deve essere interpretato il numero. Si supponga, ad esempio, che sull'asse dei valori siano visualizzati gli stipendi e che le etichette riportino 30.000, 50.000 e 70.000. È possibile utilizzare un fattore di

scala pari a 1000 per visualizzare 30, 50 e 70. Sarà quindi necessario modificare il titolo dell'asse dei valori aggiungendo il testo in migliaia.

Parentesi per -vo. Consente di scegliere se racchiudere i valori negativi tra parentesi.

Raggruppamento. Consente di scegliere se visualizzare un carattere tra gruppi di cifre. Il carattere utilizzato per il raggruppamento delle cifre è determinato dalle impostazioni internazionali in uso nel computer.

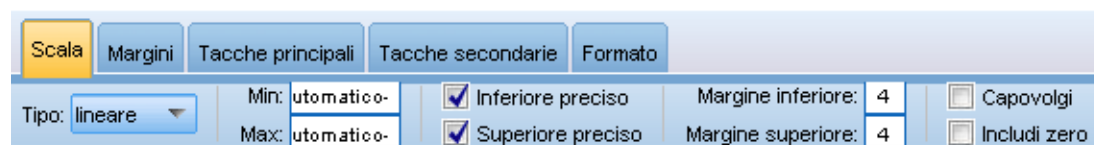
Modifica delle impostazioni degli assi e delle scale

Esistono varie opzioni per la modifica degli assi e delle scale.

Come modificare le impostazioni degli assi e delle scale

- ▶ Selezionare un qualsiasi elemento dell'asse (per esempio l'etichetta dell'asse o le etichette dei segni di graduazione).
- ▶ Modificare le impostazioni dell'asse e della scala mediante le schede Scala, Segni di graduazione maggiori e Segni di graduazione minori nella palette delle proprietà.

Figura 5-89
Palette delle proprietà



Scheda Scala

Nota: la scheda Scala non viene visualizzata per i grafici in cui i dati sono pre-aggregati (ad esempio, istogrammi).

Tipo. Indica se la scala è lineare o trasformata. Le trasformazioni di scala facilitano la comprensione dei dati o la formulazione dei presupposti necessari per l'inferenza statistica. Nei grafici a dispersione si può utilizzare una scala trasformata se la relazione tra variabili o campi indipendenti e dipendenti è non lineare. Le trasformazioni di scala si possono utilizzare anche per migliorare la simmetria di un istogramma asimmetrico al fine di renderlo simile a una distribuzione normale. Si noti che la trasformazione riguarda solo la scala con cui vengono rappresentati i dati, non i dati effettivi.

- **lineare.** Indica una scala lineare, non trasformata.
- **logaritmo.** Specifica una scala trasformata con un logaritmo in base 10. Per considerare anche i valori zero e negativi, questa trasformazione utilizza una versione modificata della funzione logaritmica. Questa funzione “logaritmica sicura” è definita in $\text{sign}(x) * \log(1 + \text{abs}(x))$. Quindi $\text{safeLog}(-99)$ è uguale a:

$$\text{sign}(-99) * \log(1 + \text{abs}(-99)) = -1 * \log(1 + 99) = -1 * 2 = -2$$
- **potenza.** Specifica una scala trasformata mediante potenza utilizzando un esponente di 0.5. Per considerare anche i valori negativi, questa trasformazione utilizza una versione modificata della funzione potenza. Questa funzione “potenza sicura” è definita in $\text{sign}(x) * \text{pow}(\text{abs}(x), 0.5)$. Quindi $\text{safePower}(-100)$ è uguale a:

$$\text{sign}(-100) * \text{pow}(\text{abs}(-100), 0.5) = -1 * \text{pow}(100, 0.5) = -1 * 10 = -10$$

Min/Max/Inferiore preciso/Superiore preciso. Specificano l’intervallo della scala. Se si seleziona Inferiore preciso e Superiore preciso, l’applicazione è in grado di selezionare una scala adeguata in base ai dati. Il minimo e il massimo sono “precisi” perché in genere sono valori interi maggiori o minori dei valori dei dati massimo e minimo. Per esempio, se l’intervallo di dati è compreso fra 4 e 92, i valori inferiore e superiore preciso per la scala possono essere 0 e 100 anziché il minimo e il massimo dei dati effettivi. Prestare attenzione a non impostare un intervallo troppo piccolo che nasconda elementi importanti. Notare inoltre che non è possibile impostare un minimo e un massimo espliciti se è selezionata l’opzione Includi zero.

Margine inferiore/Margine superiore. Consente di creare margini all’estremità inferiore e/o superiore dell’asse. Il margine risulta perpendicolare all’asse selezionato. L’unità è il pixel a meno che non ne venga indicata una diversa, per esempio cm o in. Per esempio, se si imposta il Margine superiore su 5 per l’asse verticale, lungo la parte superiore della cornice dei dati correrà un margine orizzontale di 5 pixel.

Capovolgi. Specifica se la scala viene capovolta.

Includi zero. Indica che la scala deve includere lo 0. Questa opzione si utilizza in genere per i grafici a barre per fare in modo che le barre inizino in corrispondenza dello 0 anziché di un valore prossimo all’altezza della barra più piccola. Se è selezionata questa opzione, le opzioni Min e Max sono disattivate poiché non è possibile impostare un minimo e un massimo personalizzati per l’intervallo della scala.

Schede Segni di graduazione maggiori/Segni di graduazione minori

I segni di graduazione o tacche sono le linee riportate su un asse, che indicano i valori in corrispondenza di intervalli o categorie specifiche. **I segni di graduazione maggiori** sono i segni di graduazione dotati di etichette e sono anche più lunghi rispetto agli altri segni di graduazione. **I segni di graduazione minori** sono i segni che compaiono tra un segno di graduazione maggiore e l’altro. Alcune opzioni sono specifiche per un determinato tipo di segno di graduazione, ma la maggior parte è disponibile sia per i segni di graduazione maggiori che per quelli minori.

Mostra segni di graduazione. Consente di specificare se le tacche principali o secondarie vengono visualizzate in un grafico.

Mostra linee della griglia. Specifica se in corrispondenza dei segni di graduazione maggiori o minori vengono visualizzate le linee della griglia. Le **linee della griglia** sono linee che attraversano un intero grafico da un asse all'altro.

Posizione. Specifica la posizione dei segni di graduazione rispetto all'asse.

Lunghezza. Specifica la lunghezza dei segni di graduazione. L'unità è il pixel a meno che non venga indicata una diversa, per esempio cm o in.

Base. *Valida solo per i segni di graduazione maggiori.* Specifica il valore in corrispondenza del quale è visualizzato il primo segno di graduazione maggiore.

Delta. *Valida solo per i segni di graduazione maggiori.* Specifica la differenza tra i segni di graduazione maggiori. Questo significa che tali segni saranno visualizzati ogni n valori, dove n è il valore delta.

Divisioni. *Valida solo per i segni di graduazione minori.* Specifica il numero di divisioni con segni di graduazione minori tra i segni di graduazione maggiori. Il numero dei segni di graduazione minori è inferiore di un'unità rispetto al numero delle divisioni. Si supponga per esempio che siano presenti segni di graduazione maggiori in corrispondenza di 0 e 100. Se si inserisce 2 come numero di divisioni con segni di graduazione minori vi sarà *un* segno di graduazione minore in corrispondenza di 50, a dividere l'intervallo 0–100 e creare *due* divisioni.

Modifica delle categorie

È possibile modificare le categorie su un asse di categorie in diversi modi:

- Modificare l'ordine di visualizzazione delle categorie.
- Escludere categorie specifiche.
- Aggiunta di una categoria che non viene visualizzata nell'insieme di dati.
- Compressione/combinazione di categorie di dimensioni ridotte in un'unica categoria.

Come modificare l'ordinamento delle categorie

- ▶ Selezionare un'asse delle categorie. La palette Categorie visualizza le categorie sull'asse.

Nota: se la palette non è visibile, assicurarsi di averla attivata. Dal menu Visualizza in IBM® SPSS® Modeler, scegliere Categorie.

- ▶ Nella palette Categorie, selezionare un'opzione di ordinamento dall'elenco a discesa:

Personalizzato. Ordina le categorie in base all'ordine con cui appaiono nella palette. Utilizzare i pulsanti freccia per spostare le categorie in cima, su, giù e al fondo dell'elenco.

Dati. Ordina le categorie in base all'ordine con cui si presentano nell'insieme di dati.

Nome. Ordina le categorie in ordine alfabetico, utilizzando i nomi visualizzati nella palette. Questi potrebbero essere un valore oppure un'etichetta, a seconda che sia selezionato il pulsante della barra degli strumenti per visualizzare valori o etichette.

Valore. Consente di ordinare le categorie in base al valore dei dati sottostanti, utilizzando i valori visualizzati tra parentesi nella tavolozza. Questa opzione è supportata unicamente dalle sorgenti dati con metadati (ad esempio i file di dati IBM® SPSS® Statistics).

Statistica. Ordina le categorie in base alla statistica calcolata per ogni categoria. Esempi di statistica includono conteggi, percentuali e medie. Questa opzione è disponibile solo se nel grafico è utilizzata una statistica.

Come aggiungere una categoria

Per impostazione predefinita, sono disponibili solo le categorie visualizzate nell'insieme di dati. È possibile aggiungere una categoria alla visualizzazione, se necessario.

- ▶ Selezionare un'asse delle categorie. La palette Categorie visualizza le categorie sull'asse.

Nota: se la palette non è visibile, assicurarsi di averla attivata. Dal menu Visualizza in SPSS Modeler, scegliere Categorie.

- ▶ Nella tavolozza Categorie, fare clic sul pulsante Aggiungi categoria:

Figura 5-90
Pulsante Aggiungi categoria



- ▶ Nella finestra di dialogo Aggiungi una nuova categoria, specificare un nome per la categoria.
- ▶ Fare clic su OK.

Come escludere categorie specifiche

- ▶ Selezionare un'asse delle categorie. La palette Categorie visualizza le categorie sull'asse.

Nota: se la palette non è visibile, assicurarsi di averla attivata. Dal menu Visualizza in SPSS Modeler, scegliere Categorie.

- ▶ Nella palette Categorie, selezionare il nome di una categoria nell'elenco Includi, quindi fare clic sul pulsante X. Per ripristinare la categoria, selezionarne il nome nell'elenco Escludi, quindi fare clic sulla freccia a destra dell'elenco.

Come comprimere/combinare categorie di dimensioni ridotte

È possibile raggruppare categorie di dimensioni così ridotte da non richiedere una visualizzazione separata. Se, ad esempio, si utilizza un grafico a torta composto da molte categorie, è possibile comprimerle con una percentuale inferiore a 10. La compressione delle categorie è disponibile solo per le statistiche additive. Ad esempio, non è possibile aggiungere insieme le medie poiché le medie non sono additive. Pertanto, non è disponibile la compressione/combinazione delle categorie utilizzando una media.

- ▶ Selezionare un'asse delle categorie. La palette Categorie visualizza le categorie sull'asse.

Nota: se la palette non è visibile, assicurarsi di averla attivata. Dal menu Visualizza in SPSS Modeler, scegliere Categorie.

- Nella palette Categorie, selezionare Comprimi e specificare una percentuale. Le categorie la cui percentuale del totale è inferiore al numero specificato vengono combinate in un'unica categoria. La percentuale è basata sulla statistica mostrata nel grafico. La compressione è disponibile solo per statistiche basate su conteggio o somma.

Modifica dei riquadri di orientamento

Se nella visualizzazione si utilizzano i riquadri, è possibile modificarne l'orientamento.

Come modificare l'orientamento dei riquadri

- Selezionare una parte qualsiasi della visualizzazione.
- Fare clic sulla scheda Riquadri nella tavolozza delle proprietà.

Figura 5-91
Scheda Riquadri



- Selezionare un'opzione nell'elenco Layout:

Tabella. Dispone i riquadri come una tabella, con una riga o una colonna assegnata a ogni singolo valore.

Trasposto. Dispone i riquadri come una tabella ma in più scambia fra loro le righe e le colonne originali. Questa opzione non equivale a trasporre il grafico in sé. Si noti che l'asse delle x e quello delle y restano invariati quando si seleziona questa opzione.

Elenco. Dispone i riquadri come un elenco in cui ogni cella rappresenta una combinazione di valori. Le colonne e le righe non sono più assegnate a singoli valori. Questa opzione consente ai riquadri di andare a capo, se necessario.

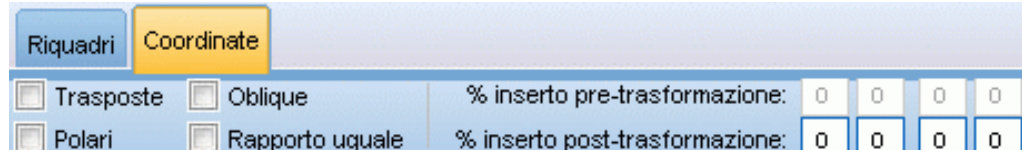
Trasformazione del sistema di coordinate

Molte visualizzazioni vengono rappresentate in un sistema di coordinate piatto di forma rettangolare. Tale sistema può essere trasformato in base alle proprie esigenze. È ad esempio possibile applicarvi una trasformazione polare, aggiungere effetti di ombreggiatura obliqui e trasporre gli assi. Se già applicate alla visualizzazione corrente, tutte queste trasformazioni possono anche essere annullate. Si supponga, ad esempio, di generare un grafico a torta in un sistema di coordinate polari. Se si desidera, è possibile annullare la trasformazione polare e visualizzare il grafico a torta sotto forma di un'unica barra sovrapposta nel sistema di coordinate rettangolare.

Come trasformare il sistema di coordinate

- ▶ Selezionare il sistema di coordinate da trasformare facendo clic sulla cornice intorno al singolo grafico.
- ▶ Fare clic sulla scheda Coordinate nella tavolozza delle proprietà.

Figura 5-92
Scheda Coordinate



- ▶ Selezionare le trasformazioni da applicare al sistema di coordinate. Una trasformazione può anche essere deselezionata per annullarla.

Trasposto. La modifica dell'orientamento degli assi, nota come **trasposizione**, si tratta di un'operazione analoga all'inversione dell'asse verticale e di quello orizzontale in una visualizzazione 2-D.

Polari. La trasformazione polare consente di disegnare gli elementi grafici a una determinata angolazione e distanza rispetto al centro del grafico. I grafici a torta sono visualizzazioni monodimensionali con una trasformazione polare che consente di disegnare le singole barre a specifiche angolazioni. I grafici a radar sono visualizzazioni 2-D con una trasformazione polare che consente di disegnare gli elementi grafici a specifiche angolazioni e distanze rispetto al centro del grafico stesso. Nelle visualizzazioni 3-D è presente un'ulteriore dimensione di profondità.

Oblique. Le trasformazioni oblique aggiungono un effetto 3-D agli elementi grafici. Viene applicata la profondità, che tuttavia è puramente decorativa e non influenzata da particolari valori dei dati.

Rapporto uguale. L'applicazione dello stesso rapporto consente di specificare che la stessa distanza in ciascuna scala rappresenta la stessa differenza tra i valori dei dati. Ad esempio, in entrambe le scale 2 cm rappresentano una differenza di 1000.

% inserto pre-trasformazione. Se, dopo la trasformazione, gli assi vengono tagliati, prima di applicarla è possibile aggiungere inserti nel grafico. Gli inserti riducono di una determinata percentuale le dimensioni prima che vengano applicate le trasformazioni al sistema di coordinate. È possibile controllare le dimensioni x inferiore, x superiore, y inferiore e y superiore in quest'ordine.

% inserto post-trasformazione. Per modificare le proporzioni del grafico è possibile aggiungere inserti dopo aver applicato la trasformazione. Gli inserti riducono di una determinata percentuale le dimensioni dopo l'applicazione di trasformazioni al sistema di coordinate. Gli inserti possono essere applicati anche in assenza di trasformazioni. È possibile controllare le dimensioni x inferiore, x superiore, y inferiore e y superiore in quest'ordine.

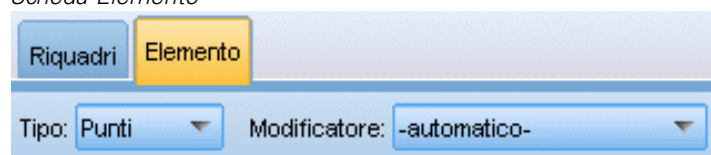
Modifica di statistiche ed elementi grafici

È possibile convertire un elemento grafico di tipo in un altro tipo, modificare le statistiche utilizzate per disegnare l'elemento grafico o specificare il modificatore di collisione che determina cosa accade in caso di elementi grafici sovrapposti.

Come convertire un elemento grafico

- ▶ Selezionare l'elemento grafico da convertire.
- ▶ Fare clic sulla scheda Elemento nella tavolozza delle proprietà.

Figura 5-93
Scheda Elemento



- ▶ Selezionare un nuovo tipo di elemento grafico dall'elenco Tipo.

Tipo di elemento grafico	Descrizione
Point	Un marker che identifica un punto dati specifico. Un elemento punto viene utilizzato nei grafici a dispersione e in altre visualizzazioni correlate.
Interval	Una forma rettangolare disegnata in corrispondenza di un valore di dati specifico e che riempie lo spazio tra un'origine e un altro valore di dati. Un elemento intervallo viene utilizzato nei grafici a barre e negli istogrammi.
Line	Una linea che collega valori di dati.
Path	Una linea che collega valori di dati nell'ordine in cui appaiono nell'insieme di dati.
Area	Una linea che collega elementi di dati con l'area tra la linea e un'origine riempita.
Polygon	Una forma a più lati che racchiude un'area di dati. Un elemento poligono potrebbe essere usato in un grafico a dispersione in bin o in una mappa.
Schema	Un elemento composto da una scatola con baffi e simboli che indicano i valori anomali. Un elemento schema viene utilizzato per i grafici a scatole.

Come modificare le statistiche

- ▶ Selezionare l'elemento grafico di cui si desidera modificare le statistiche.
- ▶ Fare clic sulla scheda Elemento nella tavolozza delle proprietà.
- ▶ Dall'elenco a discesa Riepilogo selezionare una nuova statistica. Con la selezione di una statistica i dati vengono aggregati. Se, invece, si desidera visualizzare i dati separatamente, selezionare (nessuna statistica) dall'elenco Riepilogo.

Statistiche riassuntive calcolate a partire da un campo continuo

- **Mean.** Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.
- **Median.** È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.
- **Mode.** Il valore o i valori che ricorrono più frequentemente. Se più valori condividono la maggiore ricorrenza, ognuno di essi è una moda.
- **Minimo.** Il valore più basso assunto da una variabile numerica.
- **Massimo.** Il valore più alto di una variabile numerica.
- **Intervallo.** Differenza tra i valori minimi e massimi.
- **Intervallo medio.** Il punto medio dell'intervallo, vale a dire, il valore la cui differenza rispetto al punto minimo è uguale alla differenza rispetto al punto massimo.
- **Sum.** La somma o il totale di tutti i valori non mancanti di tutti i casi.
- **Somma cumulata.** La somma cumulata dei valori. Ogni elemento grafico mostra la somma di un sottogruppo più la somma totale di tutti i gruppi precedenti.
- **Somma percentuale.** La percentuale all'interno di ogni sottogruppo basata su un campo sommato confrontato alla somma di tutti i gruppi.
- **Somma percentuale cumulata.** La percentuale cumulativa all'interno di ogni sottogruppo basata su un campo sommato confrontato alla somma di tutti i gruppi. Ogni elemento grafico mostra la percentuale di un sottogruppo più la percentuale totale di tutti i gruppi precedenti.
- **Variance.** Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.
- **Deviazione standard.** La radice quadrata della varianza. La deviazione standard è una misura della dispersione intorno alla media espressa nella stessa unità di misura delle osservazioni. In una distribuzione normale, il 68% dei casi rientra in una deviazione standard della media e il 95% dei casi rientra in due deviazioni standard. Se, ad esempio, in una popolazione con distribuzione normale l'età media fosse 45 e la deviazione standard 10, il 95% dei casi cadrebbe fra 25 e 65 anni.
- **Errore standard.** Una misura di quanto il valore di una statistica del test può variare da campione a campione. È la deviazione standard della distribuzione di campionamento di una statistica. L'errore standard della media, ad esempio, è la deviazione standard delle medie campionarie.
- **Curtosi.** Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che, rispetto a una distribuzione normale, le osservazioni sono più raggruppate intorno al centro della distribuzione e hanno code più sottili fino ai valori estremi della distribuzione; a quel punto, le code della distribuzione leptocurtica sono più spesse rispetto a una distribuzione normale. Una curtosi negativa indica che, rispetto a una distribuzione normale, le osservazioni sono meno raggruppate e hanno code più spesse fino ai valori estremi della distribuzione; a

quel punto, le code della distribuzione platicurtica sono più sottili rispetto a una distribuzione normale.

- **Asimmetria.** Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Le seguenti statistiche di regione potrebbero restituire più elementi grafici per ogni sottogruppo. Quando si utilizzano gli elementi grafici intervallo, area o limite, una statistica di regione restituisce un elemento grafico che mostra l'intervallo. Tutti gli altri elementi grafici danno come risultato due elementi separati, uno che mostra l'inizio dell'intervallo e l'altro che mostra la fine dell'intervallo.

- **Regione: Intervallo.** L'intervallo di valori tra il valore minimo e il valore massimo.
- **Regione: Intervallo di confidenza per la media al 95%.** Un intervallo di valori con il 95% di probabilità di includere la media della popolazione.
- **Regione: Intervallo di confidenza per il singolo al 95%.** Un intervallo di valori con il 95% di probabilità di includere il valore previsto dato il caso singolo.
- **Regione: 1 Deviazione standard sopra/sotto la media.** Un intervallo di valori compreso tra 1 deviazione standard sopra e sotto la media.
- **Regione: 1 Errore standard sopra/sotto la media.** Un intervallo di valori compreso tra 1 errore standard sopra e sotto la media.

Statistiche riassuntive basate sul conteggio

- **Conteggio.** Il numero di righe/casi.
- **Conteggio cumulato.** Il numero cumulato di righe/casi. Ogni elemento grafico mostra il conteggio riferito a un sottogruppo più il conteggio totale di tutti i gruppi precedenti.
- **Percentuale del conteggio.** La percentuale di righe/casi in ogni sottogruppo rispetto al numero totale di righe/casi.
- **Percentuale cumulata del conteggio.** La percentuale cumulata di righe/casi in ogni sottogruppo rispetto al numero totale di righe/casi. Ogni elemento grafico mostra la percentuale di un sottogruppo più la percentuale totale di tutti i gruppi precedenti.

Come specificare il modificatore di collisione

Il modificatore di collisione determina cosa accade in caso di sovrapposizione tra elementi grafici.

- ▶ Selezionare l'elemento grafico per il quale si desidera specificare il modificatore di collisione.
- ▶ Fare clic sulla scheda Elemento nella tavolozza delle proprietà.
- ▶ Dall'elenco a discesa Modificatore selezionare un modificatore di collisione. Con l'impostazione -automatico-, è l'applicazione a stabilire il modificatore di collisione adatto al tipo e alle statistiche dell'elemento grafico.

Sovrapposto. Consente di disegnare elementi grafici uno sopra l'altro quando hanno lo stesso valore.

Sovrapposti. Consente di impilare elementi grafici che sarebbero normalmente sovrapposti quando presentano gli stessi valori dei dati.

Alterna. Consente di spostare gli elementi grafici accanto ad altri con lo stesso valore anziché sovrapporli. Gli elementi grafici vengono disposti in modo simmetrico. In altre parole, vengono spostati dalla parte opposta rispetto a un punto centrale. L'alternanza è molto simile al raggruppamento.

Impila. Consente di spostare gli elementi grafici accanto ad altri con lo stesso valore anziché sovrapporli. Gli elementi grafici vengono disposti in modo asimmetrico. In altri termini, gli elementi grafici vengono impilati uno sopra l'altro, con quello in basso posizionato in corrispondenza di uno specifico valore sulla scala.

Jitter (normale). Consente di riposizionare in modo casuale gli elementi grafici in corrispondenza di uno stesso valore dei dati utilizzando una distribuzione normale.

Jitter (uniforme). Consente di riposizionare in modo casuale gli elementi grafici in corrispondenza di uno stesso valore dei dati utilizzando una distribuzione uniforme.

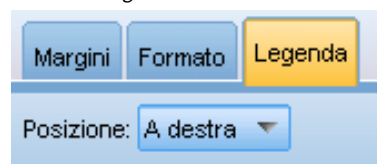
Modifica della posizione della legenda

Se nel grafico è presente una legenda, quest'ultima di solito viene visualizzata sulla destra. Se necessario, tale posizione può essere modificata.

Come modificare la posizione della legenda

- ▶ Selezionare la legenda.
- ▶ Fare clic sulla scheda Legenda nella tavolozza delle proprietà.

Figura 5-94
Scheda Legenda



- ▶ Selezionare una posizione.

Copia di una visualizzazione e dei dati delle visualizzazioni

La tavolozza Generale comprende i pulsanti per la copia della visualizzazione e dei relativi dati.

Figura 5-96
Pulsante Copia visualizzazione



Copia visualizzazione. La visualizzazione viene copiata negli Appunti come immagine (sono disponibili più formati). Quando si incolla l'immagine in un'altra applicazione, è possibile selezionare un'opzione di "incolla" speciale che consente di scegliere il formato da incollare.

Figura 5-97
Pulsante Copia dati visualizzazione



Copia dati visualizzazione. Vengono copiati i dati sottostanti utilizzati per generare la visualizzazione. I dati vengono copiati negli Appunti come solo testo o testo HTML. Quando si incollano i dati in un'altra applicazione, è possibile selezionare un'opzione di "incolla" speciale che consente di scegliere il formato da incollare.

Tasti di scelta rapida

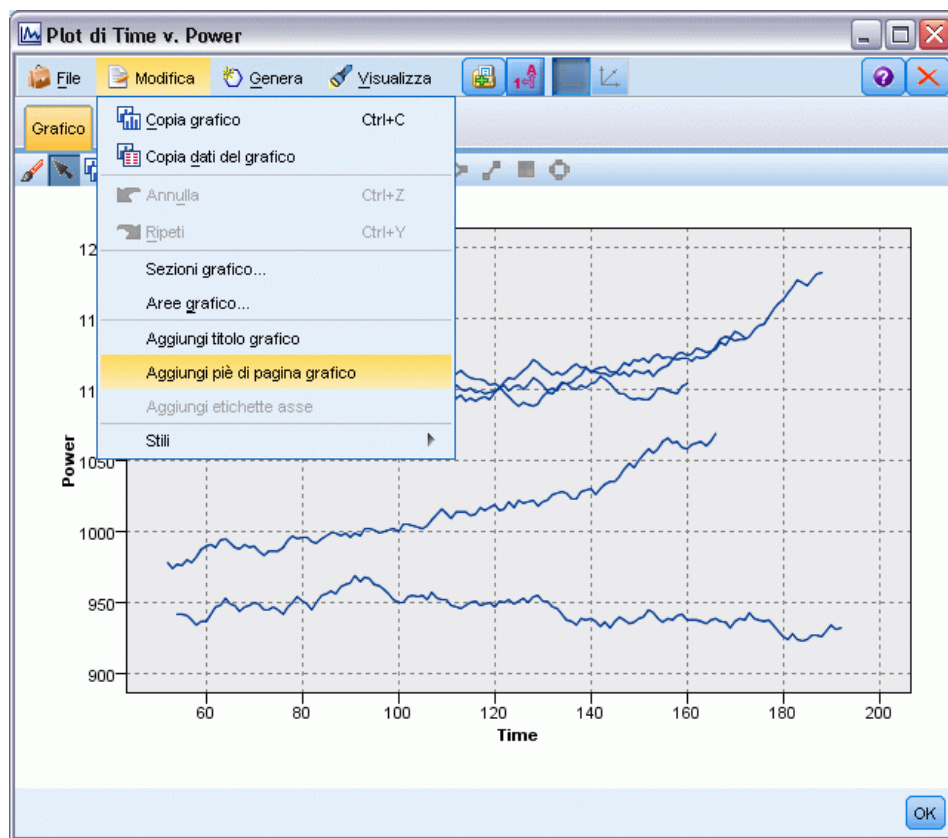
Tabella 5-3
Tasti di scelta rapida

Tasto di scelta rapida	Funzione
Ctrl+Barra spaziatrice	Consente di passare dalla modalità Esplora alla modalità Modifica e viceversa
Canc	Consente di eliminare un oggetto della visualizzazione
Ctrl+Z	Annulla
Ctrl+Y	Ripeti
F2	Visualizza la struttura per selezionare gli elementi di un grafico

Aggiunta di titoli e annotazioni

È possibile aggiungere titoli, note a piè di pagina o etichette degli assi a qualsiasi tipo di grafico in modo da facilitare l'identificazione dei dati visualizzati.

Figura 5-98
 Aggiunta di una nota a piè di pagina al grafico



Aggiunta di titoli ai grafici

- ▶ Dai menu, scegliere Modifica > Aggiungi titolo grafico. Sopra il grafico viene visualizzata una casella di testo contenente <TITOLO>.
- ▶ Controllare di essere in modalità di modifica. Dai menu, scegliere Visualizza > Modalità Modifica.
- ▶ Fare doppio clic sul testo <TITOLO>.
- ▶ Immettere il titolo richiesto e premere Invio.

Aggiunta di note a piè di pagina ai grafici

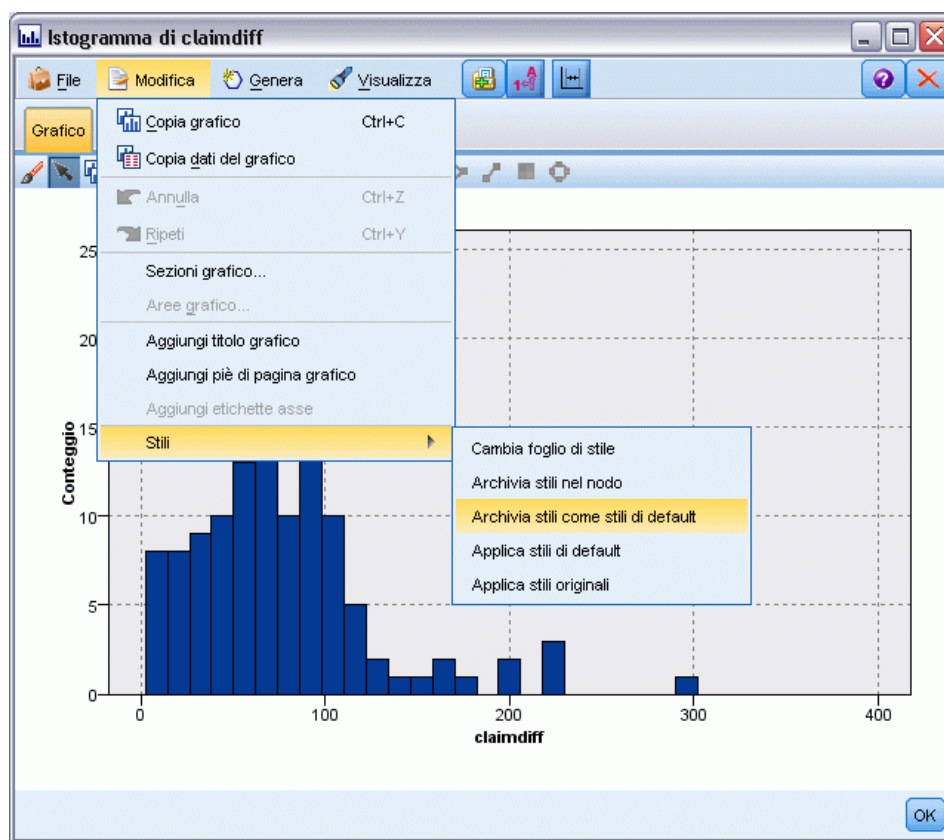
- ▶ Dai menu, scegliere Modifica > Aggiungi piè di pagina grafico. Sotto il grafico viene visualizzata una casella di testo contenente <PIÈ DI PAGINA>.
- ▶ Controllare di essere in modalità di modifica. Dai menu, scegliere Visualizza > Modalità Modifica.
- ▶ Fare doppio clic sul testo <PIÈ DI PAGINA>.
- ▶ Immettere il titolo richiesto e premere Invio.

Utilizzo dei fogli di stile dei grafici

Le informazioni di base sulla visualizzazione dei grafici, quali colori, caratteri, simboli e spessore delle righe sono controllate da un foglio di stile. Con IBM® SPSS® Modeler viene fornito un foglio di stile di default; tuttavia, se necessario, è possibile modificarlo. Per esempio, se esiste uno schema di colori aziendale specifico per le presentazioni, potrebbe essere necessario utilizzarlo anche nei grafici. [Per ulteriori informazioni, vedere l'argomento Modifica delle visualizzazioni a pag. 321.](#)

Nei nodi Grafico è possibile utilizzare la modalità di modifica per cambiare lo stile dell'aspetto di un grafico. Tramite il menu Modifica > Stili è possibile quindi salvare le modifiche come un foglio di stile da applicare a tutti i grafici generati in seguito dal nodo Grafico corrente o da adottare come nuovo foglio di stile di default per tutti i grafici generati con SPSS Modeler.

Figura 5-99
Selezione degli stili del grafico



Dall'opzione Stili del menu Modifica sono disponibili cinque diverse opzioni per i fogli di stile:

- **Alterna foglio di stile.** Visualizza un elenco di diversi fogli di stile archiviati che possono essere selezionati per modificare l'aspetto dei grafici. [Per ulteriori informazioni, vedere l'argomento Applicazione di fogli di stile a pag. 342.](#)

- **Archivia stili nel nodo.** Consente di archiviare le modifiche apportate agli stili del grafico selezionato in modo che vengano applicate a tutti i grafici creati successivamente dallo stesso nodo Grafico nello stream corrente.
- **Archivia stili come stili di default.** Consente di archiviare le modifiche apportate agli stili del grafico selezionato in modo che vengano applicate a tutti i grafici creati successivamente da qualsiasi nodo Grafico in qualsiasi stream. Dopo aver selezionato questa opzione, è possibile utilizzare *Applica stili di default* per applicare gli stessi stili a tutti gli altri grafici esistenti.
- **Applica stili di default.** Consente di modificare gli stili del grafico selezionato applicando gli stili attualmente salvati come stili di default.
- **Applica stili originali.** Consente di ripristinare gli stili del grafico forniti come stili di default originali.

Applicazione di fogli di stile

È possibile applicare un foglio di stile di visualizzazione che specifichi le proprietà stilistiche della visualizzazione. Fra le altre opzioni, il foglio di stile può definire i caratteri, i tratteggi e i colori. In un certo senso, i fogli di stile rappresentano una “scorciatoia” per le modifiche che sarebbe necessario eseguire manualmente. Si noti, tuttavia, che un foglio di stile consente di eseguire solo le modifiche di *stile*. Le modifiche di altro tipo, come la posizione della legenda o l’intervallo della scala, non vengono memorizzate nel foglio di stile.

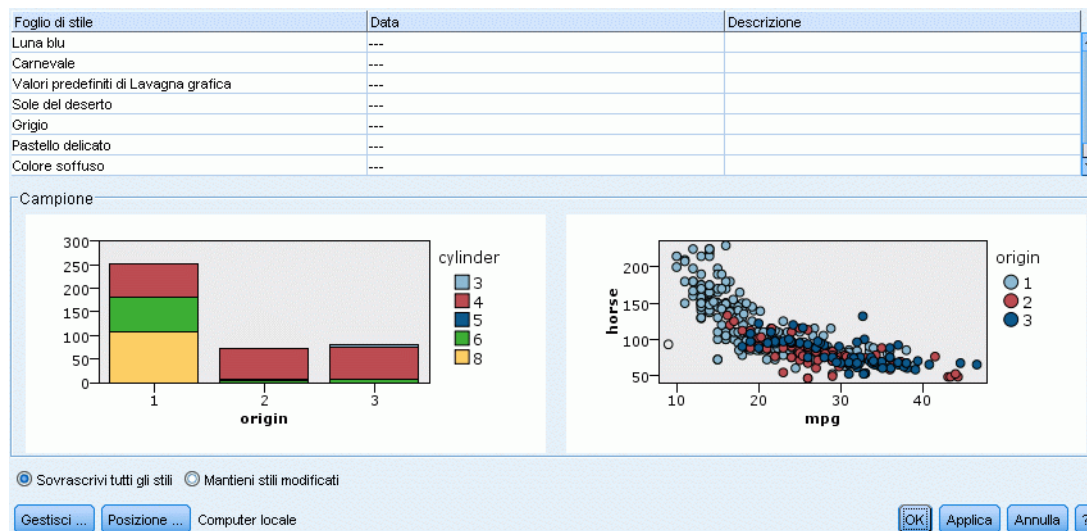
Come applicare un foglio di stile

- ▶ Dai menu, scegliere:
Modifica > Stili > Alterna foglio di stile
- ▶ Per selezionare un foglio di stile, utilizzare la finestra di dialogo *Alterna foglio di stile*.
- ▶ Fare clic su *Applica* per applicare il foglio di stile alla visualizzazione senza chiudere la finestra. Fare clic su *OK* per applicare il foglio di stile e chiudere la finestra di dialogo.

Finestra di dialogo Alterna/Seleziona foglio di stile

Figura 5-100

Finestra di dialogo Alterna foglio di stile



La tabella nella parte superiore della finestra di dialogo elenca tutti i fogli di stile di visualizzazione attualmente disponibili. Alcuni fogli di stile sono preinstallati, mentre altri potrebbero essere stati creati in IBM® SPSS® Visualization Designer (un prodotto separato).

La parte inferiore della finestra di dialogo mostra alcuni esempi di visualizzazione contenenti dei dati campione. Selezionare uno dei fogli di stile in modo da applicare gli stili alle visualizzazioni di esempio. Questi esempi consentono di capire in che modo un determinato foglio di stile influisce effettivamente sulla visualizzazione.

Nella finestra di dialogo sono presenti anche le seguenti opzioni.

Stili esistenti. Per default, un foglio di stile può sovrascrivere tutti gli stili nella visualizzazione. Questo comportamento può essere modificato.

- **Sovrascrivi tutti gli stili.** Quando si applica il foglio di stile, consente di sovrascrivere tutti gli stili della visualizzazione, inclusi quelli modificati nella visualizzazione durante la sessione di modifica corrente.
- **Mantieni stili modificati.** Quando si applica il foglio di stile, consente di sovrascrivere solo gli stili che *non* sono stati modificati nella visualizzazione durante la sessione di modifica corrente. Al contrario, gli stili che sono stati modificati durante la sessione corrente vengono mantenuti.

Gestisci. Gestire modelli e i fogli di stile di visualizzazione sul computer. È possibile importare, esportare, rinominare ed eliminare modelli e i fogli di stile di visualizzazione sul computer locale. [Per ulteriori informazioni, vedere l'argomento Gestione dei modelli, dei fogli di stile e dei file delle mappe a pag. 251.](#)

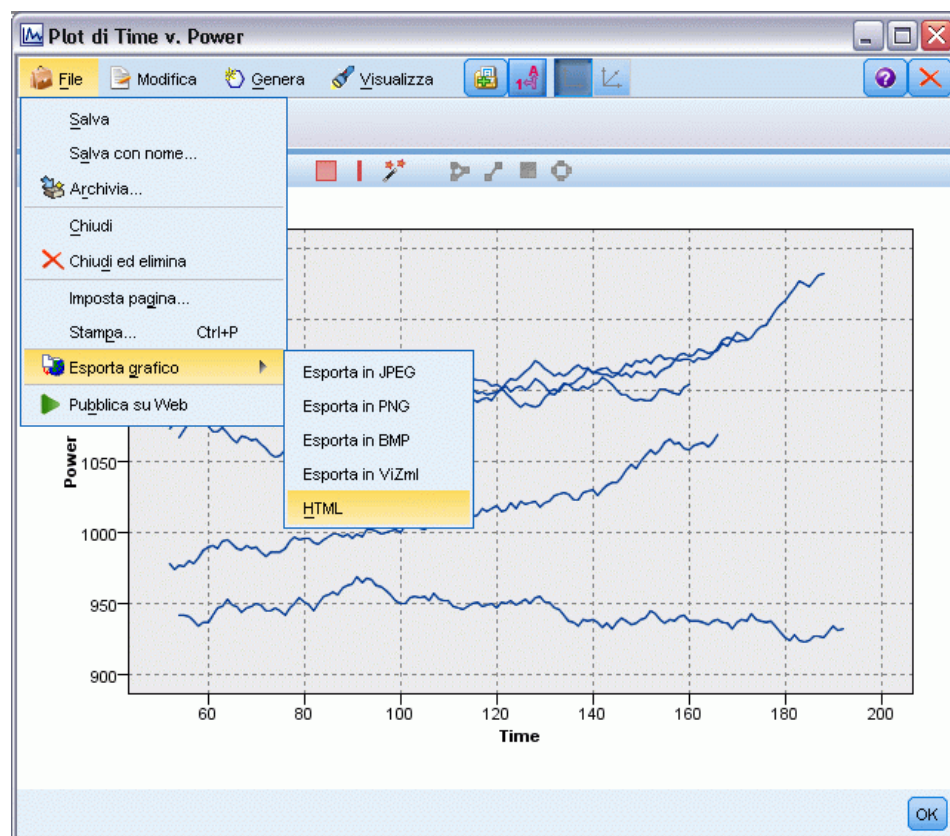
Posizione. Modificare la posizione in cui sono archiviati i modelli e i fogli di stile di visualizzazione. La posizione corrente viene elencata alla destra del pulsante. [Per ulteriori informazioni, vedere l'argomento Impostazione della posizione di modelli, fogli di stile e mappe a pag. 249.](#)

Stampa, salvataggio, copia ed esportazione di grafici

Ogni grafico dispone di numerose opzioni che ne consentono il salvataggio, la stampa o l'esportazione in un altro formato. La maggior parte di queste opzioni è accessibile dal menu File. Inoltre, nel menu Modifica è possibile scegliere di copiare il grafico o i dati contenuti al suo interno in modo da poterli utilizzare in un'altra applicazione.

Figura 5-101

Menu File e barra degli strumenti relativi alle finestre dei grafici



Stampa

- Per stampare il grafico, utilizzare l'opzione di menu o il pulsante Stampa. Prima di stampare è possibile impostare le opzioni di stampa e visualizzare un'anteprima del grafico da stampare mediante i comandi Imposta pagina e Anteprima di stampa.

Salvataggio di grafici

- ▶ Per salvare il grafico in un file di output di IBM® SPSS® Modeler (*.cou), scegliere Salva o Salva con nome dal menu File.

or

Per salvare il grafico nel repository, scegliere Memorizza output dal menu File.

Copia di grafici

- ▶ Per copiare il grafico in modo da utilizzarlo in un'altra applicazione, per esempio MS Word o MS PowerPoint, scegliere Copia grafico dal menu Modifica.

Copia di dati

- ▶ Per copiare i dati in modo da utilizzarli in un'altra applicazione, per esempio MS Excel o MS Word, scegliere Modifica > Copia dati del grafico. Per default, i dati saranno copiati in formato HTML. Utilizzare Incolla speciale... nell'altra applicazione per visualizzare le altre opzioni di formattazione quando si incollano i dati.

Esportazione di grafici

L'opzione Esporta grafico consente all'utente di esportare il grafico in uno dei seguenti formati: Bitmap (.bmp), JPEG (.jpg), PNG (.png), HTML (.html) o documento ViZml (.xml) per utilizzarlo in altre applicazioni IBM® SPSS® Statistics

- ▶ Per esportare i grafici, selezionare File > Esporta grafico e quindi scegliere il formato desiderato.

Esportazione di tabelle

L'opzione Esporta tabella consente all'utente di esportare la tabella in uno dei seguenti formati: delimitato da tabulazione (.tab), delimitato da virgola (.csv) o HTML (.html)

- ▶ Per esportare le tabelle, selezionare Esporta tabella dal menu File e quindi scegliere il formato desiderato.

Nodi Output

Panoramica sui nodi Output

I nodi Output consentono di ottenere informazioni sui dati e sui modelli. Inoltre, consentono di esportare i dati in vari formati per l'interfacciamento con gli altri strumenti software in uso.

Sono disponibili i seguenti nodi di output:



Il nodo Tabella visualizza i dati in formato tabella, che è inoltre possibile scrivere su un file. Questa funzione è utile tutte le volte che si desidera controllare i valori dei dati o esportarli in un formato di facile lettura. [Per ulteriori informazioni, vedere l'argomento Nodo Tabella a pag. 353.](#)



Il nodo Matrice crea una tabella che mostra le relazioni tra i campi. In genere viene utilizzato per mostrare le relazioni tra due campi simbolici, ma è possibile avvalersene anche per mostrare le relazioni tra campi flag o numerici. [Per ulteriori informazioni, vedere l'argomento Nodo Matrice a pag. 359.](#)



Il nodo Analisi valuta la capacità dei modelli predittivi di generare previsioni accurate. I nodi Analisi eseguono diversi confronti tra i valori previsti e i valori effettivi per uno o più insiemi di modelli. Possono inoltre confrontare i modelli predittivi fra loro. [Per ulteriori informazioni, vedere l'argomento nodo Analisi a pag. 364.](#)



Il nodo Esplora offre una prima panoramica completa dei dati, incluse statistiche riassuntive, istogrammi e distribuzione per ciascun campo, nonché informazioni su valori anomali, mancanti ed estremi. I risultati vengono visualizzati in una matrice di semplice lettura che può essere ordinata e utilizzata per generare grafici a schermo intero e nodi di preparazione dei dati. [Per ulteriori informazioni, vedere l'argomento Nodo Esplora a pag. 369.](#)



Il nodo Trasformazioni consente di selezionare e visualizzare in anteprima i risultati di trasformazioni prima di applicarli ai campi selezionati. [Per ulteriori informazioni, vedere l'argomento Nodo Trasformazioni a pag. 383.](#)



Il nodo Statistiche fornisce informazioni riassuntive di base su campi numerici. Calcola statistiche riassuntive per singoli campi e per correlazioni tra campi. [Per ulteriori informazioni, vedere l'argomento Nodo Statistiche a pag. 389.](#)



Il nodo Medie confronta le medie tra gruppi indipendenti o coppie di campi correlati per verificare se esiste una differenza significativa. Per esempio, è possibile confrontare le entrate medie prima e dopo il lancio di una promozione, oppure confrontare le entrate determinate da clienti che non hanno ricevuto la promozione con quelli che l'hanno ricevuta. [Per ulteriori informazioni, vedere l'argomento Nodo Medie a pag. 393.](#)



Il nodo Report crea report formattati che contengono sia testo fisso sia dati e altre espressioni derivate dai dati. Il formato del report viene specificato utilizzando modelli di testo per definire il testo fisso e costruzioni di output dei dati. È possibile fornire una formattazione personalizzata del testo utilizzando tag HTML nel modello e impostando apposite opzioni nella scheda Output. È possibile includere valori di dati e altro output condizionale utilizzando espressioni CLEM nel modello. [Per ulteriori informazioni, vedere l'argomento Nodo Report a pag. 400.](#)



Il nodo Globali analizza i dati e calcola i valori di riepilogo che possono essere utilizzati nelle espressioni CLEM. Per esempio, è possibile utilizzare questo nodo per calcolare le statistiche di un campo denominato *età* e utilizzare quindi la media globale dell'*età* nelle espressioni CLEM inserendo la funzione `@GLOBAL_MEAN(età)`. [Per ulteriori informazioni, vedere l'argomento Nodo Globali a pag. 403.](#)

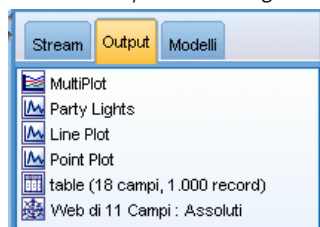
Gestione dell'output

La scheda Output di Manager mostra i diagrammi, i grafici e le tabelle generati durante una sessione di IBM® SPSS® Modeler. È sempre possibile riaprire un output semplicemente facendo doppio clic su di esso in Manager; non è necessario eseguire nuovamente il nodo o lo stream corrispondente.

Per visualizzare la scheda Output di Manager

- Aprire il menu Visualizza e scegliere Manager. Fare clic sulla scheda Output.

Figura 6-1
Scheda Output di Manager



Dalla scheda Output di Manager è possibile:

- Visualizzare gli oggetti di output esistenti, per esempio istogrammi, grafici di valutazione e tabelle.
- Rinominare gli oggetti di output.
- Salvare gli oggetti di output su disco o, se disponibile, in IBM® SPSS® Collaboration and Deployment Services Repository.
- Aggiungere file di output al progetto corrente.
- Eliminare dalla sessione corrente gli oggetti di output non salvati.
- Aprire gli oggetti di output salvati o recuperarli da IBM SPSS Collaboration and Deployment Services Repository (se disponibile).

Per accedere a queste opzioni, fare clic con il pulsante destro del mouse in un punto qualsiasi della scheda Output.

Visualizzazione dell'output

L'output su schermo viene visualizzato in una finestra di output del browser. Questa finestra dispone dei propri menu che consentono di stampare e di salvare l'output o di esportarlo in un altro formato. Si noti che le opzioni specifiche possono variare a seconda del tipo di output.

Stampa, salvataggio ed esportazione di dati. Sono disponibili ulteriori informazioni, come segue:

- Per stampare l'output, utilizzare l'opzione di menu o il pulsante Stampa. Prima di stampare è possibile impostare le opzioni di stampa e visualizzare un'anteprima del grafico da stampare mediante i comandi Imposta pagina e Anteprima di stampa.
- Per salvare l'output in un file di output di IBM® SPSS® Modeler (.cou), scegliere Salva o Salva con nome dal menu File.
- Per salvare l'output in un altro formato, per esempio testo o HTML, scegliere Esporta dal menu File. [Per ulteriori informazioni, vedere l'argomento Esportazione dell'output a pag. 351.](#)
- Per salvare l'output in un repository condiviso e consentire ad altri utenti di visualizzarlo mediante IBM® SPSS® Collaboration and Deployment Services Deployment Portal, scegliere Pubblica su Web dal menu File. Questa opzione richiede una licenza distinta per IBM® SPSS® Collaboration and Deployment Services.

Selezione di celle e di colonne. Il menu Modifica contiene diverse opzioni per la selezione, la deselegione e la copia di celle e colonne, a seconda di quanto richiesto dal tipo di output corrente. [Per ulteriori informazioni, vedere l'argomento Selezione di celle e di colonne a pag. 352.](#)

Generazione di nodi. Il menu Genera consente di generare nuovi nodi in base al contenuto del browser di output. Le opzioni variano in base al tipo di output e agli elementi attualmente selezionati nell'output. Per informazioni sulle opzioni di generazione dei nodi di un tipo specifico di output, consultare la documentazione relativa al quel tipo di output.

Pubblica sul Web

L'opzione Pubblica su Web permette di pubblicare alcuni tipi di output degli stream in un IBM® SPSS® Collaboration and Deployment Services Repository centralizzato condiviso che costituisce la base di IBM® SPSS® Collaboration and Deployment Services. Se si utilizza questa opzione, gli altri utenti che desiderano visualizzare l'output possono farlo se dispongono di un accesso a Internet e di un account IBM SPSS Collaboration and Deployment Services. Non è necessario che abbiano installato IBM® SPSS® Modeler.

Nota: per l'accesso a un repository IBM SPSS Collaboration and Deployment Services è necessaria una licenza distinta. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>

La tabella riportata di seguito elenca i nodi SPSS Modeler che supportano la funzione Pubblica su Web. L'output di questi nodi viene archiviato nel IBM SPSS Collaboration and Deployment Services Repository in formato oggetto di output (.cou) e può essere visualizzato direttamente in IBM® SPSS® Collaboration and Deployment Services Deployment Portal.

Gli altri tipi di output si possono visualizzare solo se sul computer dell'utente è installata la rispettiva applicazione (per esempio SPSS Modeler per gli oggetti stream).

Tabella 6-1
Nodi che supportano la funzione *Pubblica su Web*

Tipo di nodo	Nodo
Grafici	all
Output	Tabella
	Matrice
	Comandi di manipolazione dei nodi
	Trasforma
	Medie
	Analisi
	Statistiche
	Report (HTML)
IBM® SPSS® Statistics	Output Statistics

Pubblicazione dell'output sul Web

Per pubblicare l'output su Web:

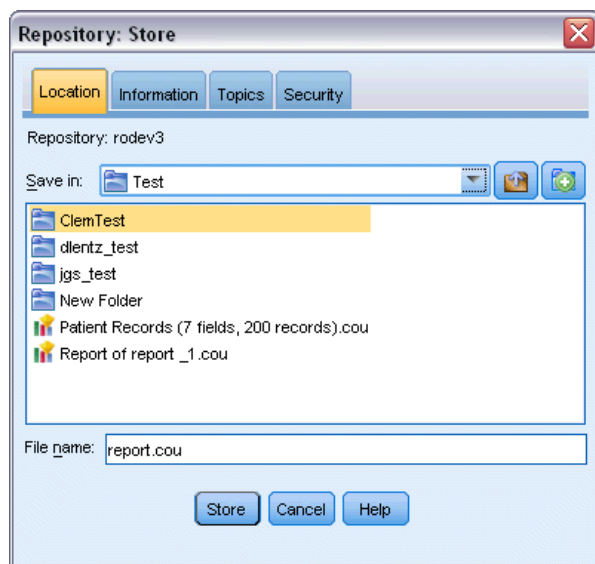
- ▶ Eseguire uno dei nodi elencati nella tabella in uno stream di IBM® SPSS® Modeler. Viene così creato un oggetto di output (per esempio un oggetto tabella, matrice o report) in una nuova finestra.
- ▶ Dalla finestra dell'oggetto di output scegliere:
File > Pubblica sul Web

Nota: se si desidera solo esportare semplici file HTML da utilizzare con un browser Web standard, scegliere *Esporta* dal menu *File* e selezionare *HTML*.

- ▶ Connessione a IBM® SPSS® Collaboration and Deployment Services Repository [Per ulteriori informazioni, vedere l'argomento Connessione a IBM SPSS Collaboration and Deployment Services Repository in il capitolo 9 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Una volta stabilita la connessione, viene visualizzata la finestra di dialogo *Repository: Archivia*, che offre diverse opzioni di archiviazione. [Per ulteriori informazioni, vedere l'argomento Archiviazione di oggetti in IBM SPSS Collaboration and Deployment Services Repository in il capitolo 9 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Figura 6-2
Finestra di dialogo Repository: Archivia



- Dopo aver scelto le opzioni di archiviazione desiderate, fare clic su Archivia.

Visualizzazione dell'output pubblicato sul Web

Per usufruire di questa funzione è necessario disporre di un account IBM SPSS Collaboration and Deployment Services. Se si dispone dell'applicazione appropriata per visualizzare il tipo di oggetto desiderato (per esempio IBM® SPSS® Modeler o IBM® SPSS® Statistics), l'output viene visualizzato all'interno dell'applicazione anziché nel browser.

Nota: per l'accesso a IBM® SPSS® Collaboration and Deployment Services è necessaria una licenza distinta. Per ulteriori informazioni, vedere <http://www.ibm.com/software/analytics/spss/products/deployment/cds/>.

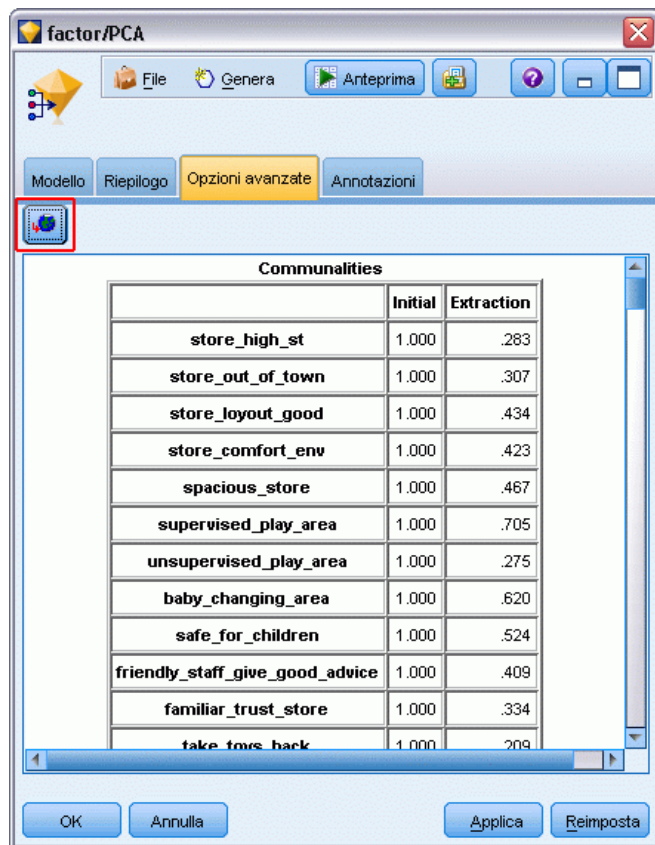
Per visualizzare l'output pubblicato sul Web:

- Dal browser, accedere a `http://<host_repos>:<port_repos>/peb` dove `host_repos` e `port_repos` sono il nome host e il numero di porta dell'host IBM SPSS Collaboration and Deployment Services.
- Digitare le credenziali di accesso per l'account IBM SPSS Collaboration and Deployment Services.
- Fare clic su Repository contenuti.
- Spostarsi sull'oggetto da visualizzare o cercarlo.
- Fare clic sul nome dell'oggetto. Per alcuni tipi di oggetti, per esempio i grafici, la visualizzazione nel browser può non essere immediata.

Visualizzazione dell'output in un browser HTML

Nella scheda Opzioni avanzate degli insiemi di modelli Lineare, Logistico e Fattoriale/PCA è possibile visualizzare le informazioni in un browser separato, per esempio Internet Explorer. Le informazioni vengono prodotte in formato HTML, il che consente di salvarle e riutilizzarle altrove, per esempio in un'intranet aziendale o in un sito Internet.

Figura 6-3
Pulsante di avvio sulla scheda Opzioni avanzate di un insieme di modelli



Per visualizzare le informazioni in un browser, fare clic sul pulsante di avvio sotto l'icona del modello nell'angolo superiore sinistro della scheda Opzioni avanzate dell'insieme di modelli.

Esportazione dell'output

Nella finestra di output del browser è possibile scegliere di esportare l'output in un altro formato, come testo o HTML. I formati di esportazione variano a seconda del tipo di output, ma solitamente sono simili alle opzioni relative al tipo di file disponibili quando si seleziona Salva nel file nel nodo utilizzato per generare l'output.

Per esportare l'output

- ▶ Nel browser di output, aprire il menu File e scegliere Esporta. Quindi, scegliere il tipo di file che si desidera creare:
 - **Delimitato da tabulazione (*.tab)** Questa opzione genera un file di testo formattato che contiene i valori dei dati. Questo stile spesso è utile per generare una rappresentazione in testo normale delle informazioni che possono essere importate in altre applicazioni. Questa opzione è disponibile per i nodi Tabella, Matrice e Medie.
 - **Delimitato da virgola (*.dat)**. Questa opzione genera un file di testo delimitato da virgola contenente i valori dei dati. Questo stile è spesso utile per generare rapidamente un file di dati che può essere importato in fogli di calcolo o in altre applicazioni di analisi dei dati. Questa opzione è disponibile per i nodi Tabella, Matrice e Medie.
 - **Dati trasposti delimitati da tabulazione (*.tab)** Questa opzione è identica all'opzione Delimitato da tabulazione, tranne per il fatto che i dati vengono trasposti in modo che le righe rappresentino i campi e le colonne rappresentino i record.
 - **Dati trasposti delimitati da virgola (*.dat)**. Questa opzione è identica all'opzione Delimitato da virgola, tranne per il fatto che i dati vengono trasposti in modo che le righe rappresentino i campi e le colonne rappresentino i record.
 - **HTML (*.html)**. Questa opzione scrive in uno o più file l'output con formattazione HTML.

Selezione di celle e di colonne

Figura 6-4
browser del nodo Tabella

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimt
1	id602	name602	north	1780	42	9	734118.000	maize	arable
2	id606	name606	southeast	1580	42	7	445785.000	maize	arable
3	id607	name607	southeast	1820	29	6	211605.000	maize	arable
4	id608	name608	southeast	1640	108	7	1167040.0...	maize	arable
5	id610	name610	southeast	600	80	6	267928.000	wheat	arable
6	id611	name611	southeast	980	38	6	222703.000	maize	arable
7	id613	name613	southeast	440	86	3	115544.000	potatoes	arable
8	id614	name614	southeast	1260	90	8	900243.000	maize	arable
9	id616	name616	midlands	1660	36	9	490617.000	rapeseed	arable
10	id620	name620	north	880	74	6	426988.000	rapeseed	arable
11	id621	name621	southwest	1160	105	4	299274.000	maize	arable
12	id622	name622	southeast	1500	61	7	687736.000	wheat	arable
13	id623	name623	southeast	1260	17	8	170279.000	maize	arable
14	id626	name626	midlands	1580	109	8	1286430.0...	wheat	arable
15	id627	name627	southeast	500	93	3	102720.000	rapeseed	arable
16	id628	name628	southeast	880	15	5	70439.800	wheat	arable
17	id630	name630	midlands	680	81	4	221391.000	potatoes	arable
18	id636	name636	southeast	1160	21	8	185939.000	potatoes	arable
19	id637	name637	midlands	940	106	6	622450.000	maize	arable
20	id638	name638	midlands	1480	64	6	586185.000	wheat	arable

Alcuni nodi, come i nodi Tabella, Matrice e Medie, generano output tabulare. Queste tabelle di output possono essere visualizzate e manipolate in modi simili, che comprendono la selezione di celle, la copia di una tabella o di una parte di essa negli Appunti, la generazione di nuovi nodi basati sulla selezione corrente e il salvataggio e la stampa della tabella.

Selezione di celle. Per selezionare una cella, fare clic su di essa. Per selezionare le celle incluse in un'area rettangolare, fare clic su un angolo dell'area desiderata, trascinare la selezione sull'angolo opposto dell'area e quindi rilasciare il pulsante del mouse. Per selezionare un'intera colonna, fare clic sulla relativa intestazione. Per selezionare più colonne, utilizzare il metodo Maiusc-clic o Ctrl-clic sulle relative intestazioni.

Quando si esegue una nuova selezione, la selezione precedente verrà cancellata. Per aggiungere una nuova selezione a una selezione esistente evitando di cancellare la selezione precedente, è sufficiente eseguire la selezione tenendo premuto il tasto Ctrl. È possibile utilizzare questo metodo per selezionare più aree non contigue della tabella. Il menu Modifica contiene anche le opzioni Seleziona tutto e Cancella selezione.

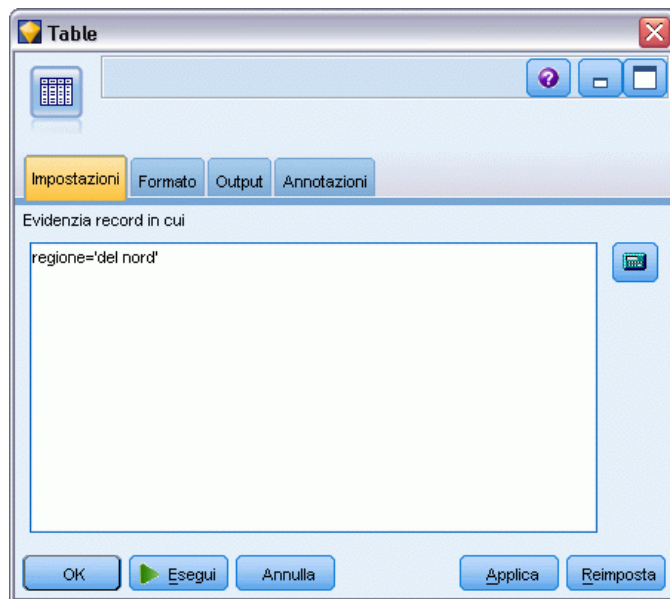
Riordinamento delle colonne. I browser di output dei nodi Tabella e Medie consentono di spostare le colonne nella tabella facendo clic sull'intestazione delle colonne e trascinandole nelle posizioni desiderate. È possibile spostare una sola colonna per volta.

Nodo Tabella

Il nodo Tabella crea una tabella contenente i valori dei dati in esame. Sono compresi tutti i campi e tutti i valori dello stream, per agevolare l'analisi dei valori dei dati o la loro esportazione in un formato facilmente leggibile. Se lo si desidera, è possibile evidenziare i record che soddisfano una determinata condizione.

Scheda Impostazioni del nodo Tabella

Figura 6-5
Nodo Tabella: scheda Impostazioni



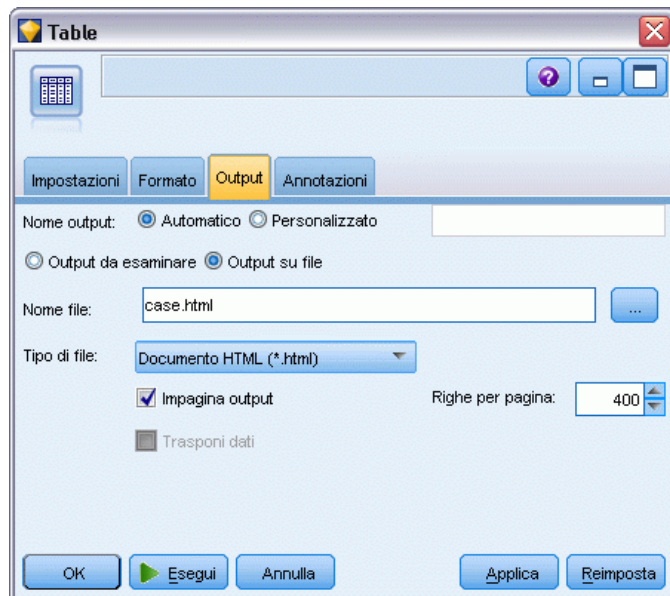
Evidenzia record in cui. Per evidenziare i record presenti nella tabella, immettere un'espressione CLEM che sia vera per tutti i record da evidenziare. Questa opzione è attivata solo se è selezionato Output su schermo.

Scheda Formato del nodo Tabella

La scheda Formato contiene le opzioni di definizione della formattazione per ognuno dei campi ed è condivisa con il nodo Tipo. [Per ulteriori informazioni, vedere l'argomento Scheda per l'impostazione dei formati dei campi a pag. 141.](#)

Scheda Output del nodo Output

Figura 6-6
Scheda Output del nodo Output



Per i nodi che generano output in stile tabella, la scheda Output consente di specificare il formato e la posizione dei risultati.

Nome output. Specifica il nome dell'output generato durante l'esecuzione del nodo. Automatico sceglie un nome in base al nodo che genera l'output. Se necessario, è possibile selezionare Personalizzato per specificare un nome diverso.

Output su schermo (impostazione di default). Crea un nuovo oggetto di output da visualizzare in linea. L'oggetto di output verrà visualizzato nella scheda Output della finestra dei manager dopo l'esecuzione del nodo di output.

Output su file. Salva l'output in un file dopo l'esecuzione del nodo. Se si sceglie questa opzione, immettere un nome di file (oppure passare a una directory e specificare un nome di file utilizzando il pulsante di selezione dei file) e selezionare un tipo di file. Si noti che alcuni tipi di file potrebbero non essere disponibili per determinati tipi di output.

I dati vengono prodotti nel formato di codifica di default del sistema, specificato nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. [Per ulteriori informazioni, vedere l'argomento Supporto Unicode in IBM SPSS Modeler in l'appendice B in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

- **Dati (delimitati da tabulazione) (*.tab)** Questa opzione genera un file di testo formattato che contiene i valori dei dati. Questo stile spesso è utile per generare una rappresentazione in testo normale delle informazioni che possono essere importate in altre applicazioni. Questa opzione è disponibile per i nodi Tabella, Matrice e Medie.

- **Dati (delimitati da virgola) (*.dat).** Questa opzione genera un file di testo delimitato da virgola contenente i valori dei dati. Questo stile è spesso utile per generare rapidamente un file di dati che può essere importato in fogli di calcolo o in altre applicazioni di analisi dei dati. Questa opzione è disponibile per i nodi Tabella, Matrice e Medie.
- **HTML (*.html).** Questa opzione scrive in uno o più file l'output con formattazione HTML. Nel caso di output tabulare (dai nodi Tabella, Matrice o Medie), un insieme di file HTML include un riquadro contenente l'elenco dei nomi dei campi e i dati in una tabella HTML. Se il numero di righe presenti nella tabella è superiore alla specifica Righe per pagina, la tabella può essere suddivisa in più file HTML. In questo caso, il riquadro del contenuto contiene i collegamenti a tutte le pagine della tabella e offre uno strumento per l'esplorazione della tabella. Se l'output non è di tipo tabulare, viene creato un solo file HTML contenente i risultati del nodo.
Nota: Se l'output HTML contiene formattazione solo per la prima pagina, selezionare Impagina output e modificare la specifica Righe per pagina in modo da includere tutto l'output in una singola pagina. Se invece il modello di output per nodi quali il nodo Report contiene tag HTML personalizzati, assicurarsi di specificare il tipo di formato Personalizzato.
- **File di testo (*.txt).** Questa opzione genera un file di testo contenente l'output. Questo stile è spesso utile per generare un output che può essere importato in altre applicazioni, per esempio elaboratori di testo o software per presentazioni. Questa opzione non è disponibile per alcuni nodi.
- **Oggetto di output (*.cou).** Gli oggetti di output salvati in questo formato possono essere aperti e visualizzati in IBM® SPSS® Modeler, aggiunti ai progetti, nonché pubblicati e registrati mediante IBM® SPSS® Collaboration and Deployment Services Repository.

Visualizzazione output. Per il nodo Medie è possibile specificare se per default deve essere visualizzato l'output semplice o avanzato. Si noti che quando si visualizza l'output generato è anche possibile passare da un tipo di visualizzazione all'altro. [Per ulteriori informazioni, vedere l'argomento Browser di output del nodo Medie a pag. 397.](#)

Formato. Per il nodo Report, è possibile scegliere se l'output deve essere formattato automaticamente oppure utilizzando la formattazione HTML inclusa nel modello. Per consentire la formattazione HTML del modello, selezionare Personalizzato.

Titolo. Per il nodo Report, è possibile specificare un titolo facoltativo che verrà visualizzato all'inizio dell'output del report.

Evidenzia testo inserito. Per il nodo Report, selezionare questa opzione per evidenziare il testo generato da espressioni CLEM nel modello Report. [Per ulteriori informazioni, vedere l'argomento Scheda Modello del nodo Report a pag. 401.](#) Questa opzione non è consigliata se si utilizza una formattazione di tipo Personalizzato.

Righe per pagina. Per il nodo Report, specificare il numero di righe da includere in ogni pagina durante la formattazione di tipo Automatico del report di output.

Trasponi dati. Questa opzione traspone i dati prima dell'esportazione in modo che le righe rappresentino i campi e le colonne rappresentino i record.

Nota: se le tabelle sono di grandi dimensioni, le opzioni sopra riportate possono risultare non completamente efficienti, soprattutto quando si utilizza un server remoto. In questi casi, si ottengono prestazioni decisamente migliori utilizzando un nodo di output File. [Per ulteriori informazioni, vedere l'argomento Nodo di esportazione File piatto in il capitolo 7 a pag. 419.](#)

Browser del nodo Tabella

Figura 6-7
browser del nodo Tabella

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimt
1	id602	name602	north	1780	42	9	734118.000	maize	arable
2	id606	name606	southeast	1580	42	7	445785.000	maize	arable
3	id607	name607	southeast	1820	29	6	211605.000	maize	arable
4	id608	name608	southeast	1640	108	7	1167040.0...	maize	arable
5	id610	name610	southeast	600	80	6	267928.000	wheat	arable
6	id611	name611	southeast	980	38	6	222703.000	maize	arable
7	id613	name613	southeast	440	86	3	115544.000	potatoes	arable
8	id614	name614	southeast	1260	90	8	900243.000	maize	arable
9	id616	name616	midlands	1660	36	9	490617.000	rapeseed	arable
10	id620	name620	north	880	74	6	426988.000	rapeseed	arable
11	id621	name621	southwest	1160	105	4	299274.000	maize	arable
12	id622	name622	southeast	1500	61	7	687736.000	wheat	arable
13	id623	name623	southeast	1260	17	8	170279.000	maize	arable
14	id626	name626	midlands	1580	109	8	1286430.0...	wheat	arable
15	id627	name627	southeast	500	93	3	102720.000	rapeseed	arable
16	id628	name628	southeast	880	15	5	70439.800	wheat	arable
17	id630	name630	midlands	680	81	4	221391.000	potatoes	arable
18	id636	name636	southeast	1160	21	8	185939.000	potatoes	arable
19	id637	name637	midlands	940	106	6	622450.000	maize	arable
20	id638	name638	midlands	1480	64	6	586185.000	wheat	arable

Il browser del nodo Tabella visualizza i dati tabulari e consente di eseguire operazioni standard quali selezionare e copiare le celle, riordinare le colonne e salvare e stampare la tabella. [Per ulteriori informazioni, vedere l'argomento Selezione di celle e di colonne a pag. 352.](#) Si tratta delle stesse operazioni che è possibile eseguire quando si visualizza l'anteprima dei dati di un nodo. [Per ulteriori informazioni, vedere l'argomento Anteprima dei dati nei nodi in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Esportazione dei dati delle tabelle. È possibile esportare i dati dal browser del nodo Tabella scegliendo:

File > Esporta

[Per ulteriori informazioni, vedere l'argomento Esportazione dell'output a pag. 351.](#)

I dati vengono esportati nel formato di codifica di default del sistema, specificato nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. [Per ulteriori informazioni, vedere l'argomento Supporto Unicode in IBM SPSS Modeler in l'appendice B in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Ricerca nella tabella. Il pulsante di ricerca (contraddistinto dall'icona con il binocolo) nella barra degli strumenti principale attiva la barra degli strumenti di ricerca, consentendo di cercare nella tabella specifici valori. È possibile cercare in avanti o all'indietro nella tabella, specificare una

ricerca con distinzione tra maiuscole e minuscole (pulsante Aa) e interrompere una ricerca in corso premendo il relativo pulsante di interruzione.

Figura 6-8

Tabella con i controlli di ricerca attivati

	id	name	region	farmsize	rainfall	landquality	farmincome	maincrop	claimt
29	id669	name669	southwest	1840	80	7	1072440.0...	wheat	arable
30	id671	name671	southeast	1020	51	5	245851.000	wheat	arable
31	id672	name672	southeast	1000	65	4	234890.000	maize	arable
32	id673	name673	midlands	900	66	6	380620.000	maize	arable
33	id675	name675	north	700	92	6	401818.000	maize	arable
34	id676	name676	southeast	740	46	7	248335.000	wheat	arable
35	id677	name677	midlands	1460	63	3	211222.000	rapeseed	arable
36	id679	name679	midlands	1380	21	8	170604.000	wheat	arable
37	id682	name682	midlands	1140	100	5	592811.000	potatoes	arable
38	id685	name685	southwest	600	48	4	108645.000	maize	arable
39	id688	name688	southwest	1480	75	3	335648.000	wheat	arable
40	id689	name689	southeast	1160	108	3	374262.000	maize	arable
41	id691	name691	southwest	920	109	9	925974.000	wheat	arable
42	id693	name693	southeast	500	76	5	181057.000	wheat	arable
43	id696	name696	southeast	1300	23	9	274389.000	maize	arable
44	id699	name699	southeast	1520	49	3	217542.000	maize	arable
45	id704	name704	southeast	1840	103	8	1588890.0...	rapeseed	arable
46	id705	name705	midlands	1800	38	7	472370.000	wheat	arable

Generazione di nodi. Il menu Genera rende disponibili operazioni di generazione dei nodi.

- **Nodo Selezione ("Record").** Genera un nodo Selezione che seleziona i record per i quali è selezionata una qualsiasi cella nella tabella.
- **Nodo Selezione ("E").** Genera un nodo Selezione che seleziona i record che contengono *tutti* i valori selezionati nella tabella.
- **Nodo Selezione ("O").** Genera un nodo Selezione che seleziona i record che contengono uno *qualsiasi* dei valori selezionati nella tabella.
- **Nuovo campo ("Record").** Genera un nodo Nuovo campo per creare un nuovo campo flag. Il campo flag contiene *T* per i record per i quali è selezionata una qualsiasi cella nella tabella e *F* per i record restanti.
- **Nuovo campo ("E").** Genera un nodo Nuovo campo per creare un nuovo campo flag. Il campo flag contiene *T* per i record che contengono *tutti* i valori selezionati nella tabella e *F* per i record restanti.
- **Nuovo campo ("O").** Genera un nodo Nuovo campo per creare un nuovo campo flag. Il campo flag contiene *T* per i record che contengono *uno qualsiasi* dei valori selezionati nella tabella e *F* per i record restanti.

Nodo Matrice

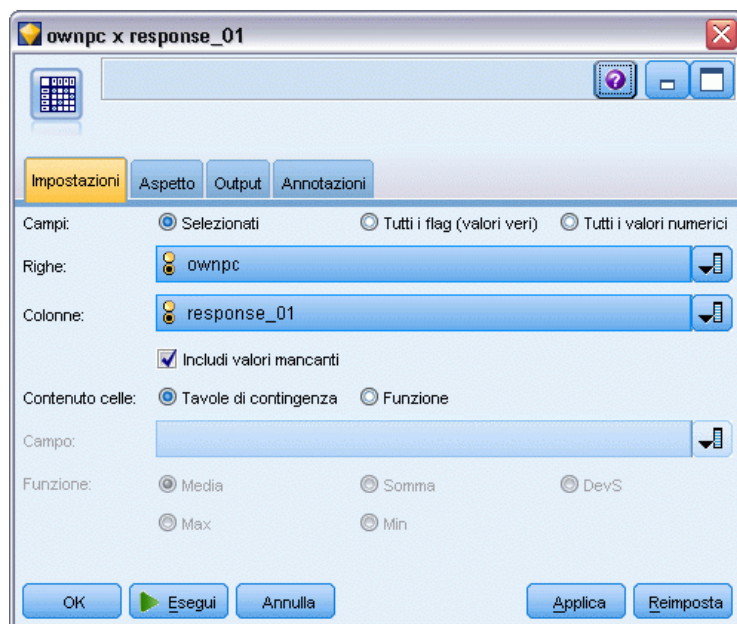
Il nodo Matrice consente di creare una tabella che mostra le relazioni tra i campi. In genere viene utilizzato per mostrare le relazioni tra due campi categoriali (flag, nominali o ordinali), ma è possibile avvalersene anche per mostrare le relazioni tra campi continui (intervallo numerico).

Scheda Impostazioni del nodo Matrice

La scheda Impostazioni permette di specificare le opzioni relative alla struttura della matrice.

Figura 6-9

Nodo Matrice: scheda Impostazioni



Campi. Selezionare una delle seguenti opzioni per specificare un tipo di selezione di campo:

- **Selezionati.** Questa opzione consente di selezionare un campo categoriale per le righe e uno per le colonne della matrice. Le righe e le colonne della matrice sono definite dall'elenco dei valori per il campo categoriale selezionato. Le celle della matrice contengono le statistiche riassuntive selezionate sotto.
- **Tutti i flag (valori veri).** Questa opzione richiede una matrice con un'unica riga e un'unica colonna per ogni campo flag nei dati. Le celle della matrice contengono i conteggi dei positivi doppi per ogni combinazione di flag. In altre parole, per una riga che corrisponde a *pane acquistato* e una colonna che corrisponde a *formaggio acquistato*, la cella che si trova nel punto di intersezione tra tale riga e tale colonna contiene il numero di record per i quali sia *pane acquistato* sia *formaggio acquistato* sono veri.
- **Tutti i valori numerici.** Questa opzione richiede una matrice con un'unica riga e un'unica colonna per ogni campo numerico. Le celle della matrice rappresentano la somma dei prodotti incrociati per la coppia di campi corrispondente. In altre parole, per ogni cella presente nella matrice, i valori per il campo riga e i valori per il campo colonna vengono moltiplicati per ogni record e poi sommati tra i vari record.

Includi valori mancanti. Consente di includere i valori mancanti definiti dall'utente (vuoti) e i valori mancanti di sistema (\$null\$) nell'output delle righe e delle colonne. Per esempio, se il valore *N/A* è stato specificato come mancante definito dall'utente per il campo della colonna selezionato, nella tabella verrà inclusa una colonna separata denominata *N/A* (a condizione che questo valore ricorra effettivamente nei dati) proprio come per tutte le altre categorie. Se l'opzione non è selezionata, la colonna *N/A* verrà esclusa, indipendentemente da quante volte ricorre il valore. [Per ulteriori informazioni, vedere l'argomento Panoramica sui valori mancanti in il capitolo 6 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Nota: l'opzione di inclusione dei valori mancanti è valida solo quando i campi selezionati sono tavole di contingenza. I valori vuoti vengono mappati a \$null\$ ed esclusi dall'aggregazione per il campo della funzione quando la modalità è Selezionati e il contenuto è impostato su Funzione e per tutti i campi numerici quando la modalità è impostata su Tutti i valori numerici.

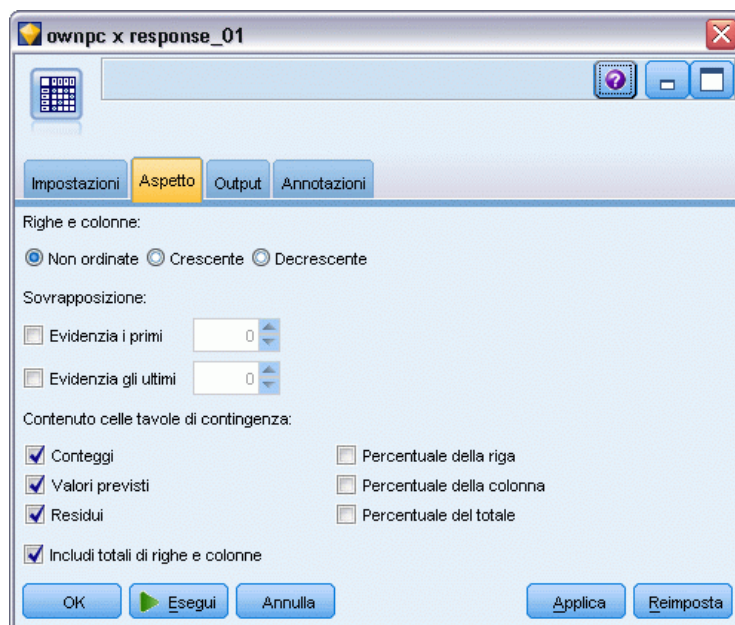
Contenuto celle. Se sono stati scelti campi Selezionati, è possibile specificare la statistica da utilizzare nelle celle della matrice. Selezionare una statistica basata sul conteggio oppure un campo di sovrapposizione per riassumere i valori di un campo numerico in base ai valori dei campi riga e colonna.

- **Tavole di contingenza.** I valori delle celle sono conteggi e/o percentuali del numero di record che includono la combinazione di valori corrispondente. È possibile specificare i riepiloghi delle tavole di contingenza desiderati, utilizzando le opzioni della scheda Aspetto. Insieme alla significatività viene visualizzato anche il valore chi-quadrato globale. [Per ulteriori informazioni, vedere l'argomento Browser di output del nodo Matrice a pag. 362.](#)
- **Funzione.** Se si seleziona una funzione riassuntiva, i valori delle celle sono rappresentati da una funzione dei valori del campo di sovrapposizione per i casi che includono i valori di riga e di colonna appropriati. Per esempio, se il campo riga è *Regione*, il campo colonna è *Prodotto* e il campo di sovrapposizione è *Entrate*, la cella di intersezione tra la riga *Nordest* e la colonna *Gadget* conterrà la somma (la media, il valore minimo o il valore massimo) delle entrate per i gadget venduti nella regione nordest. La funzione riassuntiva di default è *Media*. È possibile selezionare un'altra funzione per riassumere il campo funzione. Le opzioni includono *Media*, *Somma*, *DevS* (deviazione standard), *Max* (massimo) e *Min* (minimo).

Scheda Aspetto del nodo Matrice

La scheda Aspetto permette di controllare le opzioni di ordinamento e di evidenziazione per la matrice, nonché le statistiche presentate per le matrici delle tavole di contingenza.

Figura 6-10
Nodo Matrice: scheda Aspetto



Righe e colonne. Controlla l'ordinamento delle intestazioni di riga e colonna nella matrice. L'impostazione di default è Non ordinate. Selezionare Crescente o Decrescente per ordinare le intestazioni di riga e colonna nel modo specificato.

Sovrapposizione. Consente di evidenziare i valori estremi nella matrice. I valori sono evidenziati in base ai conteggi delle celle (per le matrici delle tavole di contingenza) o ai valori calcolati (per le matrici di funzione).

- **Evidenzia i primi.** È possibile richiedere che vengano evidenziati (in rosso) i valori massimi presenti nella matrice. Specificare il numero di valori da evidenziare.
- **Evidenzia gli ultimi.** È inoltre possibile richiedere che vengano evidenziati (in verde) i valori minimi presenti nella matrice. Specificare il numero di valori da evidenziare.

Nota: per le due opzioni di evidenziazione, i vincoli tra i numeri possono determinare l'evidenziazione di un numero di valori maggiore di quello richiesto. Per esempio, se per una matrice con sei zeri tra le celle si specifica Evidenzia gli ultimi 5, verranno evidenziati tutti e sei gli zeri.

Contenuto delle tavole di contingenza. Per le tavole di contingenza è possibile specificare le statistiche riassuntive incluse nella matrice per le matrici delle tavole di contingenza. Queste opzioni non sono disponibili quando nella scheda Impostazioni è selezionata l'opzione Tutti i valori numerici o Funzione.

- **Conteggi.** Le celle includono il numero di record contenenti il valore di riga corrispondente al valore di colonna. Questo è solo il contenuto di default delle celle.

- **Valori attesi.** Il valore previsto per il numero di record nella cella, presumendo che non esista alcuna relazione tra le righe e le colonne. I valori previsti sono basati sulla seguente formula:

$$p(\text{valore di riga}) * p(\text{valore di colonna}) * \text{numero totale di record}$$

- **Residui.** La differenza tra i valori osservati e quelli previsti.
- **Percentuale della riga.** La percentuale di tutti i record contenenti il valore di riga corrispondente al valore di colonna. Il totale delle percentuali è 100 all'interno delle righe.
- **Percentuale della colonna.** La percentuale di tutti i record contenenti il valore di colonna corrispondente al valore di riga. Il totale delle percentuali è 100 all'interno delle colonne.
- **Percentuale del totale.** La percentuale di tutti i record contenenti la combinazione di valore di colonna e valore di riga. Il totale delle percentuali è 100 nell'intera matrice.
- **Includi totali di righe e colonne.** Aggiunge alla matrice una riga e una colonna per i totali di righe e colonne.
- **Applica impostazioni.** (Solo browser di output) Consente di modificare l'aspetto dell'output del nodo Matrice senza che sia necessario chiudere e riaprire il browser di output. Apportare le modifiche in questa scheda del browser di output, fare clic su questo pulsante, quindi selezionare la scheda Matrice per visualizzare l'effetto dei cambiamenti apportati.

Browser di output del nodo Matrice

Il browser del nodo Matrice visualizza i dati delle tavole di contingenza e permette di eseguire operazioni relative alla matrice, per esempio selezionare celle, copiare negli Appunti tutta la matrice o parte di essa, generare nuovi nodi in base alla selezione della matrice, salvare e stampare la matrice. Il browser del nodo Matrice consente inoltre di visualizzare l'output di alcuni modelli, per esempio i modelli Bayes naive di Oracle.

Figura 6-11
Browser del nodo Matrice

		0	1	Totale
0	Conteggio	1611	225	1836
	Prevista	1682.510	153.490	1836
	Residui	-71.510	71.510	0
1	Conteggio	2971	193	3164
	Prevista	2899.490	264.510	3164
	Residui	71.510	-71.510	0
Totale	Conteggio	4582	418	5000
	Prevista	4582	418	5000
	Residui	0	0	0

Contenuto delle celle: tavole di contingenza dei campi (inclusi i valori mancanti)
Chi-quadrato = 57,452, df = 1, probabilità = 0

I menu File e Modifica contengono le normali opzioni di stampa, salvataggio, esportazione dell'output, selezione e copia dei dati. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output a pag. 348.](#)

Chi-quadrato. Per la tavola di contingenza di due campi categoriali, sotto alla tabella viene visualizzato anche il chi-quadrato di Pearson globale. Questo test indica la probabilità che i due campi non siano correlati in base alla differenza tra i conteggi osservati e i conteggi previsti se non esistesse alcuna relazione. Per esempio, se non esistesse alcuna relazione tra soddisfazione dei clienti e ubicazione del negozio, sarebbe normale aspettarsi percentuali di soddisfazione simili per tutti i negozi. Se però i clienti di alcuni punti vendita fanno registrare costantemente percentuali superiori rispetto agli altri, è logico sospettare che non si tratti di una coincidenza. Maggiore è la differenza, minore è la probabilità che il risultato sia dovuto semplicemente a un errore casuale di campionamento.

- Il test del chi-quadrato indica la probabilità che i due campi non siano correlati, nel qual caso le eventuali differenze tra le frequenze osservate e quelle previste sono determinate unicamente dal caso. Quando questa probabilità è estremamente bassa (in genere inferiore al 5%), la relazione fra i due campi viene definita significativa.
- In presenza di una sola colonna o di una sola riga (un test del chi-quadrato monodirezionale) i gradi di libertà sono rappresentati dal numero delle celle meno uno. Per un test del chi-quadrato bidirezionale, i gradi di libertà sono rappresentati dal numero delle righe meno uno moltiplicato per il numero delle colonne meno uno.

- La statistica del chi-quadrato deve essere interpretata con prudenza se una qualsiasi delle frequenze di cella previste è inferiore a cinque.
- Il test del chi-quadrato è disponibile solo per una tavola di contingenza di due campi. Quando nella scheda Impostazioni è selezionata l'opzione Tutti i flag o Tutti i valori numerici, il test non viene visualizzato.

Menu Genera. Il menu Genera rende disponibili operazioni di generazione dei nodi. Queste operazioni sono disponibili solo per matrici delle tavole di contingenza ed è necessario che nella matrice sia selezionata almeno una cella.

- **Nodo Seleziona.** Genera un nodo Seleziona che seleziona i record corrispondenti a qualsiasi cella selezionata nella matrice.
- **Nodo Nuovo campo (flag).** Genera un nodo Nuovo campo per creare un nuovo campo flag. Il campo flag contiene V per i record che corrispondono a qualsiasi cella selezionata nella matrice e F per i record restanti.
- **Nodo Nuovo campo (insieme).** Genera un nodo Nuovo campo per creare un nuovo campo nominale. Il campo nominale contiene una sola categoria per ogni insieme contiguo di celle selezionate nella matrice.

nodo Analisi

Il nodo Analisi consente di valutare la capacità di un modello di generare previsioni accurate. I nodi Analisi eseguono diversi confronti tra i valori previsti e i valori effettivi (il campo obiettivo) per uno o più insiemi di modelli. I nodi Analisi possono essere inoltre utilizzati per confrontare modelli predittivi con altri modelli predittivi.

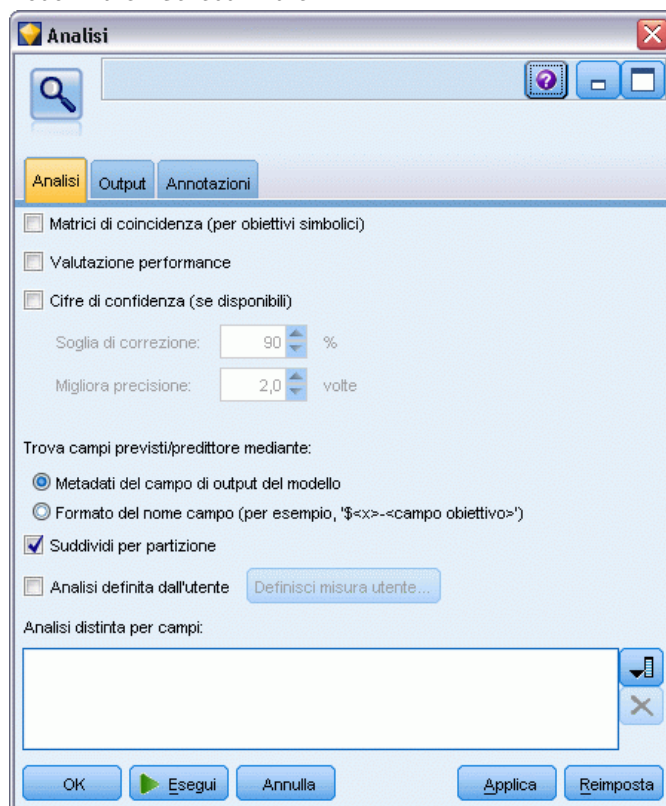
Quando si esegue un nodo Analisi, alla sezione Analisi nella scheda Riepilogo viene automaticamente aggiunto un riepilogo dei risultati dell'analisi per ogni insieme di modelli nello stream eseguito. I risultati dettagliati dell'analisi vengono visualizzati nella scheda Output della finestra dei manager, oppure possono essere scritti direttamente in un file.

Nota: poiché confrontano valori previsti e valori effettivi, i nodi Analisi sono utili solo con i modelli supervisionati (quelli che richiedono un campo obiettivo). Per i modelli non supervisionati come gli algoritmi di cluster non esistono risultati effettivi disponibili da utilizzare come base per il confronto.

Scheda Analisi del nodo Analisi

La scheda Analisi consente di specificare i dettagli dell'analisi.

Figura 6-12
Nodo Analisi: Scheda Analisi



Matrici di coincidenza (per obiettivi simbolici o categoriali). Mostra lo schema di corrispondenze tra ogni campo generato (previsto) e il campo obiettivo corrispondente per gli obiettivi categoriali (flag, nominali o ordinali). Viene visualizzata una tabella in cui le righe sono definite dai valori effettivi e le colonne dai valori predittivi, con il numero di record che presentano questo schema in ogni cella. Questa visualizzazione è utile per identificare gli errori sistematici nella previsione. Se esistono più campi generati che fanno riferimento allo stesso campo di output ma sono stati prodotti da modelli diversi, vengono conteggiate tutte le corrispondenze e le non corrispondenze di questi campi e vengono visualizzati i totali. Per i casi in cui i campi corrispondono, viene visualizzato un altro insieme di statistiche corrette/errate.

Valutazione performance. Mostra le statistiche di valutazione delle prestazioni per i modelli con output categoriali. Questa statistica, riportata per ogni categoria del campo o dei campi di output, è una misura del contenuto medio (in bit) delle informazioni del modello per la previsione dei record che appartengono a quella categoria. La statistica tiene in considerazione la difficoltà del problema di classificazione; pertanto, le previsioni più accurate per le categorie rare conseguiranno un indice di valutazione delle prestazioni più alto che le previsioni accurate per le categorie comuni. Se per una categoria il modello consente di ottenere solo semplici ipotesi, l'indice di valutazione delle prestazioni per tale categoria sarà 0.

Cifre di confidenza (se disponibili). Nel caso di modelli che generano un campo di confidenza, questa opzione consente di ottenere statistiche sui valori di confidenza e la loro relazione con le previsioni. Esistono due impostazioni per questa opzione:

- **Soglia di correzione.** Consente di ottenere informazioni sul livello di confidenza oltre il quale la precisione corrisponderà alla percentuale specificata.
- **Migliora precisione.** Consente di ottenere informazioni sul livello di confidenza oltre il quale la precisione migliora in base al fattore specificato. Per esempio, se la precisione globale è 90% e questa opzione è impostata su 2,0 il valore riportato corrisponderà alla confidenza richiesta per una precisione del 95%.

Trova campi previsti/predittore mediante. Determina in che modo i campi previsti vengono fatti corrispondere al campo obiettivo originale.

- **Metadati del campo di output del modello.** Fa corrispondere i campi previsti all'obiettivo in base alle informazioni sui campi del modello, consentendo la corrispondenza anche se un campo previsto è stato rinominato. Alle informazioni sui campi del modello è possibile accedere anche per qualsiasi campo previsto dalla finestra di dialogo Valori mediante un nodo Tipo. [Per ulteriori informazioni, vedere l'argomento Utilizzo della finestra di dialogo Valori in il capitolo 4 a pag. 132.](#)
- **Formato del nome campo.** Fa corrispondere i campi in base alla convenzione di denominazione. Per esempio, i valori previsti generati da un insieme di modelli C5.0 per un obiettivo denominato *risposta* devono trovarsi in un campo denominato *\$C-risposta*.

Suddividi per partizione. Se si utilizza un campo di partizione per suddividere i record in campioni di addestramento, di test e di validazione, selezionare questa opzione per visualizzare i risultati separatamente per ogni partizione. [Per ulteriori informazioni, vedere l'argomento Nodo Partizione in il capitolo 4 a pag. 193.](#)

Nota: se si suddivide per partizione, i record con valori nulli nel campo di partizione saranno esclusi dall'analisi. Ciò non rappresenta un problema se si utilizza un nodo Partizione, poiché i nodi Partizione non generano valori nulli.

Analisi definita dall'utente. È possibile specificare che nella valutazione del modello o dei modelli venga utilizzato un calcolo di analisi personalizzato. Utilizzare le espressioni CLEM per specificare l'oggetto da calcolare per ogni record e come combinare i punteggi a livello di record in un punteggio globale. Utilizzare le funzioni @TARGET e @PREDICTED per fare riferimento rispettivamente al valore obiettivo (output effettivo) e al valore previsto.

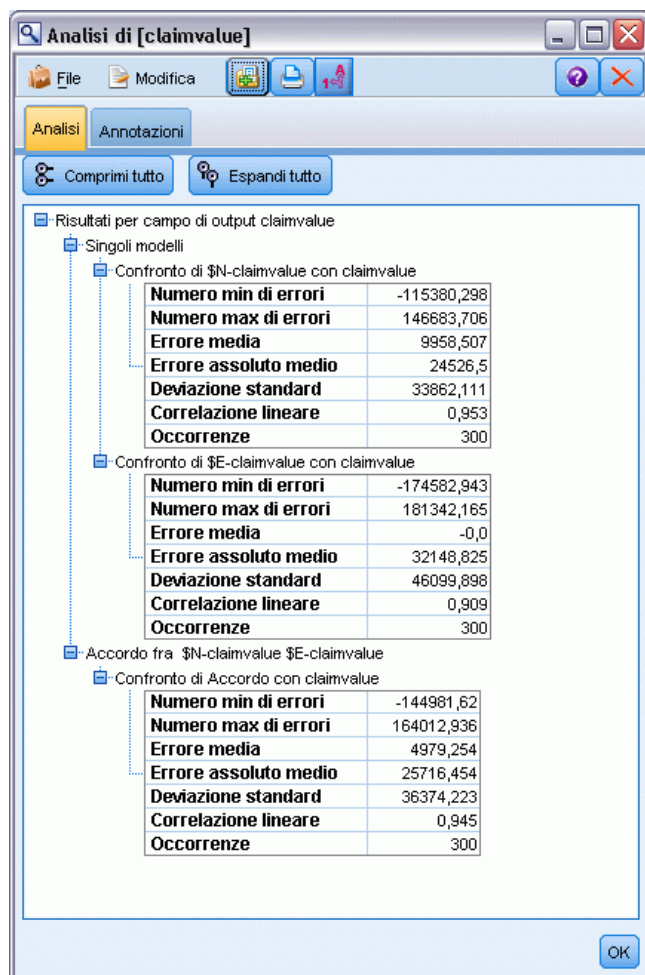
- **Se.** Specificare un'espressione condizionale se si desidera utilizzare calcoli diversi in base a condizioni determinate.
- **Allora.** Specificare il calcolo se la condizione Se è vera.
- **Altrimenti.** Specificare il calcolo se la condizione Se è falsa.
- **Utilizza.** Selezionare una statistica per calcolare il punteggio globale dai punteggi individuali.

Analisi distinta per campi. Mostra i campi categoriali disponibili per l'esecuzione di analisi distinte. Oltre all'analisi globale, verrà riportata un'analisi separata per ogni categoria di ciascun campo specificato.

Browser di output del nodo Analisi

Il browser di output del nodo di analisi permette di visualizzare i risultati dell'esecuzione del nodo Analisi. Dal menu File sono disponibili le normali opzioni di salvataggio, esportazione e stampa. Per ulteriori informazioni, vedere l'argomento [Visualizzazione dell'output a pag. 348](#).

Figura 6-13
Browser di output del nodo Analisi



Quando si visualizza per la prima volta l'output del nodo Statistiche, i risultati vengono espansi. Per nascondere i risultati, utilizzare il controllo di espansione di un elemento se si desidera comprimere solo i risultati di tale elemento oppure fare clic sul pulsante **Comprimi tutto** se si desidera nascondere tutti i risultati. Per visualizzare nuovamente i risultati dopo averli compressi, utilizzare il controllo di espansione a sinistra di un elemento per visualizzarne i risultati, oppure fare clic sul pulsante **Espandi tutto** per visualizzare tutti i risultati.

Risultati per campo di output. L'output del nodo Analisi contiene una sezione per ogni campo di output per il quale esiste un campo di previsione corrispondente creato da un modello generato.

Confronto. Nella sezione del campo di output esiste una sottosezione per ogni campo di previsione associato a tale campo di output. Nel caso di campi di output categoriali, il livello superiore di questa sezione contiene una tabella che mostra il numero e la percentuale di previsioni corrette ed errate e il numero complessivo di record nello stream. Nel caso di campi di output numerici, questa sezione mostra le seguenti informazioni:

- **Numero min di errori.** Mostra il numero minimo di errori (differenza tra i valori osservati e quelli previsti).
- **Numero max di errori.** Mostra il numero massimo di errori.
- **Errore media.** Mostra il valore medio (media) degli errori in tutti i record. La media mostra se nel modello esiste una **distorsione** sistematica (una maggiore tendenza alla stima per eccesso rispetto a quella per difetto o viceversa).
- **Errore assoluto media.** Mostra la media dei valori assoluti degli errori in tutti i record. Indica la grandezza media dell'errore, indipendentemente dalla direzione.
- **Deviazione standard.** Mostra la deviazione standard degli errori.
- **Correlazione lineare.** Mostra la correlazione lineare tra i valori previsti e quelli effettivi. Questa statistica varia tra -1.0 e 1.0 . Valori vicini a $+1.0$ indicano una forte associazione positiva: in questo caso, i valori alti previsti sono associati ai valori alti effettivi e i valori bassi previsti sono associati ai valori bassi effettivi. Valori vicini a -1.0 indicano una forte associazione negativa: in questo caso, i valori alti previsti sono associati ai valori bassi effettivi e viceversa. Valori vicini a $0,0$ indicano un'associazione debole: in questo caso, i valori previsti sono più o meno indipendenti dai valori effettivi. *Nota:* una voce vuota indica che la correlazione lineare non può essere calcolata perché i valori effettivi o previsti sono costanti.
- **Occorrenze.** Mostra il numero di record utilizzati nell'analisi.

Matrice coincidenza. Se nelle opzioni di analisi è stata richiesta una matrice di coincidenza per campi di output categoriali, tale matrice viene inclusa in questa sottosezione dell'output. Le righe rappresentano i valori effettivi osservati e le colonne rappresentano i valori previsti. La cella nella tabella indica il numero di record per ogni combinazione di valori previsti e valori effettivi.

Valutazione performance. Se nelle opzioni di analisi è stata richiesta una statistica di valutazione delle prestazioni per campi di output categoriali, i risultati di tale valutazione vengono visualizzati in questa sezione dell'output. Viene elencata ogni categoria di output con la relativa statistica di valutazione delle prestazioni.

Report Valori di confidenza. Se nelle opzioni di analisi sono stati richiesti valori di confidenza per campi di output categoriali, tali valori vengono visualizzati in questa sezione. Per i valori di confidenza del modello vengono riportate le seguenti statistiche:

- **Intervallo.** Mostra l'intervallo (valori minimi e valori massimi) dei valori di confidenza per i record dei dati nello stream.
- **Media corretta.** Mostra la confidenza media per i record classificati correttamente.
- **Media errata.** Mostra la confidenza media per i record classificati non correttamente.
- **Sempre corretto sopra.** Mostra la soglia di confidenza oltre la quale le previsioni sono sempre corrette e la percentuale dei casi che soddisfano questo criterio.
- **Sempre errato sotto.** Mostra la soglia di confidenza al di sotto della quale le previsioni sono sempre errate e la percentuale dei casi che soddisfano questo criterio.

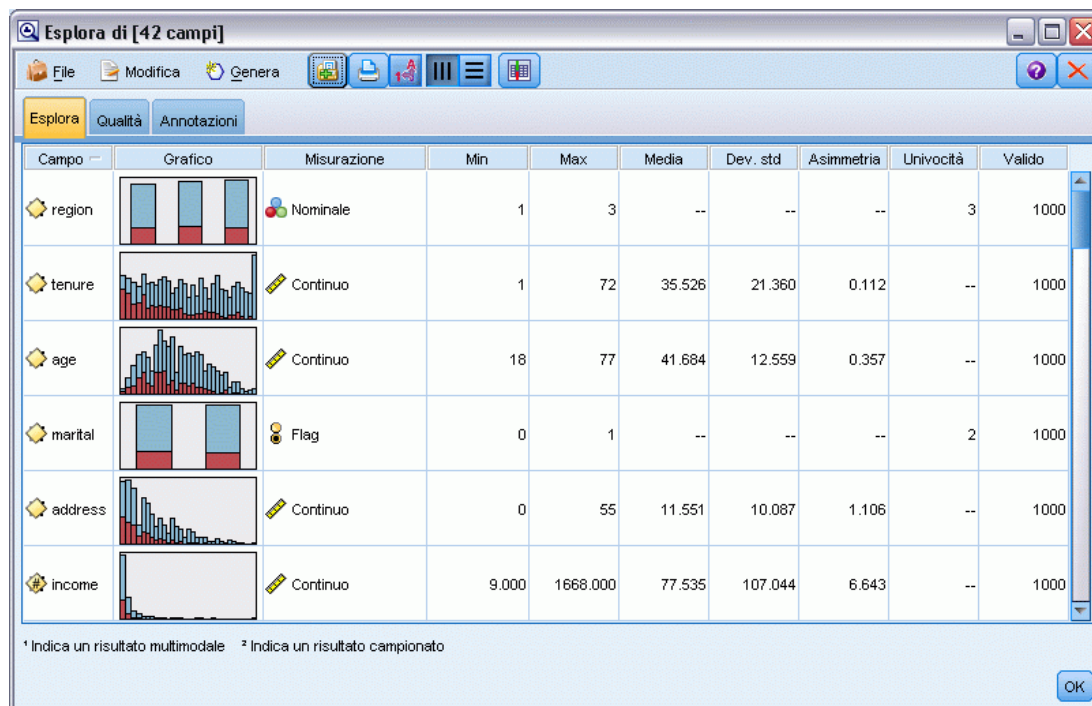
- **X% precisione al di sopra di.** Mostra il livello di confidenza al quale la precisione è X%. X è approssimativamente il valore specificato per Soglia di correzione nelle opzioni Analisi. Per alcuni modelli e insiemi di dati non è possibile scegliere un valore di confidenza che fornisca la soglia esatta specificata nelle opzioni (in genere a causa di cluster di casi simili con lo stesso valore di confidenza vicino alla soglia). La soglia visualizzata nel report corrisponde al valore più vicino al criterio di precisione specificato che può essere ottenuto con un'unica soglia del valore di confidenza.
- **X volte corrette al di sopra di.** Mostra il valore di confidenza con una precisione di X volte superiore rispetto all'intero insieme di dati. X è il valore specificato per Migliora precisione nelle opzioni Analisi.

Accordo tra. Se nello stream vengono inclusi due o più modelli generati che prevedono lo stesso campo di output, verranno visualizzate anche le statistiche sull'**accordo** tra le previsioni generate dai modelli. Le statistiche includono il numero e la percentuale di record per i quali le previsioni concordano (per campi di output categoriali) o le statistiche riassuntive degli errori (per campi di output continui). Per i campi categoriali le statistiche includono un'analisi delle previsioni rispetto ai valori effettivi per il sottoinsieme di record in cui i modelli concordano (generano lo stesso valore previsto).

Nodo Esplora

Il nodo Esplora fornisce una prima analisi completa dei dati portati all'interno di IBM® SPSS® Modeler, presentati in una matrice di semplice lettura che può essere ordinata e utilizzata per generare grafici a schermo intero e svariati nodi di preparazione dati.

Figura 6-14
Browser del nodo Esplora

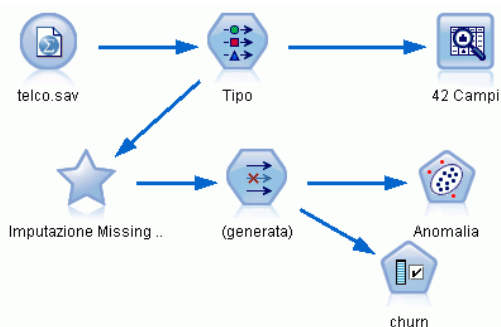


- La scheda Esplora visualizza un report che fornisce statistiche riassuntive, istogrammi e grafici di distribuzione che possono essere utili per una comprensione iniziale dei dati. Nel report, inoltre, i nomi dei campi sono preceduti dall'icona di archiviazione.
- La scheda Qualità del report di esplorazione visualizza informazioni sui valori anomali, estremi e mancanti e offre gli strumenti necessari per gestire questi valori.

Utilizzo del nodo Esplora

Il nodo Esplora può essere collegato direttamente a un nodo di input oppure collocato a valle di un nodo Tipo istanziato. Inoltre, sulla base dei risultati ottenuti è possibile generare diversi nodi di preparazione dati. Per esempio, è possibile generare un nodo Filtro che escluda i campi che presentano troppi valori mancanti per essere utili ai fini della modellazione e generare un Supernodo che assegni i valori mancanti ad alcuni o tutti i campi restanti. Proprio in questi casi viene dimostrata la vera efficacia dell'esplorazione, che consente non solo di valutare lo stato corrente dei dati, ma anche di agire sulla base di tale valutazione. [Per ulteriori informazioni, vedere l'argomento Preparazione dei dati per l'analisi \(Esplorazione dei dati\) in il capitolo 8 in Guida alle applicazioni di IBM SPSS Modeler 14.2.](#)

Figura 6-15
Stream con Supernodo valori mancanti

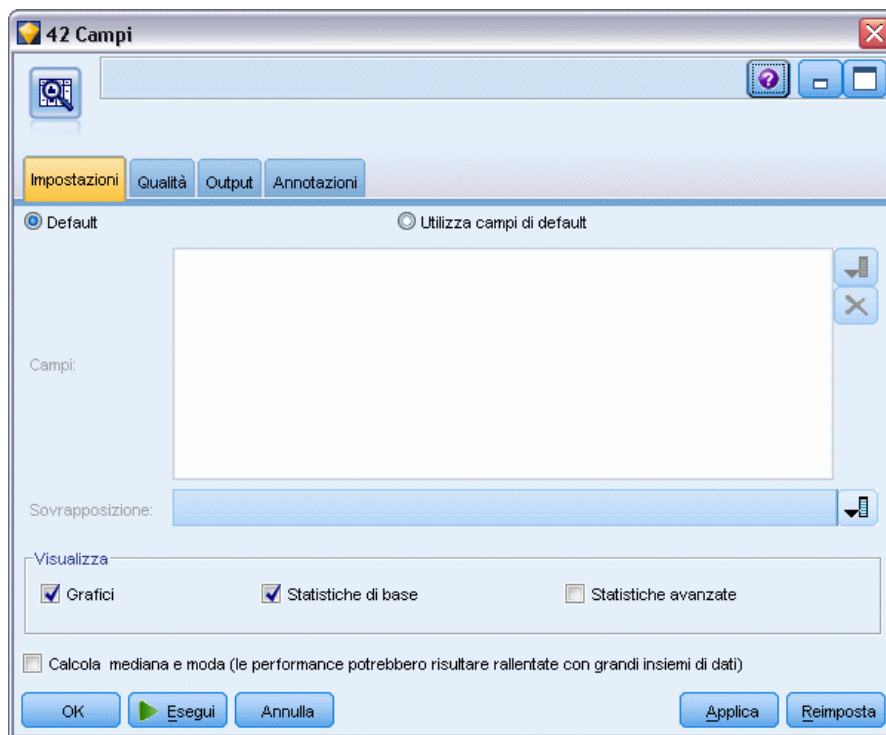


Screening o campionamento dei dati. Poiché un'esplorazione iniziale si rivela particolarmente efficace quando si ha a che fare con "grandi quantità di dati", è possibile utilizzare un nodo Campione che, mediante la selezione di un solo sottoinsieme di record, consente di ridurre i tempi di elaborazione di tale processo. Il nodo Esplora può anche essere utilizzato assieme ad altri nodi, per esempio i nodi Selezione funzioni e Rilevamento anomalie, durante le fasi preliminari dell'analisi.

Scheda Impostazioni del nodo Esplora

La scheda Impostazioni consente di specificare i parametri di base per l'esplorazione.

Figura 6-16
Nodo Esplora: scheda Impostazioni



Default. È possibile collegare semplicemente il nodo allo stream e fare clic su Esegui per generare un report di esplorazione per tutti i campi basato sulle impostazioni di default, nel modo seguente:

- In assenza di impostazioni del nodo Tipo, nel report vengono inclusi tutti i campi.
- In presenza di impostazioni del nodo Tipo, indipendentemente dalla relativa istanziazione, nella visualizzazione vengono inclusi tutti i campi *Input*, *Obiettivo* ed *Entrambi*. Se è disponibile un solo campo *Obiettivo*, questo viene utilizzato come campo di sovrapposizione. Se sono disponibili più campi *Obiettivo*, non è specificata una sovrapposizione di default.

Utilizza campi di default. Selezionare questa opzione per selezionare i campi manualmente. Utilizzare il pulsante di selezione dei campi riportato a destra per selezionare i campi singolarmente o per tipo.

Campo sovrapposto. Il campo sovrapposto viene utilizzato per disegnare i grafici in anteprima visualizzati nel report di esplorazione. Nel caso di un campo continuo (intervallo numerico), vengono calcolate anche le statistiche bivariante (covarianza e correlazione). Se è presente un solo campo *Obiettivo* in base alle impostazioni del nodo Tipo, tale campo viene utilizzato come campo di sovrapposizione di default, come descritto in precedenza. In alternativa, è possibile selezionare Utilizza campi di default per specificare una sovrapposizione.

Visualizzazione. Consente di specificare se i grafici sono disponibili nell'output e di scegliere le statistiche che vengono visualizzate per default.

- **Grafici.** Visualizza un grafico per ogni campo selezionato; a seconda dei dati, tale grafico può essere di distribuzione (a barre), un istogramma o un grafico a dispersione. I grafici vengono visualizzati come anteprime nel report iniziale, ma è anche possibile generare grafici a schermo intero e nodi di grafici. [Per ulteriori informazioni, vedere l'argomento Browser di output del nodo Esplora a pag. 374.](#)
- **Statistiche di base/Statistiche avanzate.** Specifica il livello delle statistiche visualizzate per default nell'output. Questa impostazione determina la visualizzazione iniziale, ma tutte le statistiche sono disponibili nell'output, indipendentemente da questa impostazione. [Per ulteriori informazioni, vedere l'argomento Visualizza statistiche a pag. 376.](#)

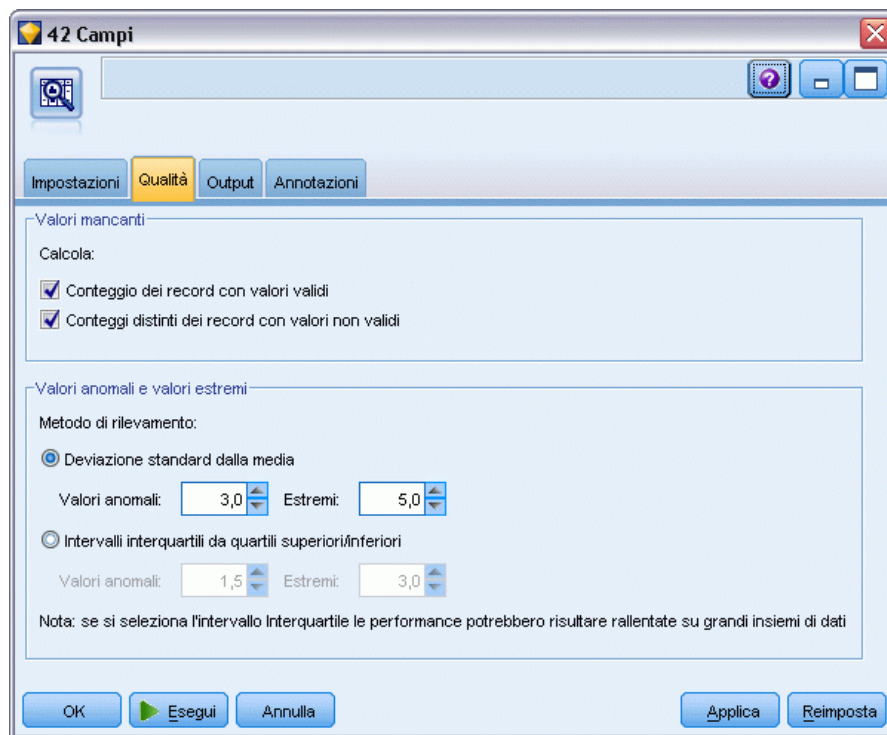
Calcola mediana e moda. Calcola la mediana e la moda di tutti i campi nel report. Nel caso di insiemi di dati di grandi dimensioni, queste statistiche potrebbero rallentare l'elaborazione in quanto richiedono più tempo di altre per il calcolo. Per quanto riguarda la mediana, il valore riportato, in alcuni casi, potrebbe essere basato su un campione di 2000 record anziché su tutto l'insieme di dati. Questo campionamento viene eseguito per ogni campo nei casi in cui si rischierebbe di superare i limiti di memoria. Quando è attivo il campionamento, i risultati verranno etichettati di conseguenza nell'output (*Mediana campione* invece che semplicemente *Mediana*). Tutte le altre statistiche vengono invece sempre calcolate utilizzando l'intero insieme di dati.

Campi vuoti o senza tipo. Quando sono utilizzati con i dati istanziati, i campi senza tipo non sono inclusi nel report di esplorazione. Per includere i campi senza tipo (compresi quelli vuoti), selezionare *Cancella tutti i valori nei nodi Tipo a monte*. Questa operazione fa in modo che i dati non vengano istanziati, determinando l'inclusione di tutti i campi nel report. Questa funzione può essere utile, per esempio, se si desidera ottenere un elenco completo di tutti i campi o generare un nodo *Filtro* che escluda i campi vuoti. [Per ulteriori informazioni, vedere l'argomento Filtraggio dei campi con dati mancanti a pag. 381.](#)

Nodo Esplora, scheda Qualità

La scheda *Qualità* del nodo *Esplora* fornisce le opzioni per la gestione di valori mancanti, anomali ed estremi.

Figura 6-17
Scheda Qualità del nodo Esplora



Valori mancanti

- **Conteggio dei record con valori validi.** Selezionare questa opzione per mostrare il numero di record con valori validi per ogni campo valutato. Si noti che i valori nulli (non definiti), i valori vuoti, gli spazi e le stringhe vuote vengono sempre trattati come valori non validi.
- **Conteggi distinti dei record con valori non validi.** Selezionare questa opzione per mostrare il numero di record con ogni tipo di valore non valido per ciascun campo.

Valori anomali ed estremi

Metodo di rilevamento dei valori anomali ed estremi. Sono supportati due metodi:

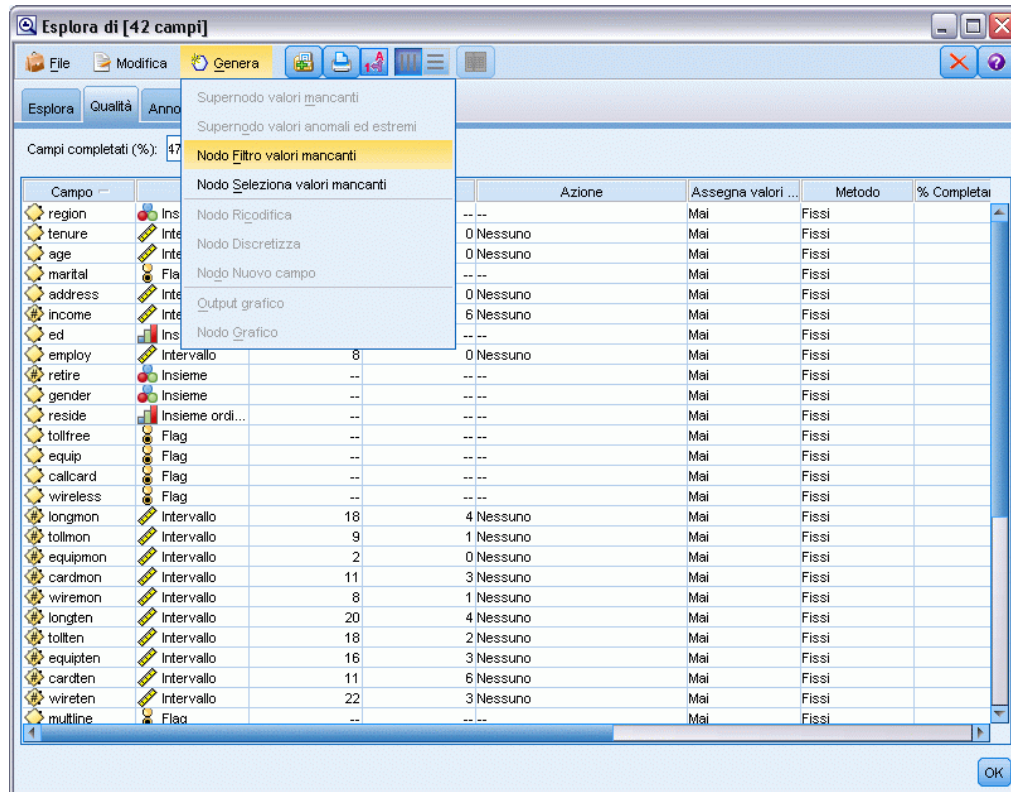
Deviazione standard dalla media. Rileva i valori anomali ed estremi in base al numero di deviazioni standard dalla media. Per esempio, se un campo ha una media di 100 e una deviazione standard di 10, è possibile specificare 3.0 per indicare che tutti i valori inferiori a 70 o superiori a 130 devono essere considerati anomali.

Intervallo interquartile. Rileva i valori anomali ed estremi in base all'intervallo interquartile, vale a dire l'intervallo in cui rientrano i due quartili centrali (fra il 25mo e il 75mo percentile). Per esempio, in base a un'impostazione di default di 1.5, la soglia inferiore per i valori anomali sarebbe $Q1 - 1.5 * IQR$ e la soglia superiore $Q3 + 1.5 * IQR$. Si noti che l'utilizzo di questa opzione su insiemi di dati di grandi dimensioni può rallentare la performance.

Browser di output del nodo Esplora

Il browser del nodo Esplora è uno strumento particolarmente efficace per ottenere una panoramica dei propri dati. La scheda Esplora visualizza grafici in anteprima, icone di archiviazione e statistiche per tutti i campi, mentre la scheda Qualità visualizza informazioni sui valori anomali, estremi e mancanti. In base ai grafici iniziali e alle statistiche riassuntive, è possibile decidere di ricodificare un campo numerico o i valori di un campo nominale oppure di derivare un nuovo campo. In alternativa, può rivelarsi utile approfondire l'esplorazione utilizzando una visualizzazione più avanzata. A questo scopo, dal browser del report di esplorazione, utilizzare il menu Genera per creare un numero qualsiasi di nodi che possono essere utilizzati per trasformare o visualizzare i dati.

Figura 6-18
Generazione di un nodo Filtro valori mancanti



- Ordinare le colonne facendo clic sulla relativa intestazione oppure trascinandole e rilasciandole. È supportata anche la maggior parte delle operazioni di output. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output a pag. 348.](#)
- Visualizzare valori e intervalli per i campi facendo doppio clic sulle colonne Misurazione o Univocità di un campo.

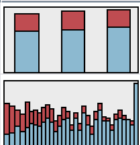
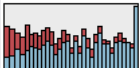
- Utilizzare la barra degli strumenti o il menu Modifica per visualizzare o nascondere le etichette dei valori oppure per scegliere le statistiche da visualizzare. [Per ulteriori informazioni, vedere l'argomento Visualizza statistiche a pag. 376.](#)
- Verificare le icone di archiviazione a sinistra dei nomi dei campi. L'archiviazione definisce le modalità di archiviazione dei dati in un campo. Per esempio, un campo con valori 1 e 0 è un campo in cui sono archiviati dati di tipo intero. Si distingue dal livello di misurazione, che descrive l'utilizzo dei dati e non influisce sull'archiviazione. [Per ulteriori informazioni, vedere l'argomento Impostazione dell'archiviazione e della formattazione dei campi in il capitolo 2 a pag. 27.](#)

Visualizzazione e generazione dei grafici

Se non è selezionata una sovrapposizione, la scheda Esplora visualizza i grafici a barre (per campi nominali o flag) o gli istogrammi (campi continui).

Figura 6-19

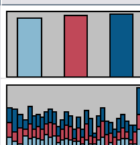
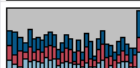
Estratto dei risultati dell'esplorazione senza campo di sovrapposizione

Campo	Grafico	Tipo	Min	Max	Media	Dev. std	Asimmetria	Univocità	Valido
region		Insieme	1	3	--	--	--	3	1000
tenure		Intervallo	1	72	35.526	21.360	0.112	--	1000

In caso di sovrapposizione di un campo nominale o flag, i grafici vengono colorati in base ai valori della sovrapposizione.

Figura 6-20

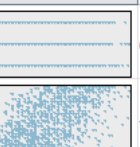

Estratto dei risultati dell'esplorazione con sovrapposizione di un campo nominale

Campo	Grafico	Tipo	Min	Max	Media	Dev. std	Asimmetria	Univocità	Valido
region		Insieme	1	3	--	--	--	3	1000
tenure		Intervallo	1	72	35.526	21.360	0.112	--	1000

In caso di sovrapposizione di un campo continuo, anziché barre e istogrammi monodimensionali vengono generati grafici a dispersione bidimensionali. In questo caso, l'asse x viene mappato al campo di sovrapposizione, consentendo di visualizzare la stessa scala su tutti gli assi x mentre si procede con la lettura della tabella.

Figura 6-21

Estratto dei risultati dell'esplorazione con sovrapposizione di un campo continuo

Campo	Grafico	Tipo	Min	Max	Media	Correlazione	Correlazione T	Correlazione T df.
region		Insieme	1	3	--	--	--	--
tenure		Intervallo	1	72	35.526	0.490	17.768	998.000

- Per i campi flag o nominali, tenere il cursore del mouse su una barra in modo da visualizzare l'etichetta o il valore sottostante in una descrizione.

- Per i campi flag o nominali, utilizzare la barra degli strumenti per cambiare da orizzontale a verticale l'orientamento dei grafici in anteprima.
- Per generare un grafico a schermo intero da un'anteprima, fare doppio clic su di essa oppure selezionare un'anteprima e scegliere Output grafico dal menu Genera. *Nota:* Se era stato creato un grafico in anteprima in base a dati campionati e lo stream di dati originale è ancora aperto, il grafico generato conterrà tutti i casi.

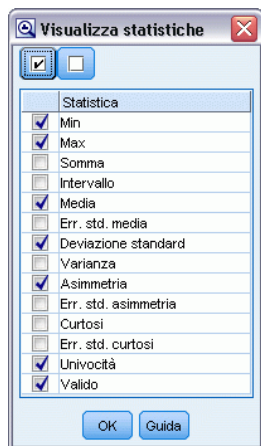
Il grafico si può generare solo se il nodo Esplora che ha generato l'output è collegato allo stream.

- Per generare un nodo di grafico corrispondente, selezionare uno o più campi nella scheda Esplora e scegliere Nodo Grafico nel menu Genera. Il nodo risultante viene aggiunto all'area di disegno dello stream e può essere utilizzato per ricreare il grafico ogni volta che viene eseguito lo stream.
- se un insieme di sovrapposizione contiene più di 100 valori, verrà visualizzato un avviso e non verrà inclusa la sovrapposizione.

Visualizza statistiche

La finestra di dialogo Visualizza statistiche consente di scegliere quali statistiche vengono visualizzate nella scheda Esplora. Le impostazioni iniziali vengono specificate nel nodo Esplora. [Per ulteriori informazioni, vedere l'argomento Scheda Impostazioni del nodo Esplora a pag. 370.](#)

Figura 6-22
Visualizza statistiche



Minimo. Il valore più basso assunto da una variabile numerica.

Massimo. Il valore più alto di una variabile numerica.

Somma. La somma o il totale di tutti i valori non mancanti di tutti i casi.

Intervallo. La differenza tra il valore massimo ed il valore minimo di una variabile numerica.

Media. Una misura di tendenza centrale. La somma dei valori di tutte le osservazioni divisa per il numero di osservazioni. Viene anche detta media aritmetica.

Errore standard della media. Una misura di quanto può variare il valore della media da campione a campione per campioni estratti dalla stessa distribuzione. Può essere utilizzata per confrontare genericamente la media osservata rispetto a un valore ipotizzato (ovvero, è possibile concludere che i due valori sono diversi se il rapporto della differenza rispetto all'errore standard è inferiore a -2 o maggiore di +2).

Deviazione standard. Una misura di dispersione intorno alla media, uguale alla radice quadrata della varianza. L'unità di misura della deviazione standard è la stessa della variabile originale.

Varianza. Una misura della dispersione dei valori intorno alla media. È calcolata come somma dei quadrati degli scostamenti dalla media, divisa per il numero totale delle osservazioni valide meno 1. La varianza è espressa in quadrati dell'unità di misura della variabile.

Asimmetria. Una misura dell'asimmetria di una distribuzione. La distribuzione normale è simmetrica e ha un valore di asimmetria pari a 0. Una distribuzione con una notevole asimmetria positiva ha una lunga coda a destra. Una distribuzione con asimmetria negativa ha una coda a sinistra. In generale un'asimmetria con valore più che doppio dell'errore standard indica lo scostamento dalla normale simmetria.

Errore standard dell'asimmetria. Il rapporto fra l'asimmetria di una distribuzione e il suo errore standard viene usato come test di normalità. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per l'asimmetria indicata una coda a destra lunga; un valore negativo estremo indica una coda a sinistra lunga.

Curtosi. Una misura di quanto le osservazioni si trovino raggruppate nelle code. Per la distribuzione normale, il valore della statistica di curtosi è zero. Una curtosi positiva indica che, rispetto a una distribuzione normale, le osservazioni sono più raggruppate intorno al centro della distribuzione e hanno code più sottili fino ai valori estremi della distribuzione; a quel punto, le code della distribuzione leptocurtica sono più spesse rispetto a una distribuzione normale. Una curtosi negativa indica che, rispetto a una distribuzione normale, le osservazioni sono meno raggruppate e hanno code più spesse fino ai valori estremi della distribuzione; a quel punto, le code della distribuzione platicurtica sono più sottili rispetto a una distribuzione normale.

Errore standard della curtosi. L'ipotesi di normalità può essere rifiutata se questo rapporto è maggiore di 2 in valore assoluto. Un valore positivo elevato per la curtosi indica che le code della distribuzione sono più lunghe di quelle di una distribuzione normale; un valore negativo per la curtosi indica code più corte, simili a quelle di una distribuzione uniforme a forma di scatola.

Simultaneo. Valuta tutti gli effetti simultaneamente, correggendo ogni effetto per tutti gli altri effetti di qualunque tipo.

Valido. I casi validi il cui valore non è né il valore mancante di sistema né un valore definito come mancante dall'utente.

Mediana. È il valore sopra il quale e sotto il quale ricade la metà dei casi, il 50-esimo percentile. Se il numero di casi è pari, la mediana è pari alla media dei due casi centrali quando questi sono ordinati secondo l'ordine ascendente o discendente. La mediana è una misura di tendenza centrale non sensibile ai valori anomali, a differenza della media che può essere influenzata da valori eccezionalmente bassi o alti.

Moda. Il valore o i valori che ricorrono più frequentemente. Se più valori condividono la maggiore ricorrenza, ognuno di essi è una moda.

Si noti che mediana e moda sono escluse per default al fine di migliorare la performance; tuttavia, è possibile selezionarle nella scheda Impostazioni del nodo Esplora. [Per ulteriori informazioni, vedere l'argomento Scheda Impostazioni del nodo Esplora a pag. 370.](#)

Statistiche per le sovrapposizioni

Se viene utilizzato un campo di sovrapposizione continuo (intervallo numerico), sono disponibili anche le seguenti statistiche:

Covarianza. Una misura non standardizzata di associazione tra due variabili, pari alla deviazione del prodotto degli scarti divisa per $N-1$.

Browser Esplora, scheda Qualità

Figura 6-23
Report Qualità nel browser del nodo Esplora

Campo	Misurazione	Valori anomali	Estremi	Azione	Assegna valori ...	Metodo	% Completati
region	Nominale	--	--		Mai	Fissi	
tenure	Continuo	0	0	Nessuno	Mai	Fissi	
age	Continuo	0	0	Nessuno	Mai	Fissi	
marital	Flag	--	--		Mai	Fissi	
address	Continuo	12	0	Nessuno	Mai	Fissi	
income	Continuo	9	6	Nessuno	Mai	Fissi	
ed	Ordinale	--	--		Mai	Fissi	
employ	Continuo	8	0	Nessuno	Mai	Fissi	
retire	Nominale	--	--		Mai	Fissi	
gender	Nominale	--	--		Mai	Fissi	
reside	Ordinale	--	--		Mai	Fissi	
tollfree	Flag	--	--		Mai	Fissi	
equip	Flag	--	--		Mai	Fissi	
calcard	Flag	--	--		Mai	Fissi	
wireless	Flag	--	--		Mai	Fissi	
longmon	Continuo	18	4	Nessuno	Mai	Fissi	
tollmon	Continuo	9	1	Nessuno	Mai	Fissi	
equipmon	Continuo	2	0	Nessuno	Mai	Fissi	
cardmon	Continuo	11	3	Nessuno	Mai	Fissi	

La scheda Qualità nel browser del nodo Esplora visualizza i risultati dell'analisi della qualità dei dati e consente di specificare il modo in cui devono essere gestiti i valori anomali, estremi e mancanti.

Assegnazione dei valori mancanti

Il report di esplorazione elenca la percentuale dei record completi per ciascun campo, nonché il numero dei valori validi, nulli e vuoti. È possibile assegnare i valori mancanti a campi specifici secondo le proprie esigenze e quindi generare un Supernodo per applicare queste trasformazioni.

- Nella colonna Assegna valori mancanti, specificare il tipo di valori che si desidera assegnare. Si può decidere di assegnare i valori vuoti, nulli o entrambi, oppure è possibile specificare una condizione o un'espressione personalizzata per selezionare i valori da assegnare.

Esistono vari tipi di valori mancanti riconosciuti da IBM® SPSS® Modeler:

- **Valori nulli o mancanti di sistema.** Sono valori diversi da stringhe che vengono lasciati vuoti nel database o nel file di origine e che non sono specificati in modo esplicito come “mancanti” in un nodo Tipo o in un nodo di input. I valori mancanti di sistema vengono visualizzati come \$null\$. In SPSS Modeler le stringhe vuote non vengono considerate valori nulli, ma possono essere considerate valori nulli da alcuni database.
- **Stringhe vuote e spazi bianchi.** I valori di stringhe vuote e spazi bianchi (stringhe prive di caratteri visibili) sono trattati in modo diverso rispetto ai valori nulli. Nella maggior parte dei casi, le stringhe vuote vengono considerate equivalenti agli spazi bianchi. Per esempio, se si seleziona l’opzione che consente di trattare gli spazi bianchi come valori vuoti in un nodo Tipo o di input, questa impostazione risulta valida anche per le stringhe vuote.
- **Valori vuoti o mancanti definiti dall’utente.** Sono valori quali sconosciuto, 99 o –1 che vengono definiti in modo esplicito come mancanti in un nodo Tipo o in un nodo di input. Se lo si desidera, si può anche decidere di considerare i valori nulli e gli spazi bianchi come valori vuoti: in questo modo è possibile evidenziarli in modo da poter essere elaborati diversamente ed esclusi dalla maggior parte dei calcoli. Per esempio, è possibile utilizzare la funzione @BLANK per trattare questi valori, insieme ad altri tipi di valori mancanti, come valori vuoti. [Per ulteriori informazioni, vedere l’argomento Utilizzo della finestra di dialogo Valori in il capitolo 4 a pag. 132.](#)

- Nella colonna Metodo, indicare il metodo da utilizzare.

Per l’immissione di valori mancanti sono disponibili i seguenti metodi:

Fissi. Sostituisce con un valore fisso (la media di un campo, il valore intermedio dell’intervallo o una costante specificata dall’utente).

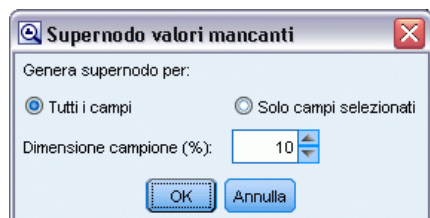
Casuale. Sostituisce con un valore casuale in base a una distribuzione normale o uniforme.

Espressione. Consente di specificare un’espressione personalizzata. Per esempio, è possibile sostituire valori con una variabile globale creata dal nodo Globali.

Algoritmo. Sostituisce con un valore previsto da un modello basato sull’algoritmo C&RT. Per ogni campo immesso utilizzando questo metodo, ci sarà un modello C&RT separato e un nodo Riempimento che sostituisce i valori vuoti e nulli con il valore previsto dal modello. Viene quindi utilizzato un nodo Filtro per rimuovere i campi di previsione generati dal modello.

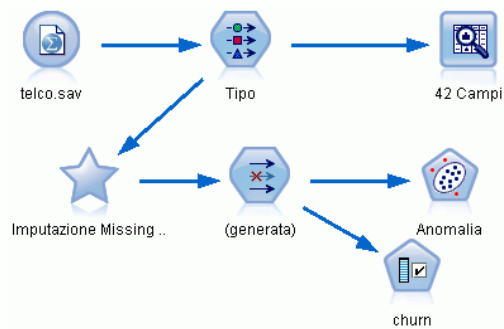
- Per generare un Supernodo valori mancanti, dai menu scegliere:
Genera > Supernodo valori mancanti

Figura 6-24
Finestra di dialogo Supernodo valori mancanti



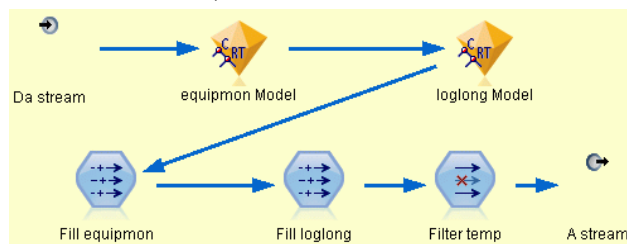
- Selezionare Tutti i campi o Solo campi selezionati e, se lo si desidera, specificare la dimensione del campione (il campione specificato è un valore percentuale; per default viene campionato il 10% di tutti i record).
- Fare clic su OK per aggiungere il Supernodo generato all'area di disegno dello stream.
- Collegare il Supernodo allo stream in modo da applicare le trasformazioni.

Figura 6-25
Aggiunta del Supernodo allo stream



Nel Supernodo viene utilizzata, a seconda dei casi, una combinazione di insieme di modelli, nodi Riempimento e Filtro. Per comprenderne il funzionamento, è possibile modificare il Supernodo e fare clic su Zoom avanti; inoltre, è possibile aggiungere, modificare o rimuovere nodi specifici all'interno del Supernodo per modificarne il comportamento secondo le esigenze.

Figura 6-26
Zoom avanti nel Supernodo



Gestione dei valori anomali ed estremi

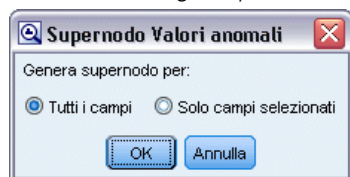
Il report di esplorazione elenca il numero dei valori anomali ed estremi per ciascun campo in base alle opzioni di rilevamento specificate nel nodo Esplora. [Per ulteriori informazioni, vedere l'argomento Nodo Esplora, scheda Qualità a pag. 372.](#) A seconda dei casi, si può decidere di forzare, scartare o annullare questi valori mancanti per campi specifici e quindi generare un Supernodo per applicare queste trasformazioni.

- Nella colonna Azione, indicare in che modo si desidera gestire i valori anomali ed estremi per campi specifici.

Per la gestione di valori anomali ed estremi sono disponibili le seguenti azioni:

- **Forza.** Consente di sostituire i valori anomali ed estremi con il valore più vicino che non sia considerato estremo. Per esempio, se un valore viene definito anomalo quando è superiore o inferiore a 3 deviazioni standard, tutti i valori anomali vengono sostituiti con il valore più alto o più basso di questo intervallo.
 - **Scarta.** Consente di scartare i record con valori anomali o estremi per il campo specificato.
 - **Annulla.** Sostituisce i valori anomali ed estremi con il valore nullo o mancante di sistema.
 - **Forza valori anomali / scarta estremi.** Scarta soltanto i valori estremi.
 - **Forza valori anomali / annulla estremi.** Annulla soltanto i valori estremi.
- Per generare il Supernodo, dai menu scegliere:
Genera > Supernodo valori anomali ed estremi

Figura 6-27
Finestra di dialogo Supernodo valori anomali



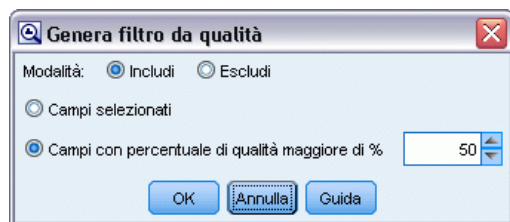
- Selezionare Tutti i campi o Solo campi selezionati, quindi fare clic su OK per aggiungere il Supernodo generato all'area di disegno dello stream.
- Collegare il Supernodo allo stream in modo da applicare le trasformazioni.

Se lo si desidera, è possibile modificare il Supernodo e ingrandirlo per esaminarlo o modificarlo. All'interno del Supernodo, i valori vengono scartati, forzati o annullati mediante una serie di nodi Seleziona e/o Riempimento, secondo necessità.

Filtraggio dei campi con dati mancanti

Dal browser del nodo Esplora è possibile creare un nuovo nodo Filtro in base ai risultati dell'analisi Qualità.

Figura 6-28
Finestra di dialogo Genera filtro da qualità



Modalità. Selezionare l'operazione desiderata per i campi specificati, ovvero Includi o Escludi.

- **Campi selezionati.** Il nodo Filtro includerà/escluderà i campi selezionati nella scheda Qualità. Per esempio, è possibile ordinare la tabella in base alla colonna % Complete, utilizzare il metodo Maiusc-clc per selezionare i campi con la percentuale di completamento più bassa e generare un nodo Filtro che escluda tali campi.
- **Campi con percentuale di qualità maggiore di.** Il nodo Filtro includerà/escluderà i campi in cui la percentuale di record completi è superiore alla soglia specificata. La soglia di default è 50%.

Filtraggio dei campi vuoti o senza tipo

Dopo che sono stati istanziati i valori dei dati, i campi senza tipo o vuoti vengono esclusi dai risultati dell'esplorazione e dalla maggior parte degli output in IBM® SPSS® Modeler. Questi campi vengono ignorati ai fini della modellazione, ma potrebbero "sovraffollare" e confondere inutilmente i dati. In questi casi, è possibile utilizzare il browser del nodo Esplora per generare un nodo Filtro che rimuova questi campi dallo stream.

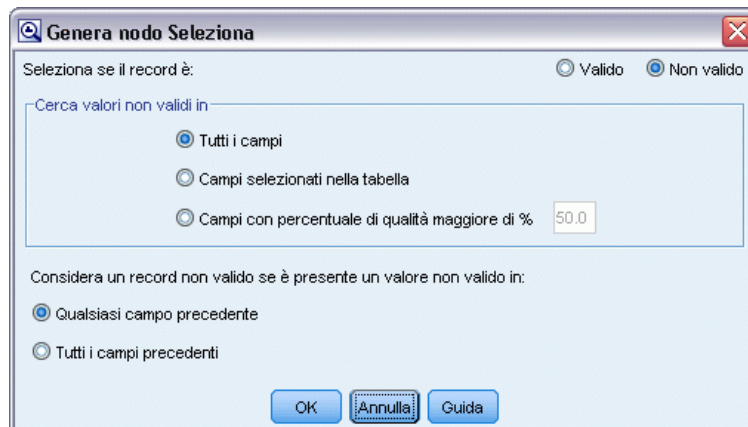
- ▶ Per assicurarsi che tutti i campi vengano inclusi nell'esplorazione, compresi i campi vuoti o senza tipo, selezionare Cancella tutti i valori nel nodo di input o nel nodo Tipo a monte o impostare i Valori su <Passa> per tutti i campi.
- ▶ Nel browser del nodo Esplora, ordinare in base alla colonna % Complete, selezionare i campi che hanno zero valori validi (o qualche altra soglia) e utilizzare il menu Genera per creare un nodo Filtro che possa essere aggiunto allo stream.

Selezione di record con dati mancanti

Dal browser del nodo Esplora, è possibile creare un nuovo nodo Seleziona in base ai risultati dell'analisi Qualità.

- ▶ Dal browser del nodo Esplora, selezionare la scheda Qualità.
- ▶ Dal menu, scegliere:
Genera > Valori mancanti nodo Seleziona

Figura 6-29
Finestra di dialogo Genera nodo Seleziona



Seleziona se il record è: Specificare se conservare i record quando sono di tipo Valido o Non valido.

Cerca valori non validi in. Specificare se cercare valori non validi.

- **Tutti i campi.** Il nodo Seleziona controlla tutti i campi per verificare se contengono valori non validi.
- **Campi selezionati nella tabella.** Il nodo Seleziona esegue il controllo solo nei campi selezionati nella tabella di output Qualità.
- **Campi con percentuale di qualità maggiore di.** Il nodo Seleziona esegue il controllo nei campi in cui la percentuale di record completi è maggiore della soglia specificata. La soglia di default è 50%.

Considera un record non valido se è presente un valore non valido in. Specificare la condizione per identificare un record come non valido.

- **Qualsiasi campo precedente.** Il nodo Seleziona considererà un record non valido se uno *qualsiasi* dei campi specificati sopra contiene un valore non valido per tale record.
- **Tutti i campi precedenti.** Il nodo Seleziona considererà un record non valido solo se *tutti* i campi specificati sopra contengono valori non validi per tale record.

Generazione di altri nodi per la preparazione dei dati

Numerosi nodi utilizzabili nella fase di preparazione dei dati (quali i nodi Ricodifica, Discretizza e Nuovo campo) possono essere generati direttamente dal browser del nodo Esplora. Per esempio:

- È possibile derivare un nuovo campo in base ai valori di *valore_domanda* e *reddito*, selezionando entrambi nel report di esplorazione e scegliendo Nuovo campo dal menu Genera. Il nuovo nodo verrà aggiunto all'area di disegno dello stream.
- Analogamente, in base ai risultati dell'esplorazione si potrebbe determinare che la ricodifica di *reddito* in bin basati su percentili fornisce un'analisi più mirata. Per generare un nodo Discretizza, selezionare la riga del campo nella visualizzazione e scegliere Discretizza dal menu Genera.

Dopo la generazione del nodo e la relativa aggiunta all'area di disegno dello stream, è necessario collegare il nodo allo stream e aprirlo per specificare le opzioni per i campi selezionati.

Nodo Trasformazioni

La normalizzazione dei campi di input rappresenta un'operazione importante prima dell'utilizzo delle tecniche di calcolo dei punteggi tradizionali, quali regressione, regressione logistica e analisi discriminante. Queste tecniche implicano delle congetture relative alle normali distribuzioni dei dati che potrebbero non essere vere per molti file di dati grezzi. Un modo per gestire i dati reali consiste nell'applicare trasformazioni che avvicinano un elemento di dati grezzo a una distribuzione più normale. Inoltre, i campi normalizzati possono essere facilmente confrontati fra loro: per esempio, reddito ed età si trovano su scale completamente diverse in un file di dati grezzi ma, mediante la normalizzazione, l'impatto relativo di ciascuno può essere facilmente interpretato.

Il nodo Trasformazioni è dotato di un Visualizzatore output che consente di valutare visivamente e rapidamente la migliore trasformazione da utilizzare. A prima vista è possibile capire se le variabili sono distribuite normalmente e, quando necessario, scegliere la trasformazione desiderata e applicarla. È possibile scegliere più campi ed eseguire una trasformazione per campo.

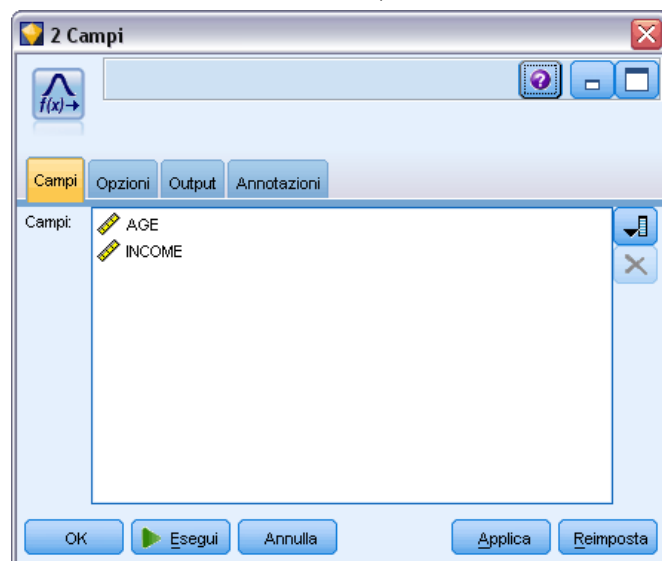
Dopo avere selezionato le trasformazioni desiderate per i campi, è possibile generare nodi Nuovo campo o Riempimento che eseguono le trasformazioni e collegare questi nodi allo stream. Il nodo Nuovo campo crea campi nuovi, mentre il nodo Riempimento trasforma i campi esistenti. [Per ulteriori informazioni, vedere l'argomento Generazione di grafici a pag. 388.](#)

Scheda Campi del nodo Trasformazioni

Nella scheda Campi è possibile specificare quali campi dei dati utilizzare per visualizzare le possibili trasformazioni e applicarle. Solo i campi numerici possono essere trasformati. Fare clic sul pulsante di selezione dei campi e selezionare uno o più campi numerici dall'elenco visualizzato.

Figura 6-30

Nodo Trasformazioni: Scheda Campi



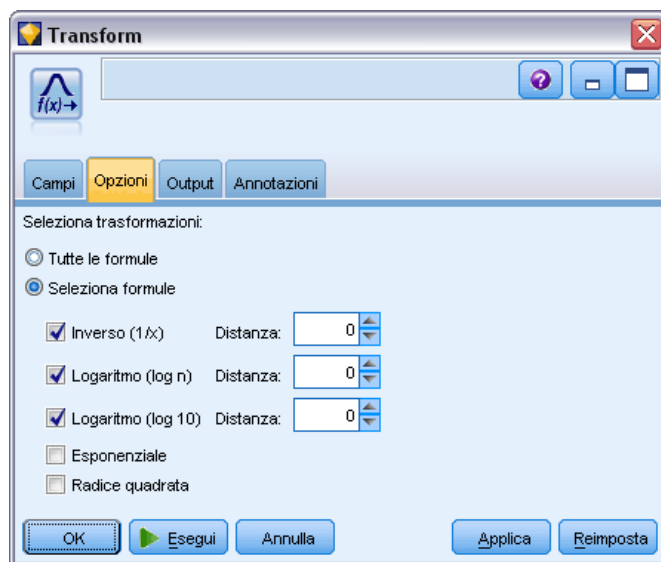
Scheda Opzioni del nodo Trasformazioni

La scheda Opzioni consente di specificare il tipo di trasformazioni che si desidera includere. Si può decidere di includere tutte le trasformazioni disponibili, oppure selezionare le singole trasformazioni.

In questo secondo caso, è anche possibile immettere un numero per distanziare i dati per le trasformazioni inversa e logaritmica. Questa operazione è utile nei casi in cui un'alta proporzione di zeri nei dati potrebbe condizionare i risultati di media e deviazione standard.

Per esempio, si supponga di avere un campo denominato *SALDO* che contiene alcuni valori zero e di volervi applicare la trasformazione inversa. Per evitare alterazioni e condizionamenti indesiderati, in questo caso sarebbe opportuno selezionare Inverso ($1/x$) e immettere 1 nel campo Utilizza una distanza tra i dati. (Questa distanza non è correlata a quella applicata dalla funzione di sequenza @OFFSET in IBM® SPSS® Modeler.)

Figura 6-31
Nodo Trasformazioni: Scheda Opzioni



Tutte le formule. Indica che tutte le trasformazioni disponibili devono essere calcolate e visualizzate nell'output.

Seleziona formule. Consente di selezionare le trasformazioni da calcolare e visualizzare nell'output.

- **Inverso (1/x).** Indica che la trasformazione inversa deve essere visualizzata nell'output.
- **Logaritmo (log n).** Indica che la trasformazione logaritmica $_n$ deve essere visualizzata nell'output.
- **Logaritmo (log 10).** Indica che la trasformazione logaritmica $_{10}$ deve essere visualizzata nell'output.
- **Esponenziale.** Indica che la trasformazione esponenziale (e^x) deve essere visualizzata nell'output.
- **Radice quadrata.** Indica che la trasformazione radice quadrata deve essere visualizzata nell'output.

Scheda Output del nodo Trasformazioni

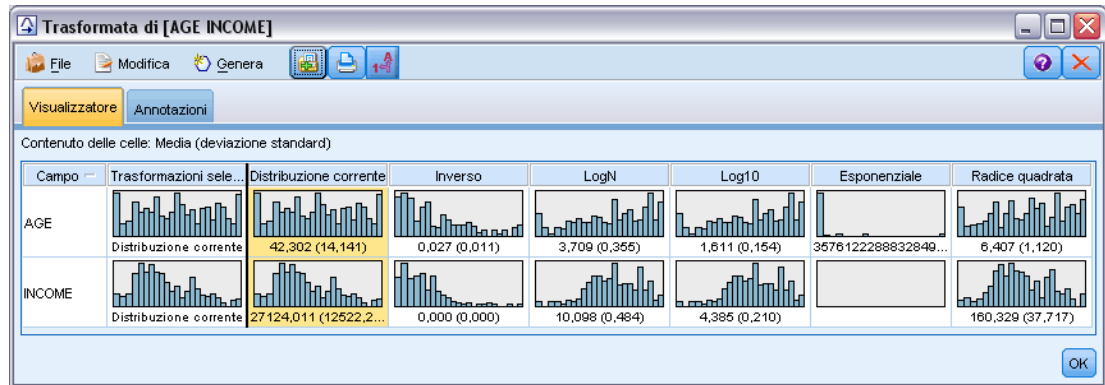
La scheda Output consente di specificare il formato e la posizione dell'output. Si può scegliere di visualizzare i risultati sullo schermo o di inviarli a uno dei tipi di file standard. [Per ulteriori informazioni, vedere l'argomento Scheda Output del nodo Output a pag. 355.](#)

Visualizzatore output del nodo Trasformazioni

Il Visualizzatore output permette di visualizzare i risultati dell'esecuzione del nodo Trasformazioni. Si tratta di uno strumento molto efficace che visualizza più trasformazioni per ogni campo sotto forma di anteprime, consentendo un rapido confronto visivo dei campi. Per

salvare, esportare o stampare l'output, utilizzare le relative opzioni nel menu File. Per ulteriori informazioni, vedere l'argomento [Visualizzazione dell'output](#) a pag. 348.

Figura 6-32
Visualizzazione delle trasformazioni disponibili per campo



Sotto ogni trasformazione, ad eccezione di Trasformazione selezionata, viene visualizzata una legenda nel formato:

Mean (Standard deviation)

Generazione di nodi per le trasformazioni

Il Visualizzatore output costituisce un'utile base di partenza per la fase di preparazione dei dati. Per esempio, si può decidere di normalizzare il campo *ETÀ* in modo da potere utilizzare una tecnica di calcolo del punteggio (come la regressione logistica o l'analisi discriminante) che presuppone una distribuzione normale. In base ai grafici iniziali e alle statistiche riassuntive, è possibile decidere di trasformare il campo *ETÀ* secondo una distribuzione specifica (per esempio, logaritmica). Dopo avere selezionato la distribuzione desiderata, è possibile generare un nodo Nuovo campo con una trasformazione standardizzata da utilizzare per il calcolo del punteggio.

Dal Visualizzatore output è possibile creare i seguenti nodi Operazioni su campi:

- Nuovo campo
- Riempimento

Un nodo Nuovo campo crea campi nuovi con le trasformazioni desiderate, mentre il nodo Riempimento trasforma i campi esistenti. I nodi vengono posti nell'area di disegno sotto forma di Supernodo.

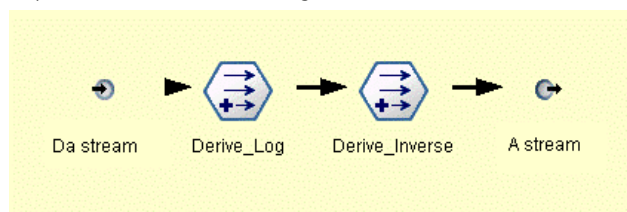
Se si seleziona la stessa trasformazione per diversi campi, un nodo Nuovo campo o Riempimento contiene le formule per quel tipo di trasformazione per tutti i campi a cui si applica quella trasformazione. Per esempio, si supponga di avere selezionato i seguenti campi e trasformazioni per generare un nodo Nuovo campo:

Campo	Trasformazione
<i>ETÀ</i>	Distribuzione corrente
<i>REDDITO</i>	Registro

<i>SALDO_APERT</i>	Inverso
<i>SALDO</i>	Inverso

Il Supernodo contiene i nodi seguenti:

Figura 6-33
Supernodo nell'area di disegno



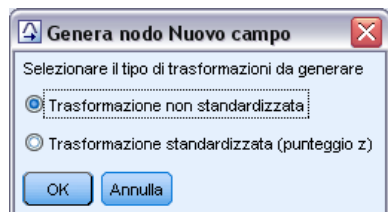
In questo esempio, il nodo Nuovo campo_Logaritmica presenta la formula logaritmica per il campo *REDDITO*, mentre il nodo Nuovo campo_Inversa presenta le formule inverse per i campi *SALDO_APERT* e *SALDO*.

Per generare un nodo

- ▶ Selezionare la trasformazione desiderata per ciascun campo nel Visualizzatore output.
- ▶ Dal menu Genera, scegliere Nodo Nuovo campo o Nodo Riempimento, secondo necessità.

Viene visualizzata la finestra di dialogo Genera nodo Nuovo campo o Genera nodo Riempimento, a seconda di quanto selezionato.

Figura 6-34
Scelta della trasformazione standardizzata o non standardizzata



Scegliere Trasformazione non standardizzata o Trasformazione standardizzata (punteggio z), secondo necessità. La seconda opzione applica un punteggio z alla trasformazione; i punteggi z rappresentano i valori come funzione della distanza della variabile dalla media nelle deviazioni standard. Per esempio, se si applica la trasformazione logaritmica al campo *ETÀ* e si sceglie la trasformazione standardizzata, l'equazione finale del nodo generato sarà:

$$(\log(\text{AGE}) - \text{Mean}) / \text{SD}$$

Dopo che il nodo è stato generato ed è visualizzato nell'area di disegno dello stream:

- ▶ Collegarlo allo stream.
- ▶ Nel caso di un Supernodo, se lo si desidera, è possibile fare doppio clic per visualizzarne il contenuto.

- ▶ Se necessario, fare doppio clic sul nodo Nuovo campo o Riempimento per modificare le opzioni relative ai campi selezionati.

Generazione di grafici

È possibile generare output di istogrammi a schermo intero da un istogramma in anteprima nel Visualizzatore output.

Per generare un grafico

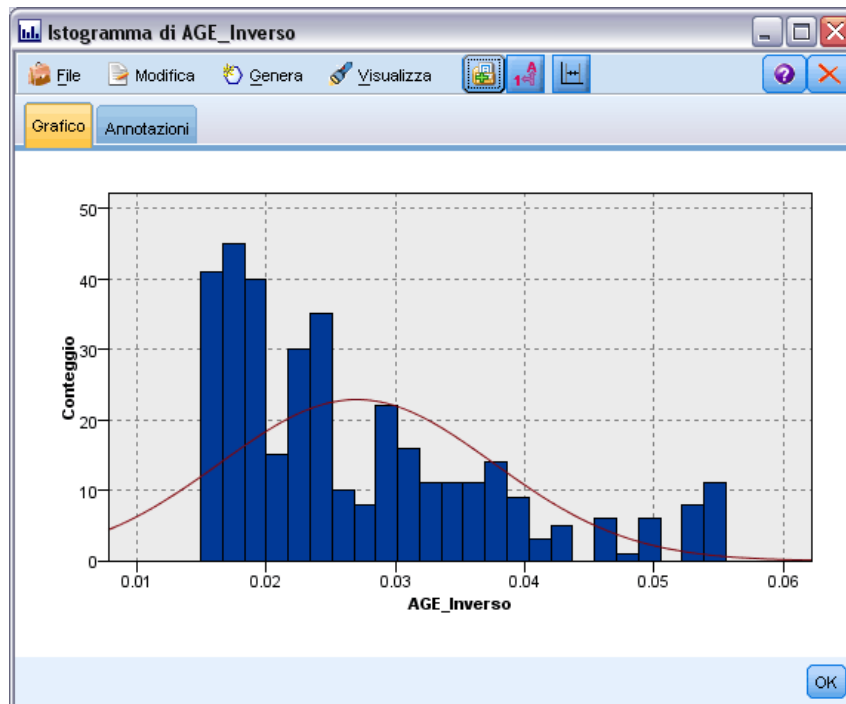
- ▶ Fare doppio clic su un grafico in anteprima nel Visualizzatore output.
- o
- ▶ Selezionare un grafico in anteprima nel Visualizzatore output.
- ▶ Dal menu Genera, scegliere Output grafico.

Viene visualizzato l'istogramma con una normale curva di distribuzione sovrapposta. Questo consente di confrontare l'effettiva corrispondenza tra ogni trasformazione disponibile e una normale distribuzione.

Nota: il grafico si può generare solo se il nodo Trasformazioni che ha generato l'output è collegato allo stream.

Figura 6-35

Istogramma di trasformazione con una curva di distribuzione normale sovrapposta



Altre operazioni

Dal Visualizzatore output è anche possibile:

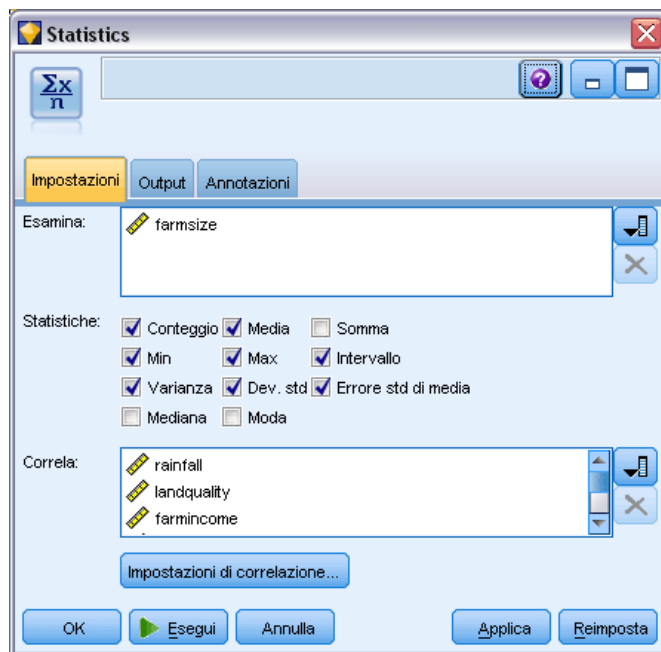
- Ordinare la griglia di output in base alla colonna Campo.
- Esportare l'output in un file HTML. [Per ulteriori informazioni, vedere l'argomento Esportazione dell'output a pag. 351.](#)

Nodo Statistiche

Il nodo Statistiche fornisce informazioni riassuntive di base su campi numerici. È possibile ottenere statistiche riassuntive per singoli campi e per correlazioni tra campi.

Scheda Impostazioni del nodo Statistiche

Figura 6-36
Nodo Statistiche: scheda Impostazioni



Esamina. Selezionare il campo o i campi per i quali si desiderano statistiche riassuntive individuali. È possibile selezionare più campi.

Statistiche. Selezionare le statistiche da inserire nel report. Le opzioni disponibili includono Conteggio, Media, Somma, Min, Max, Intervallo, Varianza, Dev. std, Errore std di media, Mediana e Modalità.

Correla. Selezionare il campo o i campi che si desidera correlare. È possibile selezionare più campi. Una volta selezionati i campi di correlazione, nell'output verrà elencata la correlazione tra ogni campo Esamina e il campo o i campi di correlazione.

Impostazioni di correlazione. È possibile impostare una serie di opzioni per visualizzare l'intensità delle correlazioni nell'output.

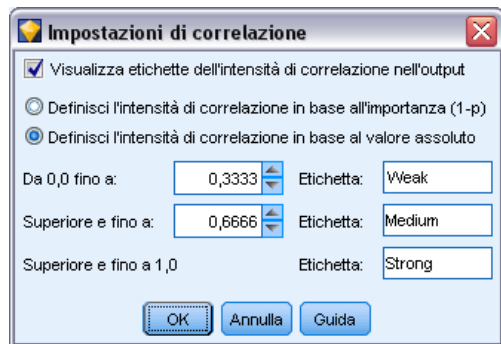
Impostazioni di correlazione

IBM® SPSS® Modeler è in grado di caratterizzare le correlazioni tramite etichette descrittive in modo da sottolineare le relazioni importanti. La **correlazione** misura l'intensità della relazione tra due campi continui (intervallo numerico). Accetta valori compresi tra -1.0 e 1.0 . I valori vicini a $+1.0$ indicano una forte associazione positiva: in questo caso, i valori alti in un campo sono associati ai valori alti nell'altro campo e i valori bassi sono associati ai valori bassi. Valori vicini a -1.0 indicano una forte associazione negativa: in questo caso, i valori alti in un campo sono associati ai valori bassi per l'altro campo e viceversa. Valori vicini a 0.0 indicano un'associazione debole, in questo caso i valori nei due campi sono più o meno indipendenti.

È possibile controllare la visualizzazione delle etichette di correlazione, modificare le soglie che definiscono le categorie e le etichette utilizzate per ogni intervallo. Dal momento che la caratterizzazione dei valori di correlazione dipende in modo rilevante dall'ambito del problema, è possibile che si desideri personalizzare gli intervalli e le etichette in base a una specifica situazione.

Figura 6-37

Finestra di dialogo Impostazioni di correlazione



Visualizza etichette dell'intensità di correlazione nell'output. Questa opzione è selezionata per default. Deselezionarla per omettere le etichette descrittive nell'output.

Intensità di correlazione. Esistono due opzioni per definire e identificare l'intensità delle correlazioni:

- **Definisci l'intensità di correlazione in base all'importanza (1-p).** Le correlazioni vengono identificate in base all'importanza, definita come 1 meno la significatività o 1 meno la probabilità che la differenza di media possa essere spiegata solo dal caso. Quanto più il valore si avvicina a 1 , tanto maggiore è la possibilità che i due campi *non* siano indipendenti o, in altre parole, che tra di loro esista un qualche tipo di relazione. In genere si consiglia di assegnare etichette alle correlazioni a seconda dell'importanza piuttosto che a seconda del valore assoluto, poiché in questo modo viene inclusa anche la variabilità dei dati: per esempio, un coefficiente di 0.6 può essere estremamente significativo in un insieme di dati e assolutamente non significativo in un altro. Per default, un'importanza con valore compreso

tra 0,0 e 0,9 viene definita di tipo *Debole*, una con valore compreso tra 0,9 e 0,95 viene definita di tipo *Medio* e una con valore compreso tra 0,95 e 1,0 viene definita di tipo *Forte*.

- **Definisci l'intensità di correlazione in base al valore assoluto.** Le correlazioni vengono identificate in base al valore assoluto del coefficiente di correlazione di Pearson, che è compreso fra -1 e 1 , come descritto in precedenza. Più il valore assoluto di questa misura si avvicina a 1 , più intensa è la correlazione. Per default, le correlazioni tra $0,0$ e $0,3333$ (in valore assoluto) sono etichettate come di tipo *Debole*, quelle tra $0,3333$ e $0,6666$ sono etichettate come di tipo *Medio* e quelle tra $0,6666$ e $1,0$ sono etichettate come di tipo *Forte*. Si noti, tuttavia, che è difficile generalizzare la significatività di qualsiasi valore da un insieme di dati a un altro; per questo motivo, nella maggior parte dei casi è consigliabile definire le correlazioni in base alla probabilità anziché in base al valore assoluto.

Browser di output del nodo Statistiche

Il browser di output del nodo Statistiche visualizza i risultati dell'analisi statistica e permette di eseguire determinate operazioni, per esempio selezionare campi, generare nodi in base alla selezione, salvare e stampare i risultati. Le normali opzioni di salvataggio, esportazione e stampa sono disponibili dal menu File e le normali opzioni di modifica sono disponibili dal menu Modifica. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output a pag. 348.](#)

Quando si visualizza per la prima volta l'output del nodo Statistiche, i risultati vengono espansi. Per nascondere i risultati, utilizzare il controllo di espansione di un elemento se si desidera comprimere solo i risultati di tale elemento oppure fare clic sul pulsante **Comprimi tutto** se si desidera nascondere tutti i risultati. Per visualizzare nuovamente i risultati dopo averli compressi, utilizzare il controllo di espansione a sinistra di un elemento per visualizzarne i risultati, oppure fare clic sul pulsante **Espandi tutto** per visualizzare tutti i risultati.

Figura 6-38
Browser di output del nodo Statistiche

Statistiche		
Conteggio		300
Media		1210.067
Min		400
Max		1980
Intervallo		1580
Varianza		203228.089
Deviazione standard		450.808
Errore standard della media		26.027

Correlazioni Pearson		
rainfall	-0.024	Weak
landquality	0.046	Weak
farmincome	0.543	Medium
claimvalue	0.486	Medium
\$N-claimvalue	0.493	Medium
\$E-claimvalue	0.534	Medium

L'output include una sezione per ogni campo *Esamina*, contenente una tabella delle statistiche richieste.

- **Conteggio.** Numero di record con valori validi per il campo.
- **Media.** Valore medio (media) del campo in tutti i record.
- **Somma.** Somma dei valori del campo in tutti i record.
- **Min.** Valore minimo del campo.
- **Max.** Valore massimo del campo.
- **Intervallo.** Differenza tra i valori minimi e massimi.
- **Varianza.** Misura della variabilità dei valori di un campo. Viene ottenuta calcolando la differenza tra ogni valore e la media globale, elevandola al quadrato, sommando tutti i valori e dividendo il risultato per il numero di record.
- **Deviazione standard.** Ulteriore misura della variabilità nei valori di un campo, calcolata come radice quadrata della varianza.
- **Errore standard della media.** Misura dell'incertezza nella stima della media del campo. Tale misura risulta utile se si presume che la media venga applicata a dati nuovi.
- **Mediana.** Valore "centrale" del campo, ovvero il valore che divide la metà superiore dei dati dalla metà inferiore (in base ai valori del campo).
- **Modalità.** Valore singolo più comune nei dati.

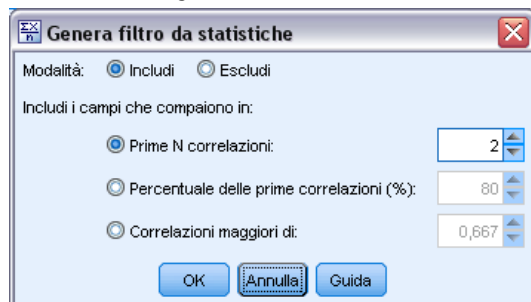
Correlazioni. Se sono stati specificati campi correlati, l'output contiene inoltre una sezione che elenca la correlazione di Pearson tra il campo Esamina e ogni campo correlato, nonché etichette descrittive facoltative per i valori di correlazione. [Per ulteriori informazioni, vedere l'argomento Impostazioni di correlazione a pag. 390.](#)

Menu Genera. Il menu Genera rende disponibili operazioni di generazione dei nodi.

- **Filtro.** Genera un nodo Filtro per filtrare i campi di output non correlati o con una correlazione debole con altri campi.

Generazione di un nodo Filtro da analisi statistiche

Figura 6-39
Finestra di dialogo Genera filtro da statistiche



Il nodo Filtro generato da un browser di output del nodo Statistiche filtrerà i campi in base alle relative correlazioni con altri campi. Il nodo Filtro generato ordina le correlazioni in base al valore assoluto mediante la creazione di un filtro che passa tutti i campi inclusi nelle correlazioni maggiori (secondo il criterio impostato nella finestra di dialogo).

Modalità. Specificare la modalità di selezione delle correlazioni. Includi consente di conservare i campi inclusi nelle correlazioni specificate. Escludi consente di filtrare tali campi.

Includi/Escludi i campi che compaiono in. Specificare il criterio per la selezione delle correlazioni.

- **Prime N correlazioni.** Seleziona il numero specificato di correlazioni e include/esclude i campi presenti in tali correlazioni.
- **Percentuale delle prime correlazioni (%).** Seleziona la percentuale specificata ($n\%$) di correlazioni e include/esclude i campi presenti in tali correlazioni.
- **Correlazioni maggiori di.** Seleziona le correlazioni maggiori in termini di valore assoluto rispetto alla soglia specificata.

Nodo Medie

Il nodo Medie confronta le medie tra gruppi indipendenti o coppie di campi correlati per verificare se esiste una differenza significativa. Per esempio, è possibile confrontare le entrate medie prima e dopo il lancio di una promozione, oppure confrontare le entrate determinate da clienti che non hanno ricevuto la promozione con quelli che l'hanno ricevuta.

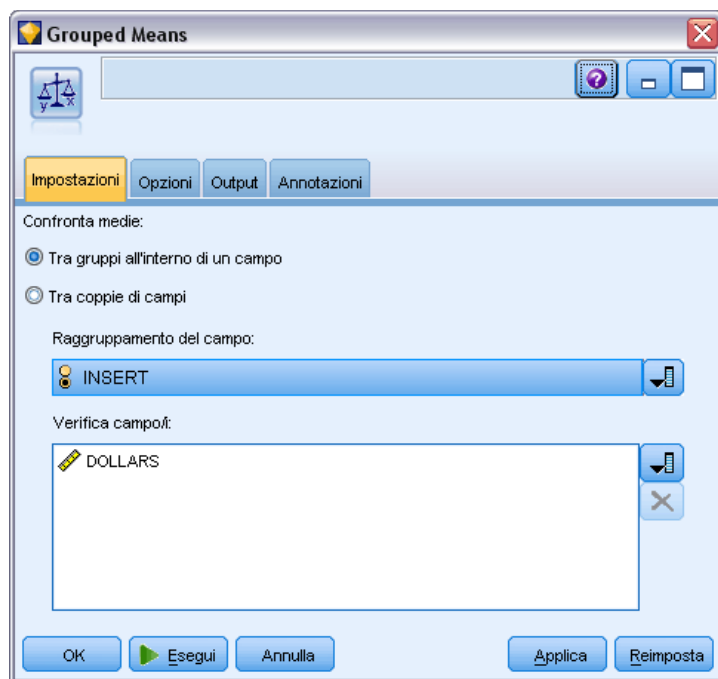
Le medie possono essere confrontate in due diversi modi, a seconda dei dati disponibili:

- **Tra gruppi all'interno di un campo.** Per confrontare gruppi indipendenti, selezionare un campo di verifica e un campo di raggruppamento. Per esempio, si potrebbe escludere un campione di clienti “di controllo” dall’invio di una promozione e confrontare le entrate medie del gruppo di controllo con quelle di tutti gli altri gruppi. In questo caso si imposterebbe un solo campo di verifica che indica le entrate per ogni singolo cliente, con un campo flag o nominale che indica se il cliente ha ricevuto l’offerta. I campioni sono indipendenti nel senso che ogni record viene assegnato all’uno o all’altro gruppo e non esiste un modo per collegare un membro specifico di un gruppo a un membro specifico di un altro. È possibile anche specificare un campo nominale con più di due valori per confrontare le medie di più gruppi. Quando viene eseguito, il nodo calcola un’analisi della varianza (ANOVA) monodirezionale sui campi selezionati. Nei casi in cui vi siano solo due gruppi di campi, i risultati dell’analisi ANOVA monodirezionale sono essenzialmente uguali a quelli ottenuti con un t test per campioni indipendenti. [Per ulteriori informazioni, vedere l'argomento Confronto di medie per gruppi indipendenti a pag. 394.](#)
- **Tra coppie di campi.** Quando si confrontano le medie di due campi correlati, è necessario che i gruppi siano in qualche modo accoppiati per ottenere risultati significativi. Per esempio, si potrebbero confrontare le entrate medie provenienti dallo stesso gruppo di clienti prima e dopo l’invio di una promozione, o confrontare le percentuali d’uso di un servizio tra coppie marito-moglie per verificare eventuali differenze. Ogni record contiene due misure distinte ma correlate che si possono confrontare in modo significativo. Quando viene eseguito, il nodo calcola un t test per campioni accoppiati su ogni coppia di campi selezionati. [Per ulteriori informazioni, vedere l'argomento Confronto delle medie tra campi accoppiati a pag. 395.](#)

Confronto di medie per gruppi indipendenti

Selezionare Tra gruppi all'interno di un campo nel nodo Medie per confrontare la media di due o più gruppi indipendenti.

Figura 6-40
Confronto delle medie tra gruppi all'interno di un campo



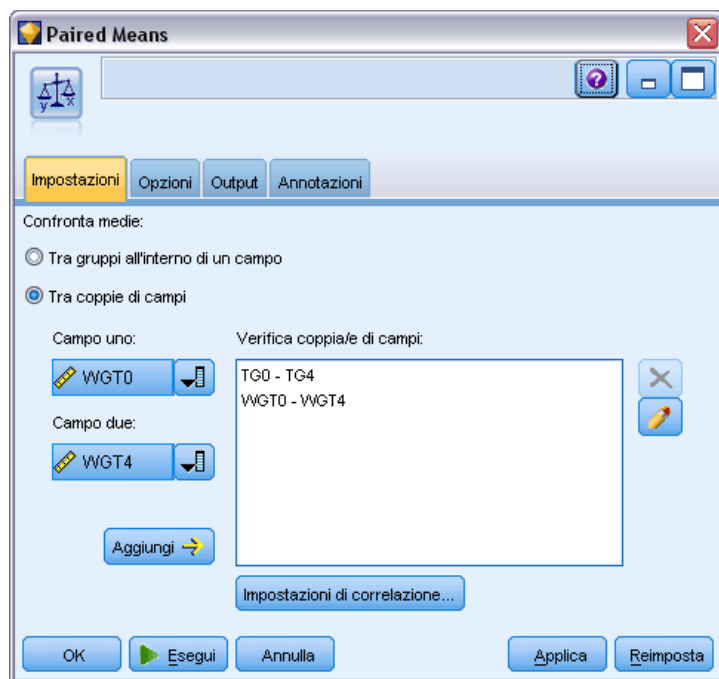
Campo di raggruppamento. Selezionare un flag numerico o un campo nominale con due o più valori distinti che divide i record nei gruppi da confrontare, per esempio i soggetti che hanno ricevuto un'offerta e quelli che non l'hanno ricevuta. Indipendentemente dal numero dei campi di verifica, è possibile selezionare un solo campo di raggruppamento.

Verifica campo/i. Selezionare uno o più campi numerici contenenti le misure da verificare. Per ogni campo selezionato verrà eseguita una verifica separata. Per esempio, si potrebbe verificare l'impatto di una determinata promozione sull'uso, sulle entrate e sul tasso di abbandono.

Confronto delle medie tra campi accoppiati

Selezionare Tra coppie di campi nel nodo Medie per confrontare le medie tra campi separati. Affinché i risultati siano significativi, i campi devono essere in qualche modo correlati (per esempio, entrate prima e dopo una promozione). È possibile anche selezionare più coppie di campi.

Figura 6-41
Confronto delle medie tra campi accoppiati



Campo uno. Selezionare un campo numerico contenente la prima delle misure da confrontare. In un'indagine di tipo "prima e dopo", questo sarebbe il campo "prima".

Campo due. Selezionare il secondo campo da confrontare.

Aggiungi. Aggiunge la coppia selezionata all'elenco Verifica coppia/e di campi.

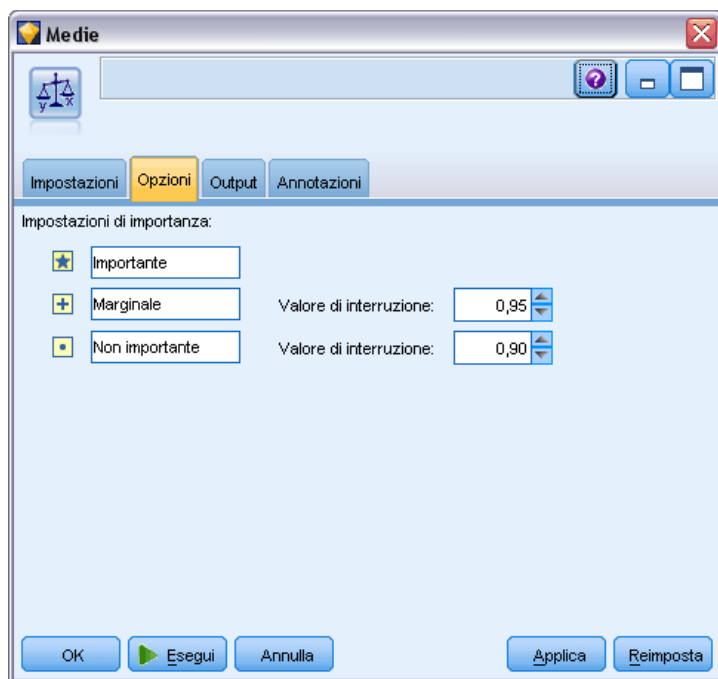
Se necessario, ripetere la selezione di campi per aggiungere più coppie all'elenco.

Impostazioni di correlazione. Consente di impostare le opzioni per identificare l'intensità delle correlazioni. [Per ulteriori informazioni, vedere l'argomento Impostazioni di correlazione a pag. 390.](#)

Opzioni del nodo Medie

La scheda Opzioni consente di impostare i valori p di soglia utilizzati per identificare i risultati come importanti, marginali o non importanti. È inoltre possibile modificare l'etichetta di ogni classificazione. L'importanza viene misurata su una scala percentuale e, in termini generici, può essere definita come 1 meno la probabilità di ottenere un risultato (quale la differenza tra le medie di due campi) altrettanto o più estremo di quello osservato determinato solo dal caso. Per esempio, un valore p superiore a 0,95 indica una probabilità inferiore al 5% che il risultato possa essere dovuto solamente al caso.

Figura 6-42
Impostazioni di importanza



Etichette di importanza. È possibile modificare le etichette utilizzate per identificare ogni coppia o gruppo di campi nell'output. Le etichette di default sono *importante*, *marginale* e *non importante*.

Valori di interruzione. Specifica la soglia per ogni rango. In genere, i valori p maggiori di 0,95 classificano gli elementi come importanti, mentre quelli minori di 0,9 sono classificati come non importanti, ma queste soglie si possono modificare in funzione delle esigenze.

Nota: le misure di importanza sono disponibili in numerosi nodi. I calcoli specifici dipendono dal nodo e dal tipo di obiettivo e di campi di input utilizzati, ma i valori possono essere comunque confrontati in quanto sono tutti misurati in base a una scala percentuale.

Browser di output del nodo Medie

Il browser di output del nodo Medie visualizza i dati delle tavole di contingenza e consente di eseguire operazioni standard quali selezionare e copiare la tabella una riga per volta, ordinarla in base a qualsiasi colonna, salvarla e stamparla. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output a pag. 348.](#)

Le informazioni specifiche contenute nella tabella dipendono dal tipo di confronto (gruppi all'interno di un campo o campi separati).

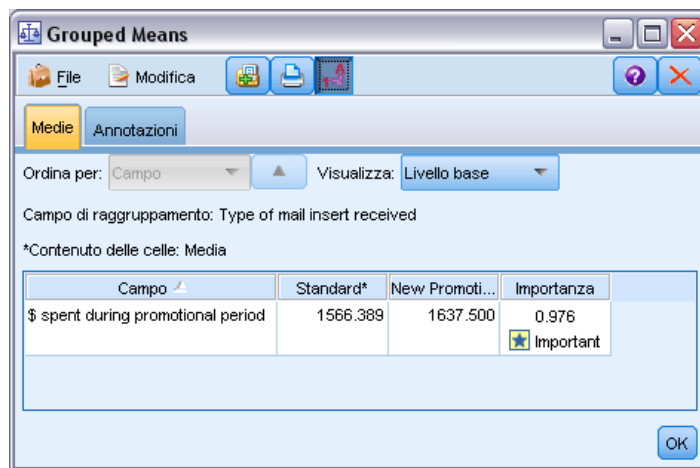
Ordina per. Consente di ordinare l'output in base a una colonna specifica. Fare clic sulla freccia su o giù per modificare la direzione dell'ordinamento. In alternativa, è possibile fare clic sull'intestazione di qualsiasi colonna per ordinare in base a quella colonna. Fare clic nuovamente per cambiare la direzione dell'ordinamento all'interno della colonna.

Visualizza. È possibile selezionare Livello base o Opzioni avanzate per controllare il livello di dettaglio visualizzato. La visualizzazione con opzioni avanzate comprende tutte le informazioni contenute in quella di livello base, ma con una serie di dettagli aggiuntivi.

Output medie - Confronto tra gruppi all'interno di un campo

Quando si confrontano gruppi all'interno di un campo, il nome del campo di raggruppamento è visualizzato sopra la tabella di output e le medie e le relative statistiche sono riportate separatamente per ogni gruppo. La tabella comprende una riga separata per ogni campo di verifica.

Figura 6-43
Confronto tra gruppi all'interno di un campo



Sono visualizzate le seguenti colonne:

- **Campo.** Elenca i nomi dei campi di verifica selezionati.
- **Medie di gruppo.** Visualizza la media per ogni categoria del campo di raggruppamento. Per esempio, è possibile confrontare i soggetti che hanno ricevuto un'offerta speciale (*Nuova promozione*) con quelli che non l'hanno ricevuta (*Standard*). Nella visualizzazione con opzioni avanzate sono visualizzati anche la deviazione standard, l'errore standard e il conteggio.
- **Importanza.** Visualizza il valore e l'etichetta dell'importanza. [Per ulteriori informazioni, vedere l'argomento Opzioni del nodo Medie a pag. 396.](#)

Output avanzato

La visualizzazione con opzioni avanzate contiene le seguenti colonne aggiuntive.

- **F-Test.** Questo test si basa sul rapporto tra varianza fra i gruppi e varianza all'interno di ogni gruppo. Se le medie sono uguali per tutti i gruppi, il rapporto F dovrebbe avvicinarsi a 1 poiché entrambi i valori sono stime della varianza della stessa popolazione. Maggiore è il rapporto, maggiori sono la variazione tra i gruppi e la probabilità che esista una differenza significativa.
- **df.** Visualizza i gradi di libertà.

Output Medie - Confronto tra coppie di campi

Quando si confrontano campi separati, la tabella di output comprende una riga per ogni coppia di campi selezionata.

Figura 6-44
Confronto tra coppie di campi

Campo uno	Campo due	Media uno*	Media due*	Correlazione	Differenza m...	Importanza
Triglyceride	Final triglyce...	138.438	124.375	-0.286 Weak	14.062	0.751 Unimportant
Weight	Final weight	198.375	190.312	0.996 Strong	8.062	1.000 Important

- **Campo uno/due.** Visualizza il nome del primo e del secondo campo di ogni coppia. Nella visualizzazione con opzioni avanzate sono visualizzati anche la deviazione standard, l'errore standard e il conteggio.
- **Media uno/due.** Visualizza la media rispettiva dei due campi.
- **Correlazione.** Misura l'intensità della relazione tra due campi continui (intervallo numerico). I valori vicini a +1.0 indicano una forte associazione positiva; i valori vicini a -1.0 indicano una forte associazione negativa. [Per ulteriori informazioni, vedere l'argomento Impostazioni di correlazione a pag. 390.](#)
- **Differenza media.** Visualizza la differenza tra le medie dei due campi.
- **Importanza.** Visualizza il valore e l'etichetta dell'importanza. [Per ulteriori informazioni, vedere l'argomento Opzioni del nodo Medie a pag. 396.](#)

Output avanzato

L'opzione Output avanzato aggiunge le seguenti colonne:

Intervallo di confidenza del 95%. Limiti inferiore e superiore dell'intervallo all'interno del quale è probabile che si trovi la media vera nel 95% di tutti i possibili campioni di quelle dimensioni e di quella popolazione.

T-Test. La statistica t si ottiene dividendo la differenza media per il suo errore standard. Maggiore è il valore assoluto di questa statistica, maggiore è la probabilità che le medie non siano uguali.

df. Visualizza i gradi di libertà della statistica.

Nodo Report

Il nodo Report consente di creare report formattati che contengono sia testo fisso sia dati e altre espressioni derivate dai dati. Il formato del report viene specificato utilizzando modelli di testo per definire il testo fisso e le costruzioni di output dei dati. È possibile fornire una formattazione personalizzata del testo utilizzando tag HTML nel modello e impostando apposite opzioni nella scheda Output. I valori dei dati e il restante output condizionale vengono inclusi nel report utilizzando espressioni CLEM nel modello.

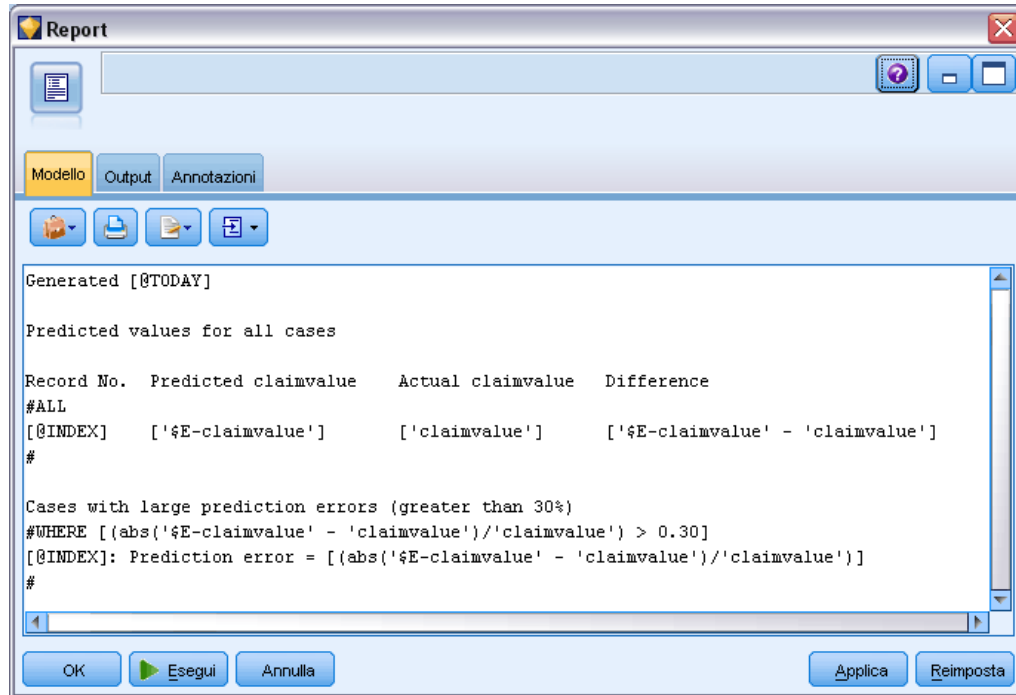
Alternative al nodo Report

Il nodo Report viene soprattutto utilizzato per elencare i record o i casi prodotti da uno stream, per esempio tutti i record che soddisfano una determinata condizione. In questo senso, può essere considerato come un'alternativa meno strutturata al nodo Tabella.

- Se si desidera produrre un report che, invece dei dati veri e propri, elenchi informazioni sui campi o altri elementi definiti nello stream (per esempio le definizioni di campi specificate in un nodo Tipo), è possibile utilizzare uno script. [Per ulteriori informazioni, vedere l'argomento Report nodo Tipo in il capitolo 6 in IBM SPSS Modeler 14.2 Guida per script e automazione.](#)
- Per generare un report che includa più oggetti di output (quali una raccolta di modelli, tabelle e grafici generati da uno o più stream) e che possano essere prodotti in più formati (come testo, HTML e Microsoft Word/Office) è possibile utilizzare un progetto IBM® SPSS® Modeler. [Per ulteriori informazioni, vedere l'argomento Introduzione ai progetti in il capitolo 11 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- Per generare un elenco di nomi di campi senza utilizzare la funzionalità di script, è possibile utilizzare un nodo Tabella preceduto da un nodo Campione che scarta tutti i record. In questo modo viene generata una tabella senza righe, che può essere trasposta nelle esportazioni per creare un elenco di nomi di campi in un'unica colonna (a questo scopo, selezionare Trasponi dati nella scheda Output del nodo Tabella).

Scheda Modello del nodo Report

Figura 6-45
Nodo Report: scheda Modello



Creazione di un modello. Per definire il contenuto del report, generare un modello nella scheda Modello del nodo Report. Il modello è formato da righe di testo, ognuna delle quali contiene una specifica relativa al contenuto del report, e da speciali righe di tag utilizzate per indicare l'ambito delle righe del contenuto. In ogni riga del contenuto, le espressioni CLEM racchiuse tra parentesi quadre ([]) vengono valutate prima dell'invio della riga al report. Per una riga del modello esistono tre possibili ambiti:

Fissa. Le righe che non hanno contrassegni di altro genere sono considerate fisse. Le righe fisse vengono copiate nel report solo una volta, dopo la valutazione delle espressioni in esse contenute. Per esempio, la riga

```
Questo report è stato generato [@TODAY]
```

copia nel report una singola riga, contenente il testo e la data corrente.

Globale (iterazione completa). Le righe tra i tag speciali #ALL e # vengono copiate nel report una volta per ogni record di dati di input. Le espressioni CLEM (racchiuse tra parentesi) vengono valutate in base al record corrente per ogni riga di output. Per esempio, le righe

```
#ALL
Per il record [@INDEX], il valore di Età è [Età]
#
```

includono una riga per ogni record che indichi il numero di record e l'età.

Per generare un elenco di tutti i record:

```
#ALL
[Età] [Sesso] [Colesterolo] [Pressione]
#
```

Condizionale (iterazione condizionale). Le righe tra i tag speciali *<condizione>* #WHERE e # vengono copiate nel report una volta per ogni record in cui la condizione specificata è vera. La condizione è un'espressione CLEM. (Nella condizione WHERE, le parentesi sono facoltative.) Per esempio, le righe

```
#WHERE [SESSO = 'M']
Maschio per il record n. [@INDEX] ha età [Età].
#
```

scrivono nel file una riga per ogni record con valore *M* per il sesso. Il report completo conterrà le righe fisse, globali e condizionali definite applicando il modello ai dati di input.

È possibile specificare le opzioni di visualizzazione o salvataggio dei risultati mediante la scheda Output, comune a diversi tipi di nodi Output. [Per ulteriori informazioni, vedere l'argomento Scheda Output del nodo Output a pag. 355.](#)

Output di dati in formato HTML o XML

I tag HTML o XML possono essere inclusi direttamente nel modello per scrivere report in uno di questi due formati. Per esempio, il seguente modello genera una tabella HTML.

This report is written in HTML.
Only records where Age is above 60 are included.

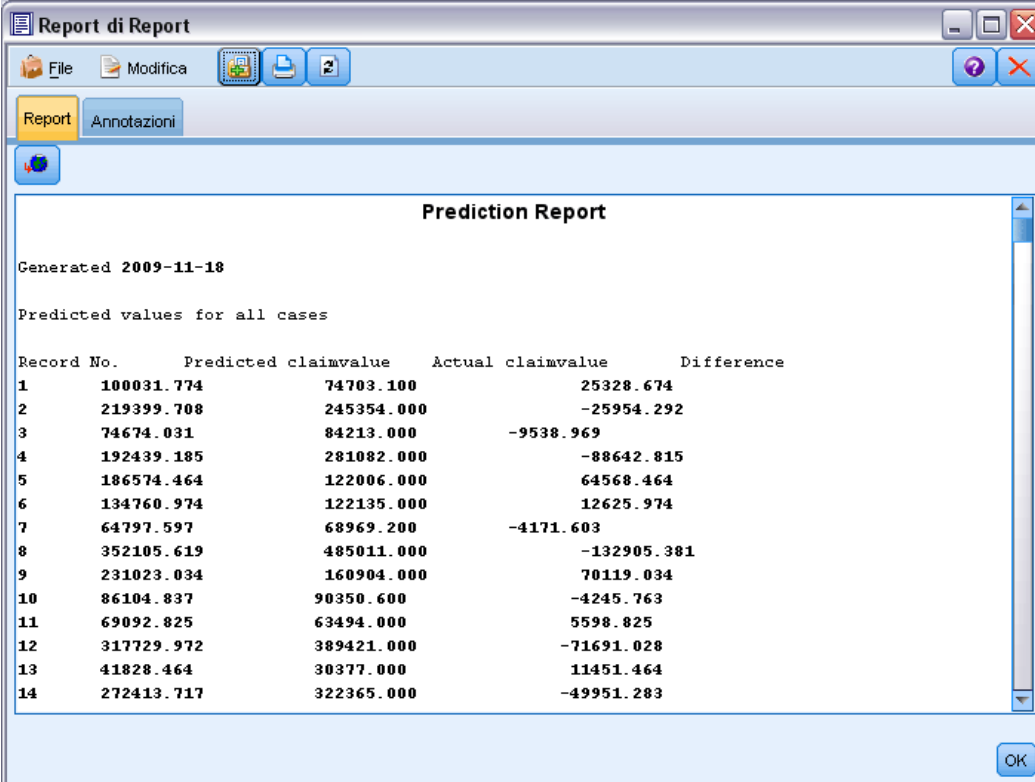
```
<HTML>
<TABLE border="2">
  <TR>
    <TD>Età</TD>
    <TD>Pressione</TD>
    <TD>Colesterolo</TD>
    <TD>Cura</TD>
  </TR>

  #WHERE Età > 60
  <TR>
    <TD>[Età]</TD>
    <TD>[Pressione]</TD>
    <TD>[Colesterolo]</TD>
    <TD>[Cura]</TD>
  </TR>
#
</TABLE>
</HTML>
```

Browser di output del nodo Report

Il browser del nodo Report mostra il contenuto del report generato. Le normali opzioni di salvataggio, esportazione e stampa sono disponibili dal menu File e le normali opzioni di modifica sono disponibili dal menu Modifica. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output a pag. 348.](#)

Figura 6-46
Browser del nodo Report



The screenshot shows a window titled 'Report di Report' with a menu bar containing 'File' and 'Modifica'. Below the menu bar are two tabs: 'Report' (selected) and 'Annotazioni'. The main content area displays a 'Prediction Report' with the following text and table:

Generated **2009-11-18**

Predicted values for all cases

Record No.	Predicted claimvalue	Actual claimvalue	Difference
1	100031.774	74703.100	25328.674
2	219399.708	245354.000	-25954.292
3	74674.031	84213.000	-9538.969
4	192439.185	281082.000	-88642.815
5	186574.464	122006.000	64568.464
6	134760.974	122135.000	12625.974
7	64797.597	68969.200	-4171.603
8	352105.619	485011.000	-132905.381
9	231023.034	160904.000	70119.034
10	86104.837	90350.600	-4245.763
11	69092.825	63494.000	5598.825
12	317729.972	389421.000	-71691.028
13	41828.464	30377.000	11451.464
14	272413.717	322365.000	-49951.283

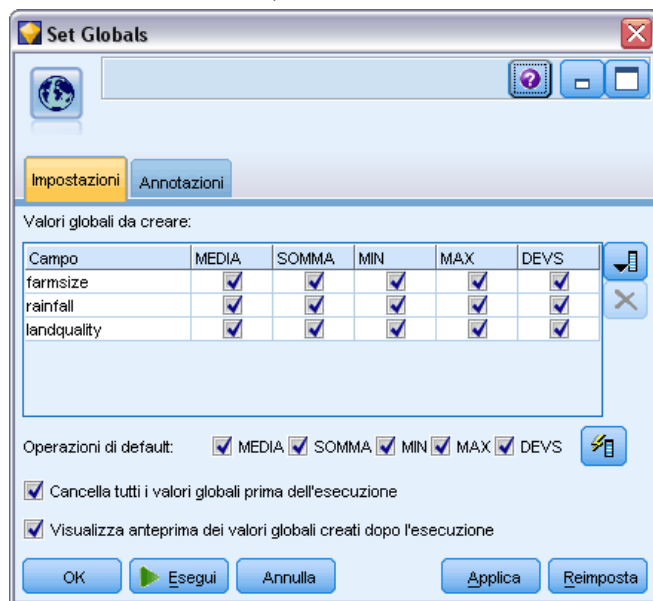
An 'OK' button is located at the bottom right of the window.

Nodo Globali

Il nodo Globali analizza i dati e calcola i valori di riepilogo che possono essere utilizzati nelle espressioni CLEM. Per esempio, è possibile utilizzare un nodo Globali per calcolare le statistiche di un campo denominato *età* e utilizzare quindi la media globale dell'*età* nelle espressioni CLEM inserendo la funzione `@GLOBAL_MEAN(age)`. [Per ulteriori informazioni, vedere l'argomento Panoramica di riferimento su CLEM in il capitolo 8 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Scheda Impostazioni del nodo Calcola globali

Figura 6-47
Nodo Globali: scheda Impostazioni



Valori globali da creare. Selezionare il campo o i campi per i quali si desidera che siano disponibili valori globali. È possibile selezionare più campi. Per ogni campo specificare le statistiche da calcolare, accertandosi che siano selezionate nelle colonne accanto al nome del campo.

- **MEDIA.** Valore medio (media) del campo in tutti i record.
- **SOMMA.** Somma dei valori del campo in tutti i record.
- **MIN.** Valore minimo del campo.
- **MAX.** Valore massimo del campo.
- **DEVS.** Deviazione standard, che misura la variabilità dei valori di un campo ed è calcolata come radice quadrata della varianza.

Operazioni di default. Le opzioni qui selezionate verranno utilizzate quando verranno aggiunti nuovi campi all'elenco Valori globali sopra riportato. Per modificare l'insieme di default delle statistiche, selezionare o deselezionare le statistiche come desiderato. È inoltre possibile utilizzare il pulsante **Applica** per applicare le operazioni di default a tutti i campi compresi nell'elenco.

Cancella tutti i valori globali prima dell'esecuzione. Selezionare questa opzione per rimuovere tutti i valori globali prima di calcolare nuovi valori. Se questa opzione non è selezionata, i nuovi valori calcolati sostituiranno i precedenti, tuttavia i valori globali non ricalcolati continueranno a essere disponibili.

Visualizza anteprima dei valori globali creati dopo l'esecuzione. Se si seleziona questa opzione, dopo l'esecuzione verrà visualizzata la scheda Valori globali della finestra di dialogo delle proprietà dello stream contenente i valori globali calcolati. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dei valori globali degli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

IBM SPSS Statistics Applicazioni di supporto

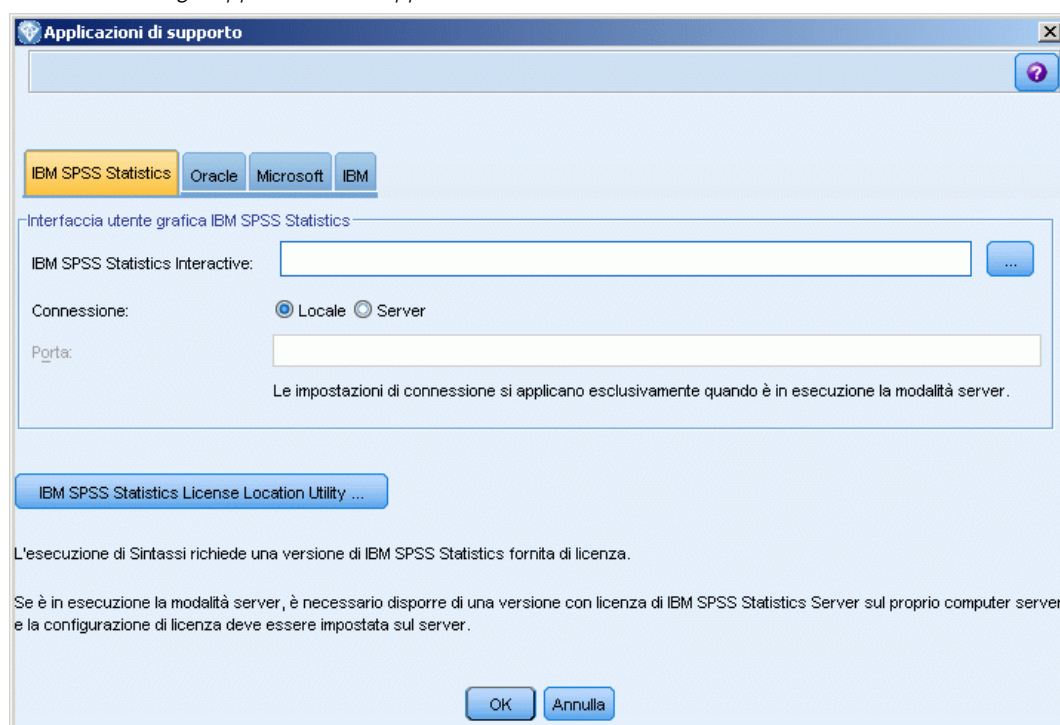
Se sul computer è installata una versione compatibile di IBM® SPSS® Statistics con relativa licenza, è possibile configurare IBM® SPSS® Modeler in modo da elaborare i dati con la funzionalità SPSS Statistics utilizzando i nodi Trasformazioni Statistics, Modelli Statistics, Output Statistics o Esporta Statistics.

- Per configurare SPSS Modeler in modo che possa funzionare con SPSS Statistics e altre applicazioni, scegliere:

Strumenti > Opzioni > Applicazioni di supporto

Figura 6-48

Finestra di dialogo Applicazioni di supporto



IBM SPSS Statistics Interactive. Immettere il percorso completo e il nome del comando (per esempio, *C:\Programmi\IBM\SPSS\Statistics\<nn>\stats.exe*) da utilizzare per l'avvio di SPSS Statistics direttamente in un file di dati prodotto dal nodo Esporta Statistics. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta Statistics in il capitolo 8 a pag. 449.](#)

Connessione. Se SPSS Statistics Server e IBM® SPSS® Modeler Server sono installati nello stesso server, è possibile attivare una connessione tra le due applicazioni, in modo da incrementare il livello di efficienza mantenendo i dati sul server durante l'analisi. Selezionare Server per attivare l'opzione Porta riportata di seguito. L'impostazione di default è Locale.

Porta: Specificare la porta del server per SPSS Statistics Server.

IBM SPSS Statistics License Location Utility. Per consentire a SPSS Modeler di utilizzare i nodi Trasformazioni Statistics, Modelli Statistics e Output Statistics, è necessario che sul computer dove viene eseguito lo stream sia installata una copia di SPSS Statistics con relativa licenza. Inoltre, in

caso di esecuzione in modalità distribuita con SPSS Modeler Server remoto, è necessario che sul computer SPSS Modeler client sia installata una copia di SPSS Statistics con relativa licenza.

- Se si esegue SPSS Modeler in modalità locale (autonoma), la copia concessa in licenza di SPSS Statistics deve trovarsi sul computer locale. Fare clic su questo pulsante per specificare la posizione dell'installazione SPSS Statistics locale da utilizzare per la licenza.
- Inoltre, se si esegue il programma in modalità distribuita con un computer SPSS Modeler Server remoto, sul computer server deve essere presente una versione di SPSS Statistics con relativa licenza e anche la configurazione della licenza deve essere effettuata sul server. Per eseguire questa operazione, passare dal prompt dei comandi alla directory SPSS Modeler Server *bin* e, per Windows, eseguire:

```
statisticsutility -location=<percorso del file di licenza di IBM SPSS Statistics Server>/bin
```

In alternativa, per UNIX eseguire:

```
./statisticsutility -location=<percorso del file di licenza di IBM SPSS Statistics Server>/bin
```

in cui *<percorso del file licenza di SPSS Statistics Server>* è la directory di installazione di una copia di SPSS Statistics Server dotata di licenza.

Se sul computer locale non è installata una copia di SPSS Statistics con relativa licenza, è comunque possibile eseguire il nodo File Statistics con una copia di SPSS Statistics server dotata di licenza, ma se si tenta di eseguire altri nodi di SPSS Statistics viene visualizzato un messaggio di errore.

Commenti

In caso di problemi nell'esecuzione dei nodi Proc SPSS Statistics, tenere presente quanto segue:

- Se i nomi dei campi utilizzati in SPSS Modeler superano gli otto caratteri (per le versioni precedenti a SPSS Statistics 12.0), superano i 64 caratteri (per SPSS Statistics 12.0 e versioni successive) oppure contengono caratteri non validi, è necessario rinominarli o troncarli prima che possano essere letti in SPSS Statistics. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics in il capitolo 8 a pag. 451.](#)
- Se SPSS Statistics è stato installato dopo SPSS Modeler, potrebbe essere necessario specificare il percorso del file licenza di SPSS Statistics, come illustrato in precedenza.

Nodi di esportazione

Panoramica sui nodi di esportazione

I nodi di esportazione consentono di esportare i dati in vari formati per l'interfacciamento con gli altri strumenti software in uso.

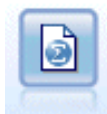
Sono disponibili i seguenti nodi di esportazione:



Il nodo Database scrive dati in una sorgente dati relazionale compatibile con ODBC. Per scrivere in una sorgente dati ODBC, è necessario utilizzare una sorgente dati esistente e disporre dell'autorizzazione in scrittura per tale sorgente. [Per ulteriori informazioni, vedere l'argomento Nodo di esportazione Database a pag. 408.](#)



Il nodo di esportazione File piatto restituisce dati in un file di testo delimitato. È utile per esportare i dati che possono essere letti da altri software di analisi o fogli di calcolo. [Per ulteriori informazioni, vedere l'argomento Nodo di esportazione File piatto a pag. 419.](#)



Il nodo Esporta Statistics restituisce i dati nel formato *.sav* di IBM® SPSS® Statistics. I file *.sav* possono essere letti da SPSS Statistics Base e altri prodotti. Questo formato viene inoltre utilizzato per i file cache di IBM® SPSS® Modeler. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta Statistics in il capitolo 8 a pag. 449.](#)



Il nodo Esportazione di IBM® SPSS® Data Collection esegue l'output di dati nel formato utilizzato dal software di ricerche di mercato Data Collection. Per utilizzare questo nodo, è necessario che sia installata Data Collection Data Library. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta IBM SPSS Data Collection a pag. 420.](#)



Il nodo Esporta SAS restituisce nel formato SAS i dati che devono essere letti in SAS o in un pacchetto software compatibile con SAS. I formati di file SAS disponibili sono tre: SAS per Windows/OS2, SAS per UNIX o SAS Versione 7/8. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta SAS a pag. 426.](#)



Il nodo Esportazione da Excel esegue l'output di dati in formato Microsoft Excel (*.xls*). Se lo si desidera, è possibile scegliere di avviare Excel automaticamente e aprire il file esportato quando si esegue il nodo. [Per ulteriori informazioni, vedere l'argomento Nodo Esportazione da Excel a pag. 427.](#)



Il nodo Esporta XML restituisce i dati in un file in formato XML. Se lo si desidera, è possibile creare un nodo di input XML per leggere nuovamente i dati esportati nello stream. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta XML a pag. 428.](#)

Nodo di esportazione Database

È possibile usare i nodi Database per scrivere dati su sorgenti di dati relazionali compatibili con ODBC, come illustrato nella descrizione del nodo di input Database. [Per ulteriori informazioni, vedere l'argomento Nodo di input Database in il capitolo 2 a pag. 13.](#)

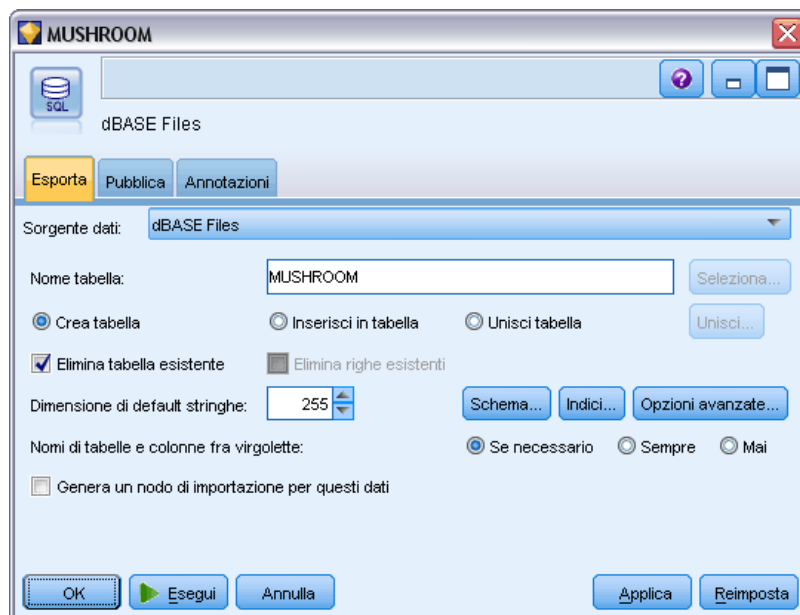
Per scrivere dati in un database, utilizzare i passaggi seguenti:

- ▶ Installare un driver ODBC e configurare una sorgente dati per il database che si desidera utilizzare.
- ▶ Nella scheda Esporta del nodo Database, specificare la sorgente dati e la tabella in cui scrivere i dati. È possibile creare una nuova tabella o inserire i dati in una tabella esistente.
- ▶ Se necessario, impostare altre opzioni.

Questi passaggi verranno descritti in dettaglio negli argomenti seguenti.

Scheda di esportazione del nodo Database

Figura 7-1
Nodo di esportazione Database, scheda Esporta



Sorgente dati. Mostra la sorgente dati selezionata. Immettere il nome o selezionarlo dall'elenco a discesa. Se nell'elenco non compare il database desiderato, selezionare *Aggiungi nuova connessione database* e individuare il database dalla finestra di dialogo *Connessioni database*. [Per ulteriori informazioni, vedere l'argomento Aggiunta di una connessione a un database in il capitolo 2 a pag. 16.](#)

Nome tabella. Immettere il nome della tabella alla quale si desidera inviare i dati. Se si seleziona l'opzione *Inserisci in tabella*, è possibile selezionare una tabella esistente del database facendo clic sul pulsante *Seleziona*.

Crea tabella. Selezionare questa opzione per creare una nuova tabella di database o per sovrascriverne una esistente.

Inserisci in tabella. Selezionare questa opzione per inserire i dati come nuove righe in una tabella di database esistente.

Unisci tabella. (Dove disponibile) Selezionare questa opzione per aggiornare le colonne del database selezionate con i valori dei campi dati di origine corrispondenti. La selezione di questa opzione abilita il pulsante Unione, che visualizza una finestra di dialogo da cui è possibile mappare i campi dati di origine alle colonne del database.

Elimina tabella esistente. Selezionare questa opzione per eliminare qualsiasi tabella esistente con lo stesso nome della nuova tabella che si desidera creare.

Elimina righe esistenti. Selezionare questa opzione per eliminare dalla tabella le righe esistenti prima dell'esportazione quando si inseriscono dati in una tabella.

Nota: se è selezionata una delle due opzioni sopra riportate, durante l'esecuzione del nodo verrà visualizzato un messaggio Avviso di sovrascrittura. Per eliminare gli avvisi, deselezionare Avviso quando un nodo sovrascrive una tabella di database nella scheda Notifiche della finestra di dialogo Opzioni utente. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di notifica in il capitolo 12 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Dimensioni di default stringhe. I campi contrassegnati come senza tipo in un nodo Tipo a monte vengono scritti nel database come campi stringa. Specificare la dimensione delle stringhe da utilizzare per i campi senza tipo.

Fare clic su Schema per visualizzare una finestra di dialogo in cui è possibile impostare tipi di dati SQL per i campi e specificare la chiave principale ai fini dell'indicizzazione del database. [Per ulteriori informazioni, vedere l'argomento Esportazione Database - Opzioni Schema a pag. 412.](#)

Fare clic su Indici per impostare le opzioni di indicizzazione della tabella esportata al fine di migliorare le prestazioni del database. [Per ulteriori informazioni, vedere l'argomento Esportazione Database - Opzioni Schema a pag. 412.](#)

Fare clic su Opzioni avanzate per specificare le opzioni di caricamento di massa e di commit nel database. [Per ulteriori informazioni, vedere l'argomento Esportazione database - Opzioni avanzate a pag. 417.](#)

Nomi di tabelle e colonne fra virgolette. Selezionare le opzioni da utilizzare al momento dell'invio di un'istruzione CREATE TABLE al database. Le tabelle e le colonne con spazi o caratteri non standard devono essere racchiuse tra virgolette.

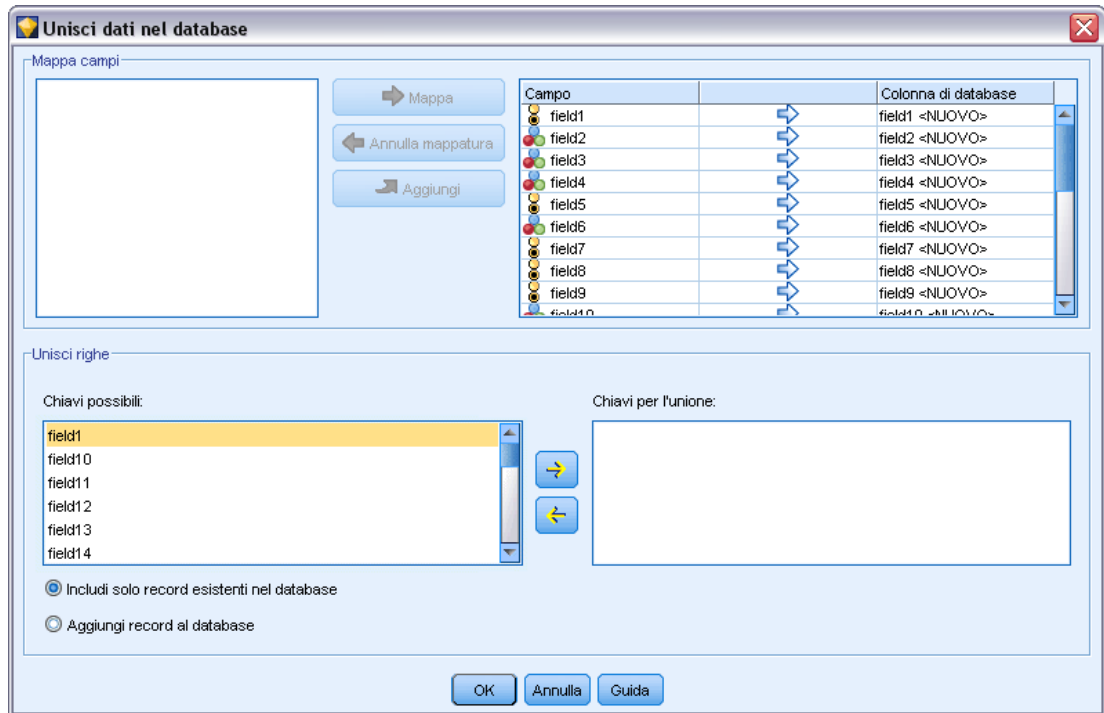
- **Se necessario.** Selezionare questa opzione per consentire a IBM® SPSS® Modeler di determinare automaticamente l'eventuale esigenza di virgolette per ogni singolo caso.
- **Sempre.** Selezionare questa opzione per racchiudere sempre i nomi di tabella e di colonna tra virgolette.
- **Mai.** Selezionare questa opzione per disattivare l'utilizzo delle virgolette.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare un nodo di input Database per i dati esportati nella sorgente dati e nella tabella specificate. Dopo l'esecuzione, questo nodo verrà aggiunto all'area di disegno dello stream.

Opzioni di unione esportazione database

Questa finestra di dialogo consente di mappare i campi dai dati di origine alle colonne della tabella di database di destinazione. Se un campo dati di origine viene mappato a una colonna del database, quando si esegue lo stream il valore della colonna viene sostituito dal valore dei dati di origine. I campi di input non mappati restano invariati nel database.

Figura 7-2
Mappatura dei campi dati di origine alle colonne del database



Mappa campi. Qui è possibile specificare la mappatura tra i campi dati di origine e le colonne del database. I campi dati di origine con lo stesso nome delle colonne del database vengono mappati automaticamente.

- **Mappa.** Mappa un campo dati di origine selezionato nell'elenco dei campi a sinistra del pulsante a una colonna del database selezionata nell'elenco a destra. È possibile mappare più campi contemporaneamente, ma il numero delle voci selezionate in entrambi gli elenchi deve essere lo stesso.
- **Annulla mappatura.** Elimina la mappatura per una o più colonne del database selezionate. Il pulsante viene attivato quando si seleziona un campo o una colonna di database nella tabella a destra della finestra di dialogo.
- **Aggiungi.** Aggiunge uno o più campi dati di origine selezionati nell'elenco dei campi a sinistra del pulsante all'elenco di destra dei campi da mappare. Il pulsante viene attivato quando si seleziona un campo nell'elenco a sinistra e non è presente alcun campo con lo stesso nome nell'elenco a destra. Se si fa clic sul pulsante, il campo selezionato viene mappato su una nuova colonna di database con lo stesso nome. Il termine <NEW> viene visualizzato dopo il nome della colonna di database a indicare che si tratta di un nuovo campo.

Unisci righe. Utilizza un campo chiave, per esempio *ID transazione*, per unire record con lo stesso valore del campo chiave. Questa opzione è equivalente a un “join di uguaglianza” (equi-join) di database. I valori chiave devono essere quelli delle chiavi principali, ovvero devono essere univoci e non possono contenere valori nulli.

- **Chiavi possibili.** Elenca tutti i campi trovati in tutte le sorgenti dati di input. Selezionare uno o più campi dall’elenco e utilizzare il pulsante con la freccia per aggiungerli come campi chiave da utilizzare per l’unione dei record. Tutti i campi della mappa a cui corrisponde una colonna del database mappata possono essere utilizzati come chiave, tranne quelli aggiunti come nuove colonne del database (contraddistinti da <NEW> dopo il nome).
- **Chiavi per l’unione.** Elenca tutti i campi utilizzati per unire i record di tutte le sorgenti dati di input in base ai valori dei campi chiave. Per rimuovere una chiave dall’elenco, selezionarla e utilizzare il pulsante con la freccia per includerla nuovamente nell’elenco Chiavi possibili. L’opzione seguente è attivata solo quando vengono selezionati più campi.
- **Includi solo record esistenti nel database.** Esegue un join parziale; se il record è presente sia nel database che nello stream, i campi mappati vengono aggiornati.
- **Aggiungi record al database.** Esegue un join esterno; tutti i record dello stream vengono uniti (se nel database è presente lo stesso record) o aggiunti (se il record non esiste ancora nel database).

Per mappare un campo dati di origine a una nuova colonna del database

- ▶ Fare clic sul nome del campo di input nell’elenco a sinistra, sotto Mappa campi.
- ▶ Fare clic sul pulsante Aggiungi per completare la mappatura.

Per mappare un campo dati di origine a una colonna del database esistente

- ▶ Fare clic sul nome del campo di input nell’elenco a sinistra, sotto Mappa campi.
- ▶ Fare clic sul nome della colonna in Colonna di database sulla destra.
- ▶ Fare clic sul pulsante Mappa per completare la mappatura.

Per rimuovere una mappatura

- ▶ Nell’elenco a destra, sotto Campo, fare clic sul nome del campo di cui si desidera rimuovere la mappatura.
- ▶ Fare clic sul pulsante Annulla mappatura.

Per deselezionare un campo in uno degli elenchi

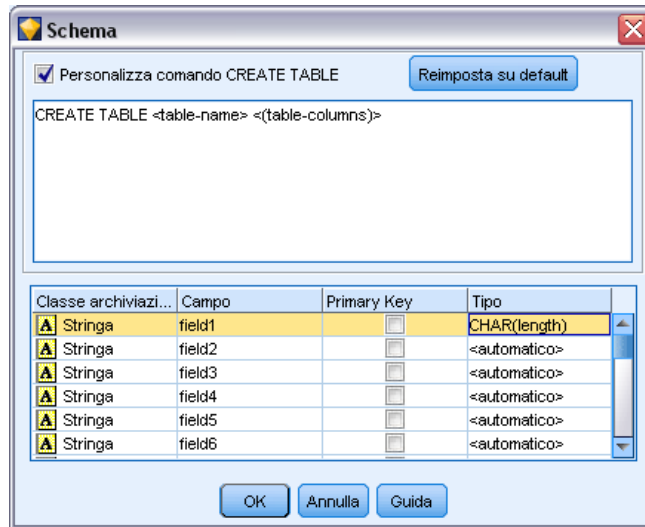
- ▶ Tenere premuto il tasto CTRL e fare clic sul nome del campo.

Esportazione Database - Opzioni Schema

La finestra di dialogo Schema di esportazione del database consente di impostare le opzioni di esportazione in un database InfoSphere Warehouse e i tipi di dati SQL per i campi, di specificare quali campi sono chiavi principali e di personalizzare l'istruzione CREATE TABLE generata durante l'esportazione.

Figura 7-3

Finestra di dialogo Schema del nodo di esportazione Database



La finestra di dialogo è composta da varie parti:

- La sezione in alto contiene le opzioni per l'esportazione in un database InfoSphere Warehouse (IBM DB2 9.7 o versioni successive). Questa sezione non è visualizzata se non si è collegati a un database di questo tipo.
- Il campo di testo al centro visualizza il modello utilizzato per generare il comando CREATE TABLE, che per default segue il formato:
CREATE TABLE <table-name> <(table columns)>
- La tabella nella parte inferiore della finestra consente di indicare il tipo di dati SQL di ogni campo e di specificare quali campi sono chiavi principali, come illustrato sotto. La finestra di dialogo genera automaticamente i valori dei parametri <table-name> e <(table columns)> in base alle specifiche contenute nella tabella.

Impostazione delle opzioni InfoSphere Warehouse

Per l'esportazione in un database InfoSphere Warehouse è possibile specificare numerose impostazioni.

Spazio di tabella. Lo spazio di tabella da utilizzare per l'esportazione. Gli amministratori di database possono creare o configurare spazi di tabella partizionati. Si consiglia di selezionare uno di questi spazi di tabella (anziché quello di default) da utilizzare per le esportazioni in database.

Partiziona i dati per campo. Specifica il campo di input da utilizzare per il partizionamento.

Utilizza compressione. Se è selezionata, questa opzione utilizza la compressione nella creazione delle tabelle per l'esportazione (per esempio, l'equivalente di CREATE TABLE MYTABLE(...) COMPRESS YES in SQL).

Personalizzazione delle istruzioni CREATE TABLE

Nella parte superiore di questa finestra di dialogo è possibile aggiungere opzioni specifiche del database all'istruzione CREATE TABLE.

- ▶ Selezionare la casella di controllo Personalizza comando CREATE TABLE per attivare la finestra di testo.
- ▶ Aggiungere all'istruzione le opzioni specifiche del database desiderate. Assicurarsi di mantenere i parametri di testo <table-name> e <table-columns> perché verranno sostituiti dal nome effettivo della tabella e dalle definizioni delle colonne in IBM® SPSS® Modeler.

Impostazione dei tipi di dati SQL

Per default, SPSS Modeler consentirà al server del database di assegnare automaticamente i tipi di dati SQL. Per ignorare il tipo automatico per un campo, trovare la riga corrispondente al campo e selezionare il tipo desiderato dall'elenco a discesa nella colonna *Tipo* della tabella dello schema. Per selezionare più righe è possibile utilizzare la combinazione Maiusc-clic.

Per i tipi a cui è associato un argomento di lunghezza, di precisione o di scala (BINARY, VARBINARY, CHAR, VARCHAR, NUMERIC e NUMBER), è consigliabile specificare una lunghezza anziché consentire al server di database di assegnare una lunghezza automatica. Se lo si desidera, per esempio, è possibile garantire la sovrascrittura del tipo di archiviazione in SPSS Modeler specificando un valore sensibile per la lunghezza, come VARCHAR(25). Per ignorare l'assegnazione automatica, selezionare Specifica dall'elenco a discesa Tipo e sostituire la definizione del tipo con l'istruzione di definizione del tipo SQL desiderato.

Figura 7-4

Finestra di dialogo Specifica tipo del nodo di output Database



Il modo più semplice per modificare l'argomento per un tipo consiste nel selezionare innanzitutto il tipo maggiormente corrispondente alla definizione del tipo desiderato e quindi Specifica per modificare tale definizione. Per esempio, per impostare il tipo di dati SQL su VARCHAR(25), impostare innanzitutto il tipo su VARCHAR(length) nell'elenco a discesa Tipo, quindi selezionare Specifica e sostituire la lunghezza del testo con il valore 25.

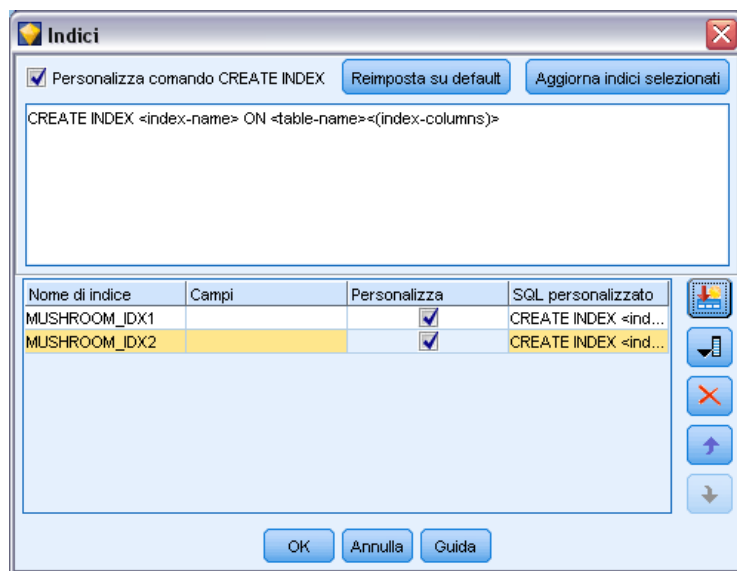
Chiavi principali

Se una o più colonne della tabella esportata devono avere un valore o una combinazione di valori univoci per ogni riga, è possibile indicarlo selezionando la casella di controllo Chiave principale per ogni campo valido. La maggior parte dei database non consente di apportare modifiche alla tabella che possano invalidare il vincolo della chiave principale e crea automaticamente un indice sulla chiave principale per favorire il rispetto di questa restrizione. Se lo si desidera, è possibile creare indici per altri campi nella finestra di dialogo Indici. [Per ulteriori informazioni, vedere l'argomento Esportazione database - Opzioni Indice a pag. 414.](#))

Esportazione database - Opzioni Indice

La finestra di dialogo Indici consente di creare indici su tabelle di database esportate da IBM® SPSS® Modeler. È possibile specificare gli insiemi di campi da includere e personalizzare il comando CREATE INDEX in funzione delle proprie esigenze.

Figura 7-5
Finestra di dialogo Indici del nodo di output Database



La finestra di dialogo è composta da due parti:

- Il campo di testo nella parte superiore visualizza un modello che può essere utilizzato per generare uno o più comandi CREATE INDEX, che per default segue il formato:
CREATE INDEX <index-name> ON <table-name>
- La tabella nella parte inferiore della finestra di dialogo consente di inserire specifiche per tutti gli indici che si desiderano creare. Per ogni indice vengono specificati il nome e i campi o le colonne da includere. Questa finestra di dialogo genera automaticamente i valori dei parametri <index-name> e <table-name> di conseguenza.

Per esempio, la stringa SQL generata per un singolo indice sui campi *empid* e *deptid* potrebbe avere il seguente aspetto:

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID)
```

È possibile aggiungere più righe per creare indici multipli. Per ogni riga viene generato un comando `CREATE INDEX` separato.

Personalizzazione del comando `CREATE INDEX`

Se lo si desidera, è possibile personalizzare il comando `CREATE INDEX` per tutti gli indici o solo per un indice specifico. Questo garantisce la flessibilità necessaria per adeguarsi ai requisiti o alle opzioni specifiche di un database e per applicare le personalizzazioni a tutti gli indici o solo a indici specifici, in funzione delle proprie esigenze.

- Selezionare Personalizza comando `CREATE INDEX` nella parte superiore della finestra di dialogo per modificare il modello utilizzato per tutti gli indici aggiunti in seguito. Si noti che le modifiche non vengono applicate automaticamente agli indici che sono già stati aggiunti alla tabella.
- Selezionare una o più righe della tabella e fare clic su **Aggiorna indici selezionati** nella parte superiore della finestra di dialogo per applicare le personalizzazioni correnti a tutte le righe selezionate.
- Selezionare la casella di controllo **Personalizza** in ogni riga per modificare il modello del comando solo per quell'indice.

Si noti che i valori dei parametri `<index-name>` e `<table-name>` sono generati automaticamente dalla finestra di dialogo a seconda delle specifiche della tabella e non possono essere modificati direttamente.

Parola chiave `BITMAP`. Se si utilizza un database di Oracle, è possibile personalizzare il modello per creare un indice bitmap anziché un indice standard, nel modo seguente:

```
CREATE BITMAP INDEX <index-name> ON <table-name>
```

Gli indici bitmap possono essere utili per indicizzare le colonne con un numero limitato di valori distinti. La stringa SQL ottenuta potrebbe avere il seguente aspetto:

```
CREATE BITMAP INDEX MYTABLE_IDX1 ON MYTABLE(COLOR)
```

Parola chiave `UNIQUE`. La maggior parte dei database supporta la parola chiave `UNIQUE` nel comando `CREATE INDEX`. Questo applica un vincolo di univocità simile a quello della chiave principale nella tabella sottostante.

```
CREATE UNIQUE INDEX <index-name> ON <table-name>
```

Si noti che, per i campi effettivamente designati come chiavi principali, questa specifica non è necessaria. La maggior parte dei database crea automaticamente un indice per tutti i campi designati come chiavi principali nel comando `CREATE TABLE`, per cui non è necessario creare esplicitamente degli indici in questi campi. [Per ulteriori informazioni, vedere l'argomento Esportazione Database - Opzioni Schema a pag. 412.](#)

Parola chiave FILLFACTOR. Alcuni parametri fisici dell'indice possono essere ulteriormente raffinati. Per esempio, SQL Server consente all'utente di raggiungere un compromesso tra le dimensioni dell'indice (dopo la creazione iniziale) e i costi di manutenzione delle future modifiche apportate alla tabella.

```
CREATE INDEX MYTABLE_IDX1 ON MYTABLE(EMPID,DEPTID) WITH FILLFACTOR=20
```

Altre osservazioni

- Se esiste già un indice con il nome specificato, la creazione dell'indice non andrà a buon fine. Gli eventuali errori vengono inizialmente considerati avvisi, consentendo la creazione degli indici successivi e la nuova segnalazione dell'errore nel registro dei messaggi dopo che il database ha tentato di creare tutti gli indici. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dei messaggi relativi alle operazioni dello stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- Per ottimizzare le prestazioni, gli indici devono essere creati dopo avere caricato i dati nella tabella. Gli indici devono contenere almeno una colonna.
- Prima di eseguire il nodo è possibile visualizzare un'anteprima della stringa SQL generata nel registro dei messaggi. [Per ulteriori informazioni, vedere l'argomento Anteprima di SQL generato in il capitolo 6 in IBM SPSS Modeler Server 14.2 Guida della performance e amministrazione.](#)
- Per le tabelle temporanee che vengono scritte nel database (quando cioè viene attivata la cache dei nodi) non sono disponibili le opzioni che consentono di indicare le chiavi principali e gli indici. Tuttavia, se necessario, il sistema può creare degli indici nella tabella temporanea, a seconda di come vengono utilizzati i dati nei nodi a valle. Per esempio, se ai dati della cache si aggiunge in seguito una colonna *DEPT*, sarebbe sensato indicizzare la tabella della cache in questa colonna. [Per ulteriori informazioni, vedere l'argomento Opzioni di copia dei dati nella cache per i nodi in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Indici e ottimizzazione delle query

In alcuni sistemi di gestione di database, dopo che una tabella è stata creata, caricata e indicizzata è necessaria un'ulteriore operazione per consentire all'ottimizzatore di utilizzare gli indici per accelerare l'esecuzione delle query nella nuova tabella. Per esempio, in Oracle, l'ottimizzatore di query basato sui costi richiede che la tabella venga prima analizzata per poterne utilizzare gli indici per l'ottimizzazione delle query. Il file delle proprietà ODBC interno per Oracle (non visibile all'utente) contiene un'opzione perché questo accada, nel modo seguente:

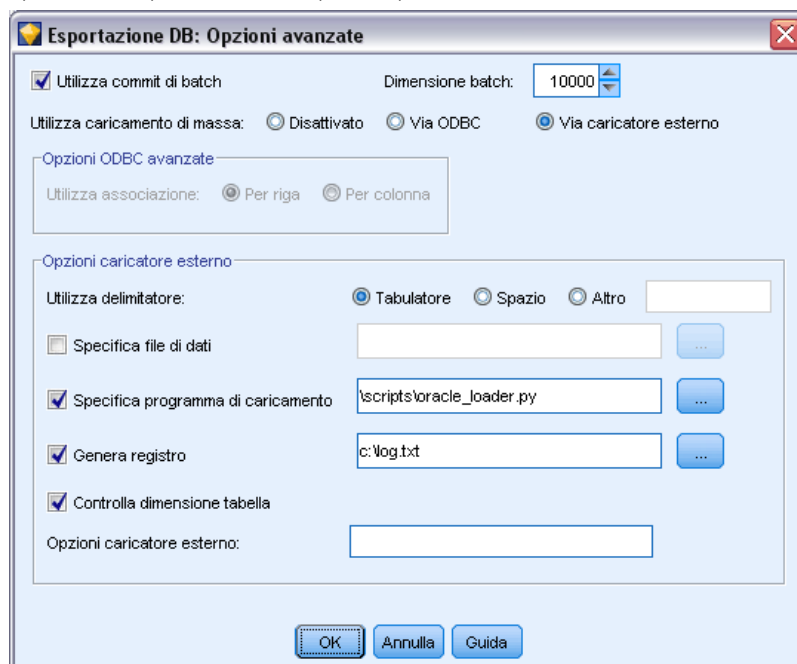
```
# Defines SQL to be executed after a table and any associated indexes
# have been created and populated
table_analysis_sql, 'ANALYZE TABLE <nome-tabella> COMPUTE STATISTICS'
```

Questa operazione viene eseguita tutte le volte che viene creata una tabella in Oracle (indipendentemente dal fatto che vengano definite delle chiavi principali o degli indici). Se necessario, il file delle proprietà ODBC per database aggiuntivi può essere personalizzato in modo analogo. Contattare l'assistenza.

Esportazione database - Opzioni avanzate

Facendo clic sul pulsante Opzioni avanzate nella finestra di dialogo Nodo di esportazione Database, viene aperta una nuova finestra di dialogo in cui è possibile specificare i dettagli tecnici per l'esportazione dei risultati in un database.

Figura 7-6
Specificazione di opzioni avanzate per l'esportazione in database



Utilizza commit di batch. Selezionare questa opzione per disattivare i commit riga per riga nel database.

Dimensione batch. Specificare il numero di record da inviare al database prima del commit nella memoria. La riduzione di questo numero garantisce una maggiore integrità dei dati, rallentando però la velocità di trasferimento. Può rivelarsi utile regolare questo numero in modo da ottenere prestazioni ottimali con il database in uso.

Opzioni InfoSphere Warehouse. Visualizzata solo se si è collegati a un database InfoSphere Warehouse (IBM DB2 9.7 o versioni successive). Non registrare gli aggiornamenti consente di evitare di registrare gli eventi quando si creano le tabelle e si inseriscono i dati.

Utilizza caricamento di massa. Selezionare un metodo per eseguire il caricamento di massa dei dati nel database direttamente da IBM® SPSS® Modeler. Può essere necessario fare qualche esperimento per determinare le opzioni di caricamento di massa più appropriate per una determinata situazione.

- **Via ODBC.** Selezionare questa opzione per utilizzare l'API ODBC in modo da eseguire inserimenti di più righe con una maggiore efficienza rispetto alla normale esportazione nel database. Nelle opzioni riportate di seguito scegliere l'associazione per riga o per colonna.
- **Via caricatore esterno.** Selezionare questa opzione per utilizzare un programma di caricamento di massa personalizzato, specifico per il database in uso. Selezionando questa opzione verranno attivate le diverse opzioni riportate di seguito.

Opzioni ODBC avanzate. Queste opzioni sono disponibili soltanto in caso di selezione di Via ODBC. Si noti che questa funzionalità potrebbe non essere supportata da tutti i driver ODBC.

- **Per riga.** Selezionare l'associazione per riga per utilizzare la chiamata a `SQLBulkOperations` per il caricamento dei dati nel database. L'associazione per riga garantisce in genere una velocità più elevata rispetto all'utilizzo di inserimenti con parametri, con cui i dati vengono inseriti record per record.
- **Per colonna.** Selezionare questa opzione per utilizzare l'associazione per colonna per il caricamento dei dati nel database. L'associazione per colonna garantisce un incremento delle prestazioni associando ogni colonna del database (in un'istruzione `INSERT` con parametri) a una matrice di valori N . Eseguendo `INSERT` una volta, nel database verranno inserite N righe. Questo metodo può determinare un notevole incremento delle prestazioni.

Opzioni caricatore esterno. Quando si specifica Via caricatore esterno, vengono visualizzate diverse opzioni che consentono di esportare l'insieme di dati in un file e di specificare ed eseguire un programma di caricamento personalizzato per il caricamento dei dati da tale file al database. SPSS Modeler può interfacciarsi con i caricatori esterni di molti dei più diffusi sistemi di database. Il software è stato corredato di numerosi script che sono disponibili insieme alla documentazione tecnica nella sottodirectory *scripts*. Si noti che, per poter utilizzare questa funzionalità, è necessario che sullo stesso computer di SPSS Modeler o di IBM® SPSS® Modeler Server sia installato Python 2.6 e che sia impostato il parametro `python_exe_path` nel file *options.cfg*.

- **Utilizza delimitatore.** Specificare il carattere di delimitazione da utilizzare nel file esportato. Selezionare Tabulatore per eseguire la delimitazione tramite tabulazione e Spazio per delimitare tramite spazi. Selezionare Altro per specificare un altro carattere, per esempio una virgola (,).
- **Specifica file di dati.** Selezionare questa opzione per immettere il percorso da utilizzare per il file di dati scritto durante il caricamento di massa. Per default, viene creato un file temporaneo nella directory `TEMP` sul server.
- **Specifica programma di caricamento.** Selezionare questa opzione per specificare un programma per il caricamento di massa. Per default, viene cercato uno script Python da eseguire per un dato database nella sottodirectory *scripts* dell'installazione di SPSS Modeler. Il software è stato corredato di numerosi script che sono disponibili insieme alla documentazione tecnica nella sottodirectory *scripts*.
- **Genera registro.** Selezionare questa opzione per generare un file di registro nella directory specificata. Il file di registro, contenente informazioni sugli errori, si rivela utile in caso di esito negativo dell'operazione di caricamento di massa.
- **Controlla dimensione tabella.** Selezionare questa opzione per eseguire il controllo della tabella e verificare così che l'incremento di dimensione della tabella corrisponda al numero di righe esportato da SPSS Modeler.
- **Opzioni caricatore esterno.** Specificare argomenti aggiuntivi per il programma di caricamento. Per gli argomenti contenenti spazi, utilizzare le virgolette doppie.

Per includere le virgolette doppie negli argomenti facoltativi, utilizzare una barra rovesciata come carattere di escape. Per esempio, l'opzione specificata come `-comment "This is a \"comment\""` include sia il flag `-comment` che il commento stesso, reso come `This is a "comment"`.

È possibile includere una singola barra rovesciata utilizzando un'altra barra rovesciata come carattere di escape. Per esempio, l'opzione specificata come `-specialdir "C:\\Test Scripts\\"` include sia il flag `-specialdir` che la directory stessa, resa come `C:\Test Scripts\`.

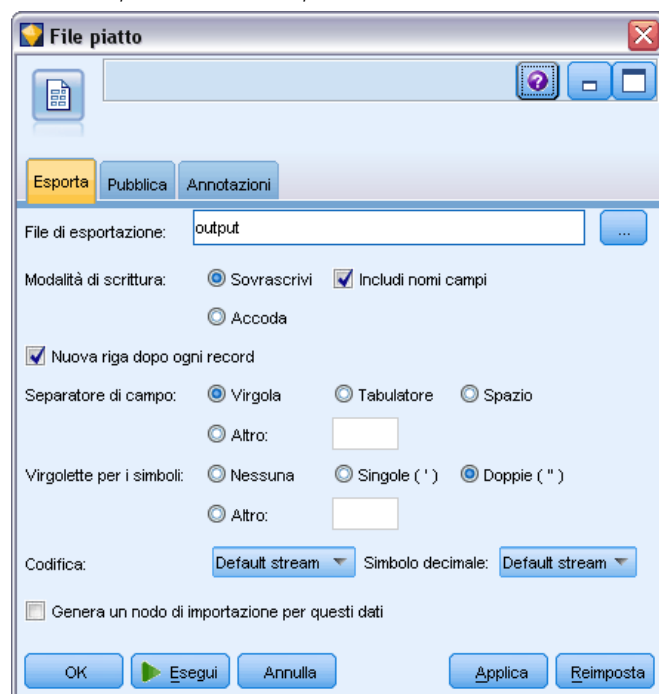
Nodo di esportazione File piatto

Il nodo di esportazione File piatto permette di scrivere i dati in un file di testo delimitato. Questa funzione è utile per esportare i dati che possono essere letti da altri software di analisi o fogli di calcolo.

Nota: non è possibile scrivere file nel precedente formato cache, poiché tale formato non viene più utilizzato per i file cache in IBM® SPSS® Modeler. I file cache di SPSS Modeler vengono attualmente salvati nel formato `.sav` di IBM® SPSS® Statistics, che può essere scritto utilizzando un nodo Esporta Statistics. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta Statistics in il capitolo 8 a pag. 449.](#)

Scheda di esportazione del nodo File piatto

Figura 7-7
Nodo File piatto, scheda Esporta



File di esportazione. Specificare il nome del file. Digitare un nome di file o fare clic sul pulsante di selezione dei file per individuare la posizione del file.

Modalità di scrittura. Se è selezionata l'opzione Sovrascrivi, tutti i dati esistenti nel file specificato verranno sovrascritti. Se è selezionata l'opzione Accoda, l'output verrà aggiunto alla fine del file esistente, preservando tutti i dati in esso contenuti.

- **Includi nomi campi.** Se questa opzione è selezionata, i nomi dei campi verranno scritti nella prima riga del file di output. Questa opzione è disponibile solo per la modalità di scrittura Sovrascrivi.

Nuova riga dopo ogni record. Se questa opzione è selezionata, ogni record verrà scritto su una nuova riga nel file di output.

Separatore di campo. Specificare il carattere da inserire tra i valori di campo nel file di testo generato. Le opzioni sono Virgola, Tabulatore, Spazio e Altro. Se si seleziona Altro, immettere il carattere o i caratteri di delimitazione desiderati nella casella di testo.

Virgolette per i simboli. Specificare il tipo di virgolette da utilizzare per i valori dei campi simbolici. Le opzioni sono Nessuna (i valori non vengono racchiusi tra virgolette), Singole ('), Doppie (") e Altro. Se si seleziona Altro, immettere i caratteri desiderati nella casella di testo.

Codifica. Specifica il metodo di codifica del testo utilizzato. È possibile scegliere tra l'impostazione di default del sistema, l'impostazione di default dello stream o UTF-8.

- L'impostazione di default del sistema è specificata nel Pannello di controllo di Windows oppure, se si esegue il programma in modalità distribuita, sul computer server. [Per ulteriori informazioni, vedere l'argomento Supporto Unicode in IBM SPSS Modeler in l'appendice B in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- L'impostazione di default dello stream è specificata nella finestra di dialogo Proprietà stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Simbolo decimale. Specificare il tipo di separatore desiderato per i decimali nei dati.

- **Default stream.** Verrà utilizzato il separatore decimale definito dall'impostazione di default dello stream corrente. In genere si tratterà del separatore decimale definito dalle impostazioni internazionali del computer.
- **Punto (.).** Come separatore di decimali verrà utilizzato il punto.
- **Virgola (,).** Come separatore di decimali verrà utilizzata la virgola.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input Testo variabile che leggerà il file di dati esportato. [Per ulteriori informazioni, vedere l'argomento Nodo Testo variabile in il capitolo 2 a pag. 20.](#)

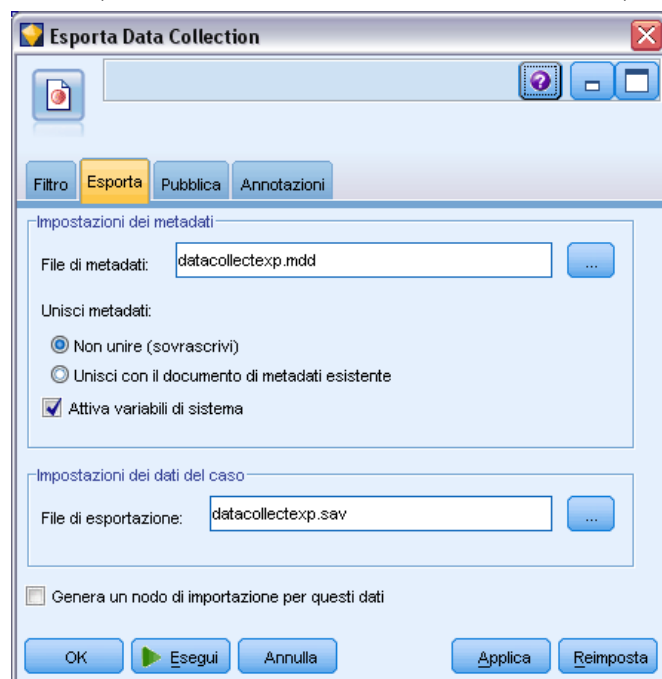
Nodo Esporta IBM SPSS Data Collection

Il nodo Esportazione di IBM® SPSS® Data Collection salva i dati nel formato utilizzato dai prodotti per ricerche di mercato Data Collection, in base al Modello dati di Data Collection. Questo formato distingue i dati del caso, cioè le risposte effettive alle domande raccolte durante un sondaggio, dai metadati che descrivono le modalità di raccolta e organizzazione dei dati del caso. I metadati sono costituiti da informazioni quali testi di domande, nomi di variabili e descrizioni, insiemi di risposte multiple, traduzioni dei vari testi, nonché dalla definizione della

struttura dei dati del caso. Per ulteriori informazioni, vedere l'argomento [Nodo Data Collection in il capitolo 2 a pag. 31](#).

Figura 7-8

Nodo Esportazione IBM SPSS Data Collection, scheda Esporta



Nota: questo nodo richiede Data Collection Data Model versione 4.0 o superiore, distribuito con il software Data Collection. Per ulteriori informazioni, vedere la pagina Web di Data Collection all'indirizzo <http://www.ibm.com/software/analytics/spss/products/data-collection/>. Fatta eccezione per l'installazione di Data Model, non è necessario eseguire altre operazioni di configurazione.

File di metadati. Specifica il nome del file di definizione del questionario (*.mdd*) in cui verranno salvati i metadati esportati. Viene creato un questionario di default in base alle informazioni sui tipi di campi. Per esempio, un campo nominale (insieme) potrebbe essere rappresentato come una domanda singola con la descrizione del campo utilizzata come testo della domanda e una casella di controllo separata per ogni valore definito.

Unisci metadati. Specifica se i metadati sovrascriveranno le versioni esistenti o saranno uniti ai metadati esistenti. Se si seleziona l'opzione di unione, ogni volta che viene eseguito lo stream viene creata una nuova versione. In questo modo è possibile tenere traccia delle versioni di un questionario via via che questo viene modificato. Ogni versione può essere considerata come un'istantanea dei metadati utilizzati per raccogliere un particolare insieme di dati del caso.

Attiva variabili di sistema. Specifica se le variabili di sistema vengono incluse nel file *.mdd* esportato. Tra queste sono comprese variabili come *Respondent.Serial*, *Respondent.Origin* e *DataCollection.StartTime*.

Impostazioni dei dati del caso. Specifica il file di dati IBM® SPSS® Statistics (.sav) in cui vengono esportati i dati del caso. Si noti che in questo caso sono valide tutte le restrizioni sui nomi delle variabili e dei valori; quindi, per esempio, può essere necessario passare alla scheda Filtro e utilizzare l'opzione "Rinomina per SPSS Statistics" del menu Filtro per correggere i caratteri non validi nei nomi dei campi.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input di Data Collection che leggerà il file di dati esportato.

Insiemi di risposte multiple. Tutti gli insiemi di risposte multiple definiti nello stream saranno automaticamente conservati quando il file viene esportato. Gli insiemi di risposte multiple si possono visualizzare e modificare da qualsiasi nodo mediante la scheda Filtro. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple in il capitolo 4 a pag. 147.](#)

Nodo di esportazione IBM Cognos BI

Il nodo di esportazione IBM Cognos BI consente di esportare dati da uno stream di IBM® SPSS® Modeler a Cognos BI, in formato UTF-8. In questo modo, Cognos BI può usufruire dei dati trasformati o dei punteggi dei dati di SPSS Modeler. Per esempio, si può utilizzare Cognos BI Report Studio per creare un report basato sui dati esportati, comprese le previsioni e i valori di confidenza. Il report si può in seguito salvare sul server Cognos BI ed essere distribuito agli utenti di Cognos BI.

Nota: è possibile esportare solo dati relazionali, non dati OLAP.

Per esportare i dati in Cognos BI è necessario specificare quanto segue:

- Connessione Cognos - la connessione al server Cognos BI
- Connessione ODBC - la connessione al server di dati Cognos utilizzato dal server Cognos BI

Le connessioni devono puntare allo stesso database e devono utilizzare lo stesso nome utente.

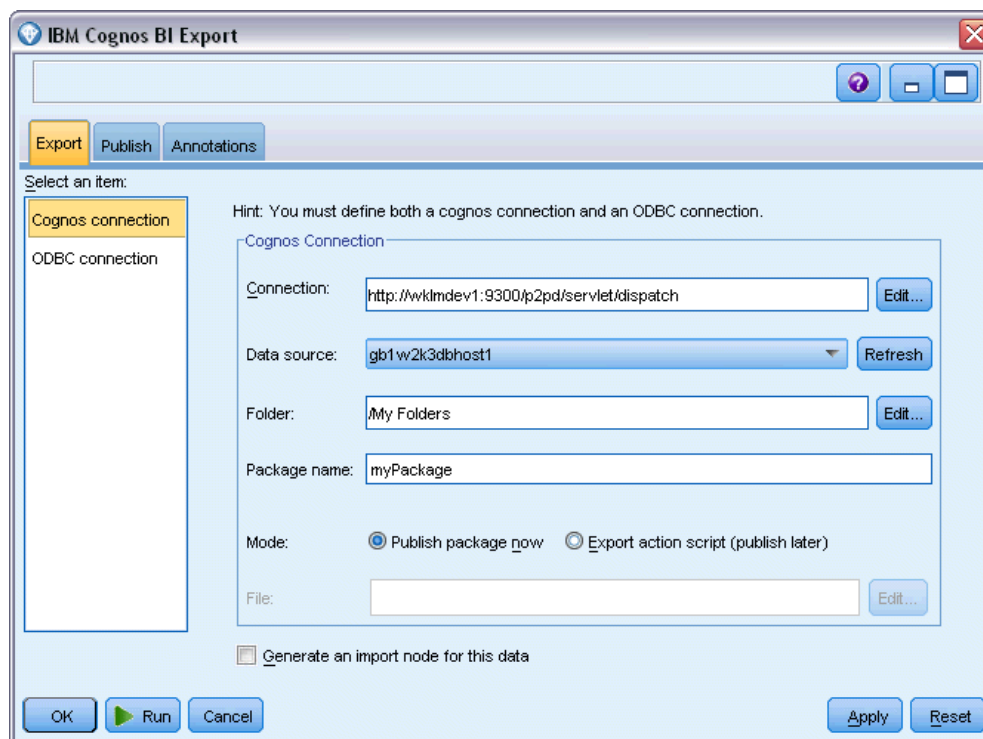
I dati dello stream vengono esportati sul server di dati, mentre i metadati del pacchetto vengono esportati sul server Cognos BI.

Come con qualsiasi altro nodo di esportazione, è possibile anche utilizzare la scheda Pubblica della finestra di dialogo del nodo per pubblicare lo stream per il deployment mediante IBM® SPSS® Modeler Solution Publisher.

Connessione Cognos

Qui è possibile specificare la connessione al server Cognos BI da utilizzare per l'esportazione. La procedura prevede l'esportazione dei metadati in un nuovo pacchetto sul server Cognos BI, mentre i dati dello stream vengono esportati sul server di dati Cognos.

Figura 7-9
Esportazione dei dati Cognos



Connessione. Fare clic sul pulsante Modifica per visualizzare una finestra di dialogo che permette di definire l'URL e gli altri dettagli del server Cognos BI su cui si desidera esportare i dati. Se si è già connessi a un server Cognos BI mediante IBM® SPSS® Modeler è possibile anche modificare i dettagli della connessione corrente. [Per ulteriori informazioni, vedere l'argomento Connessioni Cognos in il capitolo 2 a pag. 41.](#)

Sorgente dati. Il nome della sorgente dati Cognos (in genere un database) in cui saranno esportati i dati. L'elenco a discesa mostra tutte le sorgenti dati Cognos a cui è possibile accedere dalla connessione corrente. Fare clic sul pulsante Aggiorna per aggiornare l'elenco.

Cartella. Il percorso e il nome della cartella del server Cognos BI dove deve essere creato il pacchetto di esportazione.

Nome pacchetto. Il nome del pacchetto che dovrà contenere i metadati esportati all'interno della cartella specificata. Deve essere un nuovo pacchetto con un solo oggetto della query; non è possibile esportare i metadati in un pacchetto esistente.

Modalità. Specifica le modalità di esecuzione dell'esportazione:

- **Pubblica subito il pacchetto.** (impostazione di default) Esegue l'operazione di esportazione non appena si fa clic su Esegui.
- **Script di esportazione.** Crea uno script XML di esportazione eseguibile in un secondo momento (per esempio tramite Framework Manager). Digitare il percorso e il nome del file di script nel campo File, oppure specificare il nome e la posizione del file di script con il pulsante Modifica.

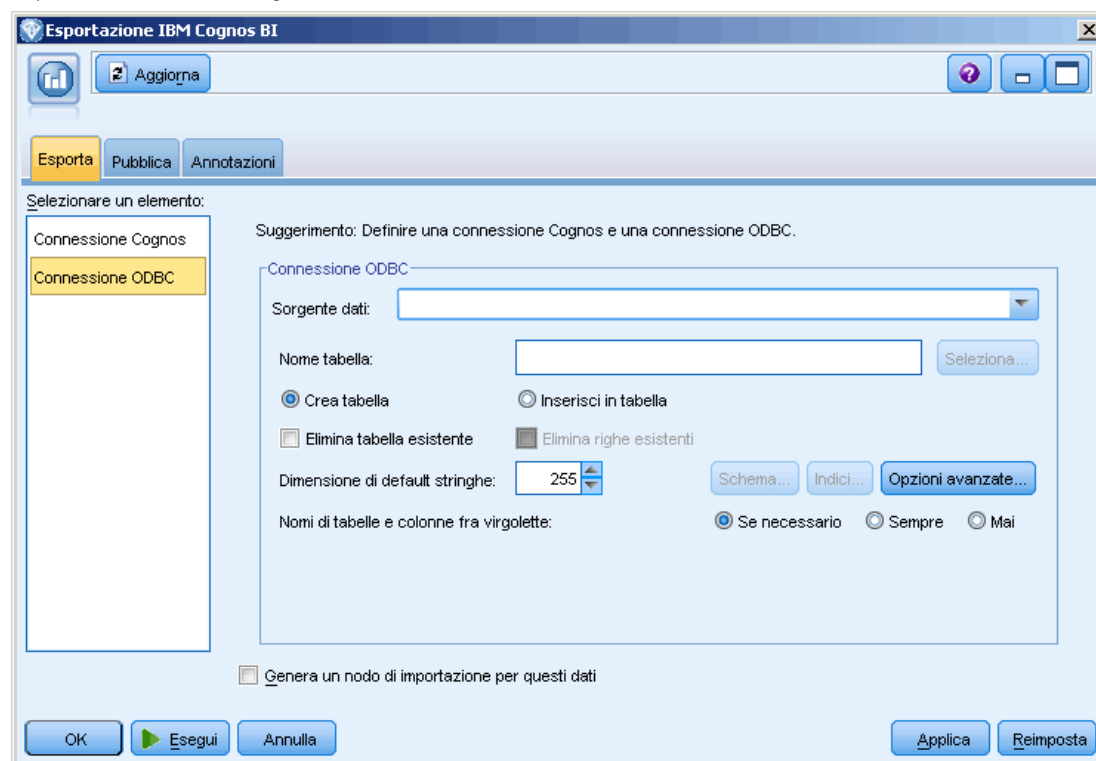
Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare un nodo di input per i dati esportati nella sorgente dati e nella tabella specificate. Quando si fa clic su Esegui, questo nodo viene aggiunto all'area di disegno dello stream.

Connessione ODBC

Qui viene specificata la connessione al server di dati Cognos (cioè al database) in cui saranno esportati i dati dello stream.

Nota: è necessario verificare che la sorgente dati specificata punti alla stessa sorgente dati specificata nel riquadro Connessioni Cognos. Controllare anche di avere utilizzato lo stesso nome utente sia per la connessione Cognos che per la connessione ODBC.

Figura 7-10
Esportazione dei dati Cognos



Sorgente dati. Mostra la sorgente dati selezionata. Immettere il nome o selezionarlo dall'elenco a discesa. Se nell'elenco non compare il database desiderato, selezionare Aggiungi nuova connessione database e individuare il database dalla finestra di dialogo Connessioni database. [Per ulteriori informazioni, vedere l'argomento Aggiunta di una connessione a un database in il capitolo 2 a pag. 16.](#)

Nome tabella. Immettere il nome della tabella alla quale si desidera inviare i dati. Se si seleziona l'opzione Inserisci in tabella, è possibile selezionare una tabella esistente del database facendo clic sul pulsante Seleziona.

Crea tabella. Selezionare questa opzione per creare una nuova tabella di database o per sovrascriverne una esistente.

Inserisci in tabella. Selezionare questa opzione per inserire i dati come nuove righe in una tabella di database esistente.

Unisci tabella. (Dove disponibile) Selezionare questa opzione per aggiornare le colonne del database selezionate con i valori dei campi dati di origine corrispondenti. La selezione di questa opzione abilita il pulsante Unione, che visualizza una finestra di dialogo da cui è possibile mappare i campi dati di origine alle colonne del database.

Elimina tabella esistente. Selezionare questa opzione per eliminare qualsiasi tabella esistente con lo stesso nome della nuova tabella che si desidera creare.

Elimina righe esistenti. Selezionare questa opzione per eliminare dalla tabella le righe esistenti prima dell'esportazione quando si inseriscono dati in una tabella.

Nota: se è selezionata una delle due opzioni sopra riportate, durante l'esecuzione del nodo verrà visualizzato un messaggio Avviso di sovrascrittura. Per eliminare gli avvisi, deselezionare Avvisa quando un nodo sovrascrive una tabella di database nella scheda Notifiche della finestra di dialogo Opzioni utente. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di notifica in il capitolo 12 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

Dimensioni di default stringhe. I campi contrassegnati come senza tipo in un nodo Tipo a monte vengono scritti nel database come campi stringa. Specificare la dimensione delle stringhe da utilizzare per i campi senza tipo.

Fare clic su Schema per visualizzare una finestra di dialogo in cui è possibile impostare tipi di dati SQL per i campi e specificare la chiave principale ai fini dell'indicizzazione del database. [Per ulteriori informazioni, vedere l'argomento Esportazione Database - Opzioni Schema a pag. 412.](#)

Fare clic su Indici per impostare le opzioni di indicizzazione della tabella esportata al fine di migliorare le prestazioni del database. [Per ulteriori informazioni, vedere l'argomento Esportazione Database - Opzioni Schema a pag. 412.](#)

Fare clic su Opzioni avanzate per specificare le opzioni di caricamento di massa e di commit nel database. [Per ulteriori informazioni, vedere l'argomento Esportazione database - Opzioni avanzate a pag. 417.](#)

Nomi di tabelle e colonne fra virgolette. Selezionare le opzioni da utilizzare al momento dell'invio di un'istruzione CREATE TABLE al database. Le tabelle e le colonne con spazi o caratteri non standard devono essere racchiuse tra virgolette.

- **Se necessario.** Selezionare questa opzione per consentire a IBM® SPSS® Modeler di determinare automaticamente l'eventuale esigenza di virgolette per ogni singolo caso.
- **Sempre.** Selezionare questa opzione per racchiudere sempre i nomi di tabella e di colonna tra virgolette.
- **Mai.** Selezionare questa opzione per disattivare l'utilizzo delle virgolette.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare un nodo di input per i dati esportati nella sorgente dati e nella tabella specificate. Quando si fa clic su Esegui, questo nodo viene aggiunto all'area di disegno dello stream.

Nodo Esporta SAS

Il nodo Esporta SAS permette di scrivere nel formato SAS i dati che devono essere letti in SAS o in un pacchetto software compatibile con SAS. È possibile eseguire l'esportazione in tre formati di file SAS: SAS per Windows/OS2, SAS per UNIX o SAS Versione 7/8.

Scheda Esporta del nodo Esporta SAS

Figura 7-11
Nodo Esporta SAS, scheda Esporta



File di esportazione. Specificare il nome del file. Digitare un nome di file o fare clic sul pulsante di selezione dei file per individuare la posizione del file.

Esporta. Specificare il formato del file di esportazione. Le opzioni sono SAS per Windows/OS2, SAS per UNIX o SAS Versione 7/8.

Esporta nomi di campo come variabile. Selezionare le opzioni per esportare i nomi e le etichette di campo da IBM® SPSS® Modeler per utilizzarli in SAS.

- **Nomi ed etichette di variabili.** Selezionare questa opzione per esportare sia i nomi che le etichette dei campi di SPSS Modeler. I nomi vengono esportati come nomi di variabili SAS, mentre le etichette vengono esportate come etichette di variabili SAS.
- **etichette.** Selezionare questa opzione per utilizzare i nomi di campi SPSS Modeler come etichette di variabile SAS. SPSS Modeler consente di includere nei nomi dei campi caratteri non validi nei nomi di variabili SAS. Per evitare di creare nomi SAS non validi, selezionare nomi.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input SAS che leggerà il file di dati esportato. [Per ulteriori informazioni, vedere l'argomento Nodo di input SAS in il capitolo 2 a pag. 42.](#)

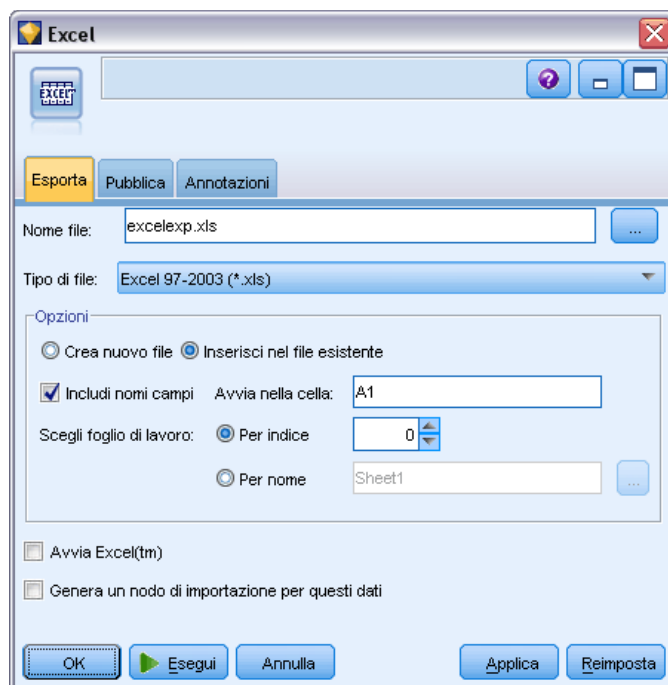
Nodo Esportazione da Excel

Il nodo Esportazione da Excel esegue l'output di dati in formato Microsoft Excel (.xls). Se lo si desidera, è possibile scegliere di avviare Excel automaticamente e aprire il file esportato quando si esegue il nodo.

Scheda di esportazione del nodo Excel

Figura 7-12

Nodo Esportazione da Excel, scheda Esporta



Nome file. Digitare un nome di file o fare clic sul pulsante di selezione dei file per individuare la posizione del file. Il nome file di default è *excelexp.xls*.

Tipo di file. Selezionare il tipo di file Excel che si sta esportando.

Crea nuovo file. Crea un nuovo file di Excel.

Inserisci nel file esistente. Il contenuto viene sostituito a partire dalla cella specificata nel campo Avvia nella cella. Le altre celle del foglio di calcolo mantengono il contenuto originale.

Includi nomi campi. Specifica se i nomi dei campi devono essere inclusi nella prima riga del foglio di lavoro.

Avvia nella cella. La posizione della cella utilizzata per il primo record di esportazione (o il nome del primo campo se è selezionato Includi nomi campi). I dati vengono compilati a destra e verso il basso rispetto a questa cella iniziale.

Scegli foglio di lavoro. Specifica il foglio di lavoro nel quale si desidera esportare i dati. È possibile individuare il foglio di lavoro in base all'indice o al nome:

- **Tramite indice.** In caso di creazione di un nuovo file, specificare un valore tra 0 e 9 per identificare il foglio di lavoro nel quale si desidera effettuare l'esportazione, iniziando con 0 per il primo foglio di lavoro, 1 per il secondo e così via. È possibile utilizzare valori maggiori o uguali a 10 solo se in questa posizione esiste già un foglio di lavoro.
- **Per nome.** In caso di creazione di un nuovo file, specificare il nome utilizzato per il foglio di lavoro. In caso di inserimento all'interno di un file esistente, se il foglio di lavoro esiste i dati vengono inseriti al suo interno, altrimenti viene creato un nuovo foglio di lavoro con questo nome.

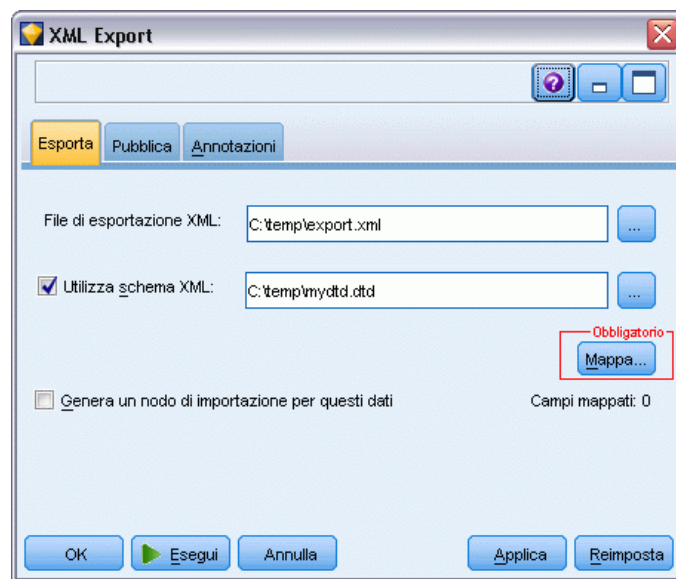
Avvia Excel. Indica se Excel deve essere avviato automaticamente con il file esportato quando viene eseguito il nodo. Si noti che, durante l'esecuzione in modalità distribuita con IBM® SPSS® Modeler Server, l'output viene salvato nel file system del server ed Excel viene avviato nell'installazione Client con una copia del file esportato.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input Excel che leggerà il file di dati esportato. [Per ulteriori informazioni, vedere l'argomento Nodo di input Excel in il capitolo 2 a pag. 44.](#)

Nodo Esporta XML

Il nodo Esporta XML consente di esportare i dati in formato XML, utilizzando la codifica UTF-8. Se lo si desidera, è possibile creare un nodo di input XML per leggere nuovamente i dati esportati nello stream.

Figura 7-13
Esportazione di dati XML



File di esportazione XML. Il percorso e il nome file completi del file XML nel quale si desidera esportare i dati.

Utilizza schema XML. Selezionare questa casella di controllo per utilizzare uno schema o DTD per controllare la struttura dei dati esportati. In questo modo viene attivato il pulsante Mappa descritto di seguito.

Se non si utilizza uno schema o un DTD, per i dati esportati viene utilizzata la seguente struttura di default:

```
<records>
  <record>
    <fieldname1>value</fieldname1>
    <fieldname2>value</fieldname2>
    :
    <fieldnameN>value</fieldnameN>
  </record>
  <record>
  :
  :
  </record>
  :
  :
</records>
```

Gli spazi in un nome campo vengono sostituiti dai caratteri di sottolineatura, per esempio “My Field” diventa <My_Field>.

Mappa. Se è stata selezionata l’opzione di utilizzo di uno schema XML, il pulsante visualizza una finestra di dialogo nella quale è possibile specificare la parte della struttura XML da utilizzare per iniziare ogni nuovo record. [Per ulteriori informazioni, vedere l'argomento Opzioni dei record di mappatura XML a pag. 430.](#)

Campi mappati. Indica il numero di campi che sono stati mappati.

Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input XML che rileggerà il file di dati esportato nello stream. [Per ulteriori informazioni, vedere l'argomento XML Source Node in il capitolo 2 a pag. 45.](#)

Scrittura dei dati XML

Quando viene specificato un elemento XML, il valore del campo viene posizionato all’interno del tag dell’elemento:

```
<element>value</element>
```

Quando viene mappato un attributo, il valore del campo viene posizionato come valore dell’attributo:

```
<element attribute="value">
```

Se un campo viene mappato a un elemento sopra all’elemento <records>, il campo viene scritto solo una volta e sarà una costante per tutti i record. Il valore di questo elemento proviene dal primo record.

Se è necessario scrivere un valore nullo, specificare un contenuto vuoto. Per gli elementi, si tratta di:

```
<element></element>
```

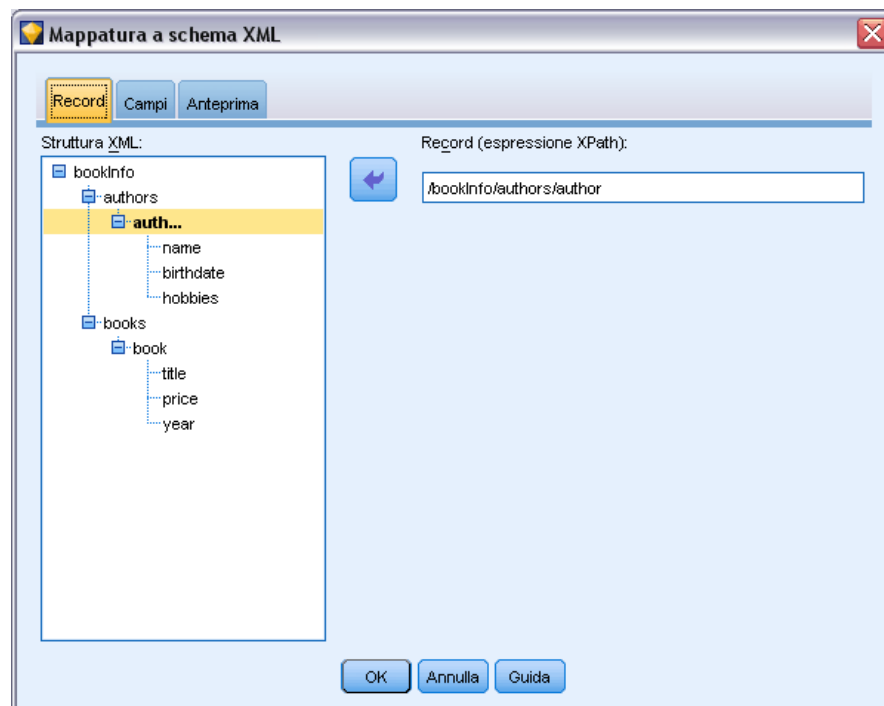
Per gli attributi, si tratta di:

```
<element attribute="">
```

Opzioni dei record di mappatura XML

La scheda Record consente di specificare quale parte della struttura XML debba essere utilizzata per iniziare ogni nuovo record. Per eseguire correttamente la mappatura a uno schema, è necessario specificare il delimitatore dei record.

Figura 7-14
Record di mappatura XML



Struttura XML. Un albero gerarchico che mostra la struttura dello schema XML specificato nella schermata precedente.

Record (espressione XPath). Per impostare il delimitatore dei record, selezionare un elemento nella struttura XML e fare clic sul pulsante con la freccia destra. Ogni volta che si incontra questo elemento nei dati di input, nel file di output viene creato un nuovo record.

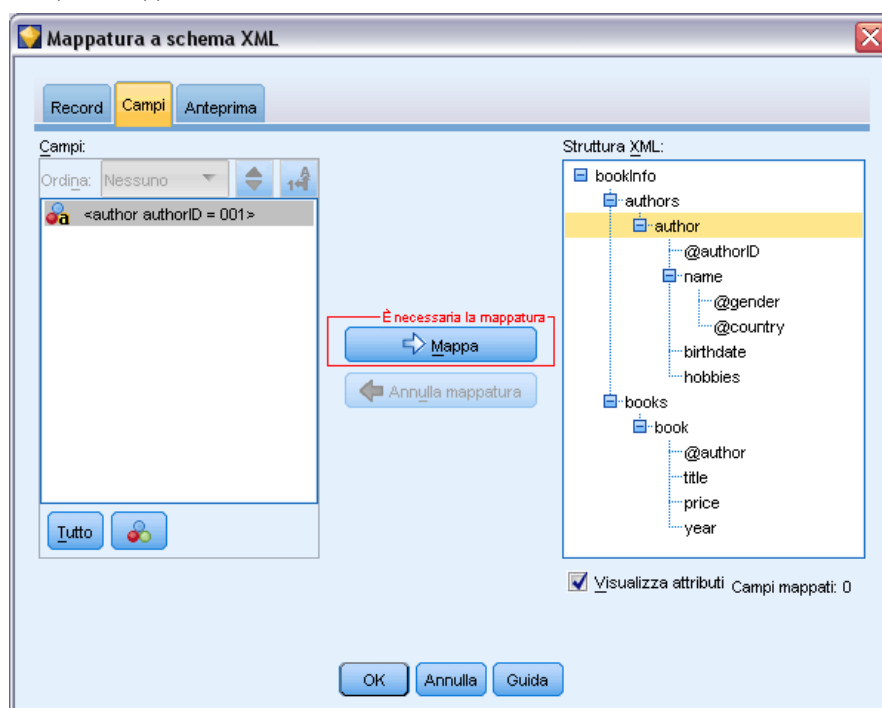
Nota: se si seleziona l'elemento radice nella struttura XML, è possibile scrivere soltanto un record; tutti gli altri record verranno saltati.

Opzioni dei campi di mappatura XML

La scheda Campi viene utilizzata per mappare i campi nell'insieme di dati a elementi o attributi nella struttura XML quando viene utilizzato un file schema.

I nomi dei campi che corrispondono al nome di un elemento o di un attributo vengono mappati automaticamente a condizione che il nome dell'elemento o dell'attributo sia univoco. Pertanto, se sono presenti sia un elemento che un attributo con il nome field1, non viene eseguita alcuna mappatura automatica. Se nella struttura è presente un solo elemento di nome field1, viene mappato automaticamente un campo con questo nome nello stream.

Figura 7-15
Campi di mappatura XML



Campi. L'elenco dei campi nel modello. Selezionare uno o più campi come parte di input della mappatura. È possibile utilizzare i pulsanti in fondo all'elenco per selezionare tutti i campi, oppure tutti i campi con un particolare livello di misurazione.

Struttura XML. Selezionare un elemento nella struttura XML come obiettivo della mappatura. Per generare la mappatura, fare clic su Mappa. La mappatura viene quindi visualizzata. Il numero di campi mappati in questo modo viene visualizzato sotto all'elenco.

Per eliminare una mappatura, selezionare l'elemento nell'elenco della struttura XML e fare clic su Annulla mappatura.

Visualizza attributi. Visualizza o nasconde gli attributi, se presenti, degli elementi XML nella struttura XML.

Anteprima mappatura XML

Nella scheda Anteprima, fare clic su **Aggiorna** per visualizzare un'anteprima dell'XML che verrà scritto.

Se la mappatura non è corretta, ritornare alla scheda **Record** o **Campi** per correggere gli errori e fare nuovamente clic su **Aggiorna** per visualizzare il risultato.

Nodi IBM SPSS Statistics

Nodi IBM SPSS Statistics - Panoramica

Come complemento a IBM® SPSS® Modeler e alle relative funzionalità di data mining, IBM® SPSS® Statistics consente di eseguire analisi statistiche più approfondite e di gestire i dati.

Se si dispone di una copia compatibile con licenza di SPSS Statistics installata, è possibile connettersi a questo programma da SPSS Modeler ed eseguire analisi e manipolazioni di dati complessi a più fasi che non sarebbero altrimenti supportate da SPSS Modeler. L'utente più esperto ha inoltre la possibilità di modificare ulteriormente le analisi tramite una sintassi di comandi. Vedere le Note di versione per informazioni sulla compatibilità delle versioni.

Se disponibili, i nodi SPSS Statistics vengono mostrati nella parte dedicata della palette dei nodi.

Nota: si consiglia di istanziare i dati in un nodo Tipo prima di utilizzare i nodi Trasformazioni, Modello o Output di SPSS Statistics. Questo è necessario anche quando si utilizza il comando di sintassi AUTORECODE.

La palette SPSS Statistics contiene i seguenti nodi:



Il nodo File Statistics legge i dati dal formato di file *.sav* utilizzato da SPSS Statistics, nonché da file della cache salvati in SPSS Modeler, che utilizzano lo stesso formato. [Per ulteriori informazioni, vedere l'argomento Nodo File Statistics a pag. 434.](#)



Il nodo Trasformazioni Statistics esegue una selezione di comandi di sintassi SPSS Statistics rispetto alle sorgenti dati in SPSS Modeler. Questo nodo richiede una copia di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento Nodo Trasformazioni Statistics a pag. 435.](#)



Il nodo Modello Statistics consente di analizzare e operare con i dati eseguendo le procedure SPSS Statistics che generano PMML. Questo nodo richiede una copia di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento Nodo Modello Statistics a pag. 440.](#)



Il nodo Output Statistics consente di chiamare una procedura SPSS Statistics per analizzare i dati di SPSS Modeler. È disponibile una vasta gamma di procedure analitiche di SPSS Statistics. Questo nodo richiede una copia di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento Nodo Output Statistics a pag. 444.](#)

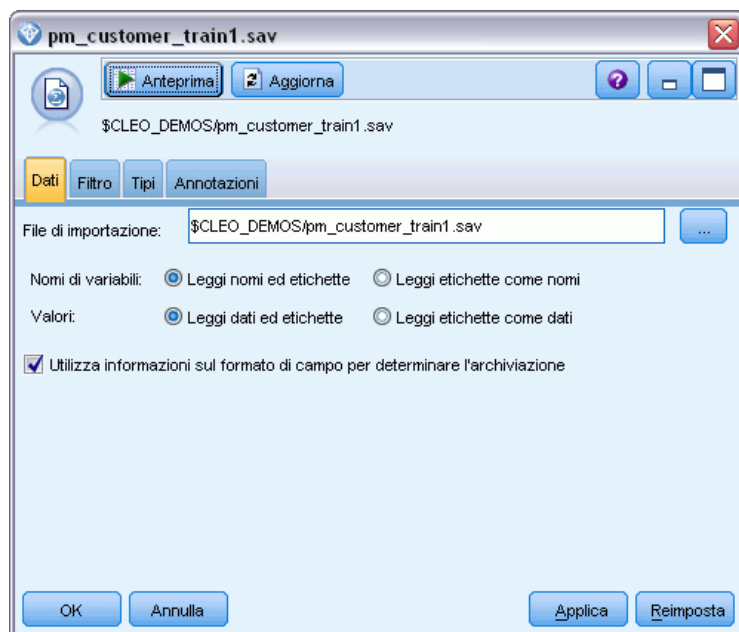


Il nodo Esporta Statistics restituisce i dati nel formato *.sav* di SPSS Statistics. I file *.sav* possono essere letti da SPSS Statistics Base e altri prodotti. Questo formato viene inoltre utilizzato per i file cache di SPSS Modeler. [Per ulteriori informazioni, vedere l'argomento Nodo Esporta Statistics a pag. 449.](#)

Nodo File Statistics

È possibile utilizzare il nodo File Statistics per leggere dati direttamente da un file IBM® SPSS® Statistics salvato (.sav). Questo formato viene attualmente utilizzato per sostituire il file cache di versioni precedenti di IBM® SPSS® Modeler. Se si desidera importare un file cache salvato, è necessario utilizzare il nodo File SPSS Statistics.

Figura 8-1
Importazione di un file .sav



File di importazione. Specificare il nome del file. È possibile immettere un nome di file o fare clic sul pulsante con i puntini di sospensione (...) per selezionare un file. Il percorso di un file viene visualizzato dopo la selezione del file.

Nomi delle variabili. Selezionare un metodo per la gestione dei nomi e delle etichette di variabile durante l'importazione da un file SPSS Statistics.sav. I metadati che si sceglie di includere vengono mantenuti per la sessione di SPSS Modeler e possono essere esportati di nuovo per essere utilizzati in SPSS Statistics.

- **Leggi nomi ed etichette.** Selezionare questa opzione per leggere sia i nomi che le etichette di variabile in SPSS Modeler. Questa opzione è selezionata per default e i nomi di variabile vengono visualizzati nel nodo Tipo. È possibile visualizzare le etichette nei grafici, nei browser dei modelli e in altri tipi di output, in base alle opzioni specificate nella finestra di dialogo delle proprietà dello stream. La visualizzazione delle etichette nell'output è disattivata per default. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)
- **Leggi etichette come nomi.** Selezionare questa opzione per leggere dal file .sav di SPSS Statistics le etichette di variabili descrittive anziché i nomi di campo brevi e utilizzare queste etichette come nomi di variabili in SPSS Modeler.

Valori. Selezionare un metodo per la gestione dei valori e delle etichette durante l'importazione da un file *.sav* di SPSS Statistics. I metadati che si sceglie di includere vengono mantenuti per la sessione di SPSS Modeler e possono essere esportati di nuovo per essere utilizzati in SPSS Statistics.

■ **Leggi dati ed etichette.** Selezionare questa opzione per leggere sia i valori effettivi che le etichette dei valori in SPSS Modeler. Questa opzione è selezionata per default e i valori vengono visualizzati nel nodo Tipo. Le etichette dei valori possono essere visualizzate nel generatore di espressioni, nei grafici, nei browser dei modelli e in altri tipi di output, in base alle opzioni specificate nella finestra di dialogo delle proprietà dello stream. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni per gli stream in il capitolo 5 in Manuale dell'utente di IBM SPSS Modeler 14.2.](#)

■ **Leggi etichette come dati.** Selezionare questa opzione se per i valori si desidera utilizzare le etichette del file *sav* anziché i codici numerici o simbolici. Per esempio, se si seleziona questa opzione per dati con un campo relativo al genere i cui valori 1 e 2 rappresentano rispettivamente *maschio* e *femmina*, il campo verrà convertito in una stringa e come valori effettivi verranno importati i valori *maschio* e *femmina*.

È importante considerare i valori mancanti nei dati SPSS Statistics prima di selezionare questa opzione. Per esempio, se un campo numerico utilizza etichette solo per i valori mancanti (0 = *Nessuna risposta*, -99 = *Sconosciuto*), la selezione di questa opzione determinerà l'importazione delle sole etichette *Nessuna risposta* e *Sconosciuto* e la conversione del campo in una stringa. In questi casi, è consigliabile importare i valori stessi e impostare i valori mancanti in un nodo Tipo.

Utilizza informazioni sul formato di campo per determinare l'archiviazione. Se questa casella è selezionata, i valori dei campi formattati come numeri interi nel file *.sav* (cioè i campi specificati come *F_n.0* nella Visualizzazione variabili di SPSS Statistics) vengono importati utilizzando il tipo di archiviazione numero intero. Tutti i valori degli altri campi tranne le stringhe vengono importati come numeri reali.

Se la casella non è selezionata (impostazione di default), tutti i valori dei campi tranne le stringhe vengono importati come numeri reali, indipendentemente dal fatto che siano o meno formattati come numeri interi nel file *.sav*.

Insiemi di risposte multiple. Tutti gli insiemi di risposte multiple definiti nel file SPSS Statistics vengono mantenuti automaticamente quando il file viene importato. Gli insiemi di risposte multiple si possono visualizzare e modificare da qualsiasi nodo mediante la scheda Filtro. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple in il capitolo 4 a pag. 147.](#)

Nodo Trasformazioni Statistics

Il nodo Trasformazioni Statistics consente di completare le trasformazioni dei dati utilizzando la sintassi di comando IBM® SPSS® Statistics. Questa funzione consente di completare una serie di trasformazioni non supportate da IBM® SPSS® Modeler e di automatizzare trasformazioni complesse e articolate in più fasi, compresa la creazione di alcuni campi da un unico nodo. È simile al nodo Output Statistics, ad eccezione del fatto che i dati vengono restituiti a SPSS

Modeler per un'ulteriore analisi, mentre nel nodo Output i dati vengono restituiti sotto forma degli oggetti di output richiesti, per esempio grafici o tabelle.

Per utilizzare questo nodo è necessario avere installata sul computer una versione compatibile di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento IBM SPSS Statistics Applicazioni di supporto in il capitolo 6 a pag. 405.](#) Vedere le Note di versione per informazioni sulla compatibilità.

Se necessario, è possibile visualizzare la scheda Filtro per filtrare o rinominare i campi in modo che risultino conformi agli standard di denominazione SPSS Statistics. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics a pag. 451.](#)

Syntax Reference. Per informazioni dettagliate su specifiche procedure SPSS Statistics, consultare la guida *SPSS Statistics Command Syntax Reference*, inclusa nella copia del software di SPSS Statistics. Per visualizzare la guida dalla scheda Sintassi, scegliere l'opzione Editor di sintassi e fare clic sul pulsante Avvia Guida sintassi SPSS Statistics.

Nota: non tutta la sintassi SPSS Statistics è supportata da questo nodo. [Per ulteriori informazioni, vedere l'argomento Sintassi consentita a pag. 438.](#)

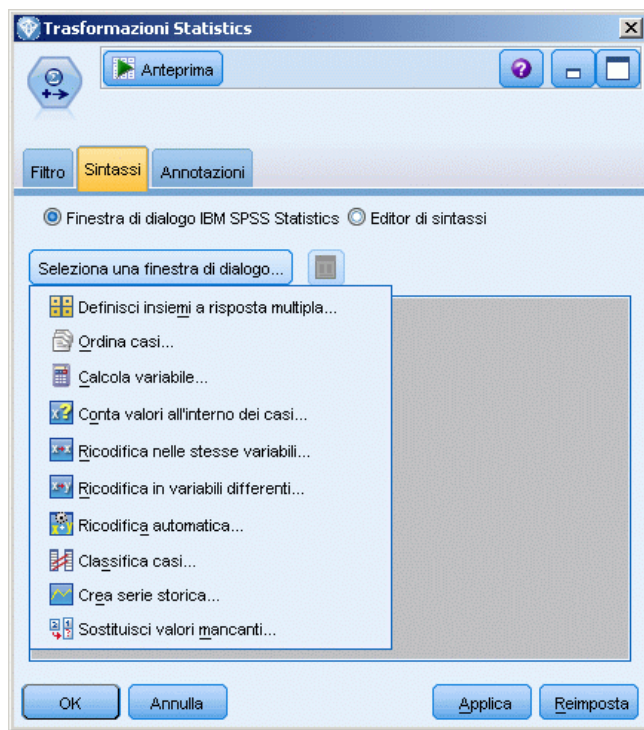
Scheda Sintassi del nodo Trasformazioni Statistics

Opzione finestra di dialogo IBM SPSS Statistics

Se non si conosce la sintassi IBM® SPSS® Statistics, il modo più semplice per creare la sintassi in IBM® SPSS® Modeler consiste nell'eseguire prima il comando in SPSS Statistics, copiare la sintassi nel nodo Trasformazioni Statistics in SPSS Modeler ed eseguire infine lo stream.

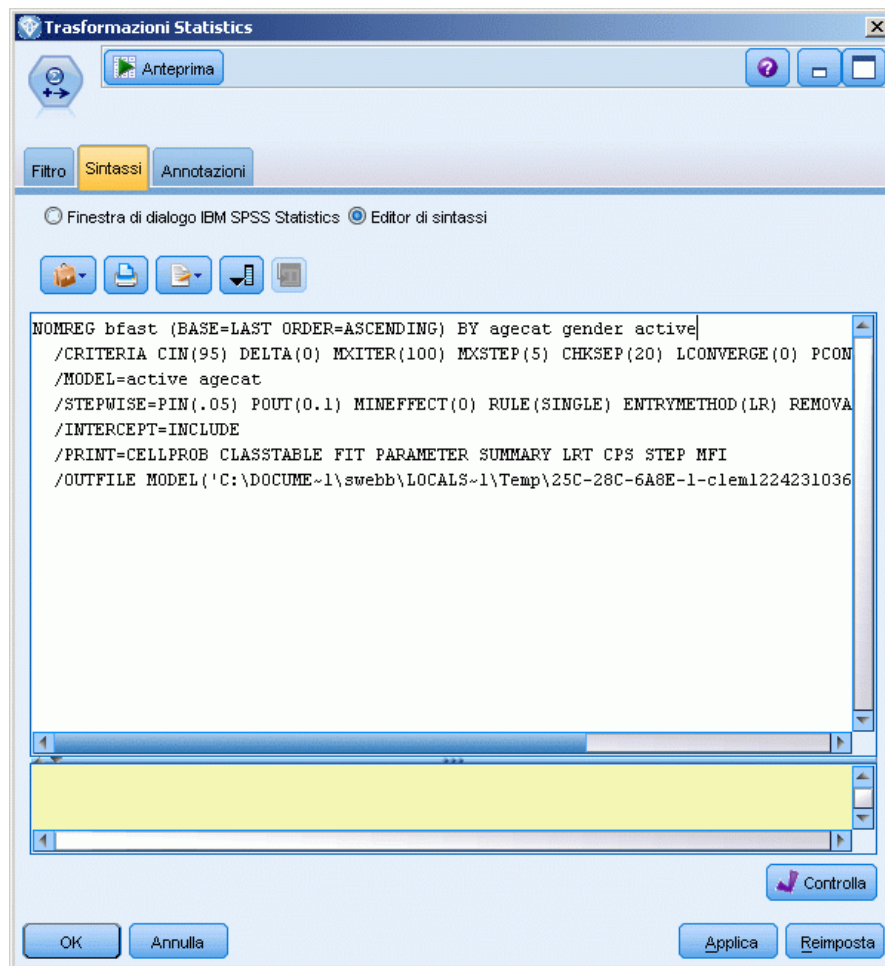
A tale scopo, fare clic sul pulsante Seleziona una finestra di dialogo... e scegliere la procedura SPSS Statistics appropriata dall'elenco. [Per ulteriori informazioni, vedere l'argomento Scheda Modello del nodo Modello Statistics a pag. 441.](#)

Figura 8-2
Nodo Trasformazioni Statistics, selezione dalla finestra di dialogo



Opzione Editor di sintassi IBM SPSS Statistics

Figura 8-3
Editor di sintassi del nodo Trasformazioni Statistics



Controllo. Dopo che sono stati immessi i comandi di sintassi nella parte superiore della finestra di dialogo, utilizzare questo pulsante per convalidare i dati immessi. Gli eventuali errori di sintassi vengono evidenziati nella parte inferiore della finestra.

Per evitare il prolungarsi del processo di verifica, quando viene convalidata la sintassi viene eseguita la verifica in base a un campione rappresentativo dei dati e non in base all'insieme di dati completo.

Sintassi consentita

Se si dispone di una grande quantità di sintassi legacy da IBM® SPSS® Statistics oppure si ha già una certa dimestichezza con le funzioni di preparazione dei dati di SPSS Statistics, è possibile utilizzare il nodo Trasformazioni Statistics per eseguire molte delle trasformazioni esistenti. Come linea guida, il nodo consente di trasformare i dati in modi prevedibili, per esempio, mediante

l'esecuzione di comandi in ciclo oppure mediante la modifica, l'aggiunta, l'ordinamento, il filtraggio o la selezione dei dati.

Segue un elenco di alcuni dei comandi che possono essere eseguiti:

■ **Calcolo di numeri casuali in base a una distribuzione binomiale:**

```
COMPUTE newvar = RV.BINOM(10000,0.1)
```

■ **Ricodifica di una variabile in una nuova variabile:**

```
RECODE Age (Lowest thru 30=1) (30 thru 50=2) (50 thru Highest=3) INTO AgeRecoded
```

■ **Sostituzione dei valori mancanti:**

```
RMV Age_1=SMEAN(Age)
```

La tabella seguente elenca la sintassi SPSS Statistics che è supportata dal nodo Trasformazioni Statistics:

Nome comando

```
ADD VALUE LABELS
APPLY DICTIONARY
AUTORECODE
BREAK
CD
CLEAR MODEL PROGRAMS
CLEAR TIME PROGRAM
CLEAR TRANSFORMATIONS
COMPUTE
COUNT
CREATE
DATE
DEFINE-!ENDDEFINE
DELETE VARIABLES
DO IF
DO REPEAT
ELSE
ELSE IF
END CASE
END FILE
END IF
END INPUT PROGRAM
END LOOP
END REPEAT
EXECUTE
FILE HANDLE
FILE LABEL
FILE TYPE-END FILE TYPE
FILTER
FORMATS
IF
INCLUDE
INPUT PROGRAM-END INPUT PROGRAM
```

Nome comando

INSERT
LEAVE
LOOP-END LOOP
MATRIX-END MATRIX
MISSING VALUES
N OF CASES
NUMERIC
PERMISSIONS
PRESERVE
RANK
RECODE
RENAME VARIABLES
RESTORE
RMV
SAMPLE
SELECT IF
SET
SORT CASES
SORT CASES
STRING
SUBTITLE
TEMPORARY
TITLE
UPDATE
V2C
VALIDATEDATA
VALUE LABELS
VARIABLE ATTRIBUTE
VARSTOCASES
VECTOR

Nodo Modello Statistics

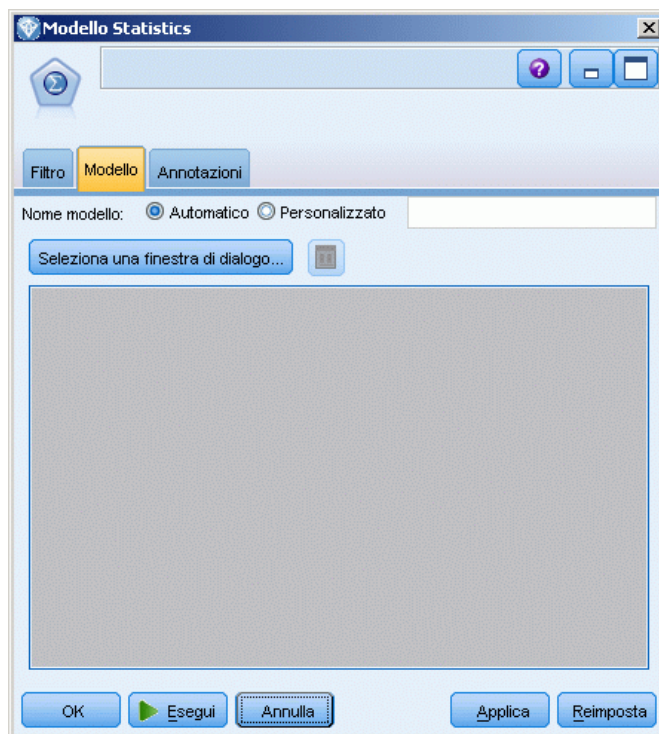
Il nodo Modello Statistics consente di analizzare e operare con i dati eseguendo le procedure IBM® SPSS® Statistics che generano PMML. L'insieme dei modelli che si creano possono essere utilizzati normalmente all'interno degli stream IBM® SPSS® Modeler per il calcolo del punteggio e altre operazioni.

Per utilizzare questo nodo è necessario avere installata sul computer una versione compatibile di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento IBM SPSS Statistics Applicazioni di supporto in il capitolo 6 a pag. 405.](#) Vedere le Note di versione per informazioni sulla compatibilità.

Le procedure analitiche SPSS Statistics disponibili dipendono dal tipo di licenza che possiede l'utente.

Scheda Modello del nodo Modello Statistics

Figura 8-4
Scheda Modello del nodo Modello Statistics

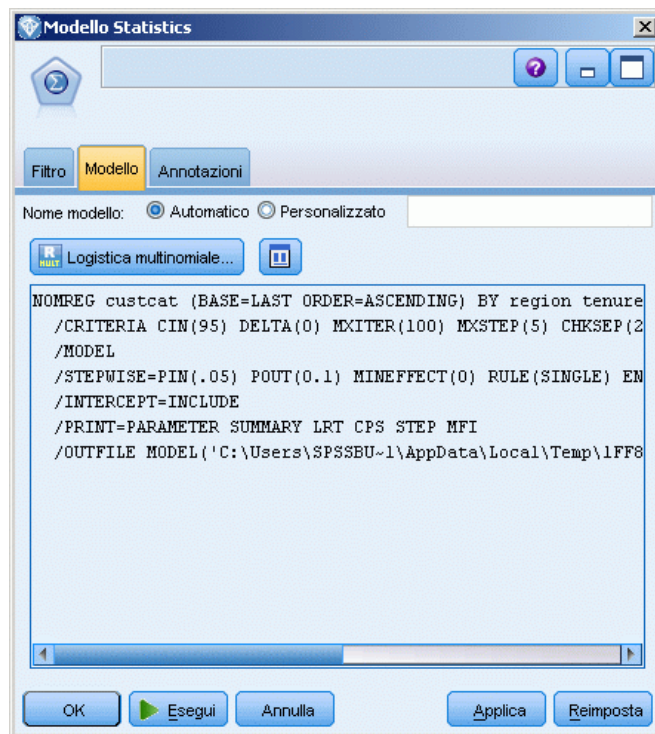


Nome modello. È possibile generare il nome del modello automaticamente in base al campo ID o obiettivo (oppure il tipo di modello nei casi non sia specificato tale campo) oppure indicare un nome personalizzato.

Seleziona una finestra di dialogo. Fare clic per visualizzare un elenco delle procedure IBM® SPSS® Statistics disponibili che è possibile selezionare ed eseguire. L'elenco include soltanto le procedure che generano PMML e per le quali si dispone di licenza e non include procedure scritte dall'utente.

- ▶ Fare clic sulla procedura richiesta; viene visualizzata la finestra di dialogo SPSS Statistics relativa.
- ▶ Nella finestra di dialogo SPSS Statistics, inserire i dettagli della procedura.
- ▶ Fare clic su OK per tornare al nodo Modello Statistics; la sintassi di SPSS Statistics viene visualizzata nella scheda Modello.

Figura 8-5
Sintassi visualizzata nella scheda Modello

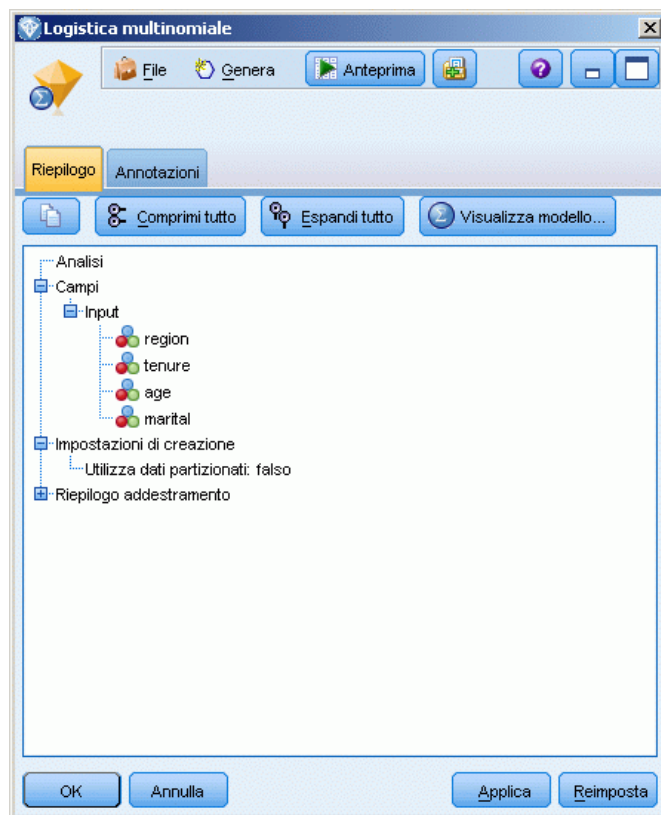


- Per tornare alla finestra di dialogo SPSS Statistics in qualsiasi momento, per esempio per modificare la propria query, fare clic sul pulsante di visualizzazione della finestra di dialogo SPSS Statistics alla destra del pulsante di selezione della procedura.

Riepilogo dell'insieme di modelli del nodo Modello Statistics

Quando si esegue il nodo Modello Statistics, questo esegue la procedura IBM® SPSS® Statistics associata e crea un insieme di modelli che è possibile utilizzare negli stream IBM® SPSS® Modeler per il calcolo del punteggio.

Figura 8-6
Insieme di modelli Statistics, scheda Riepilogo

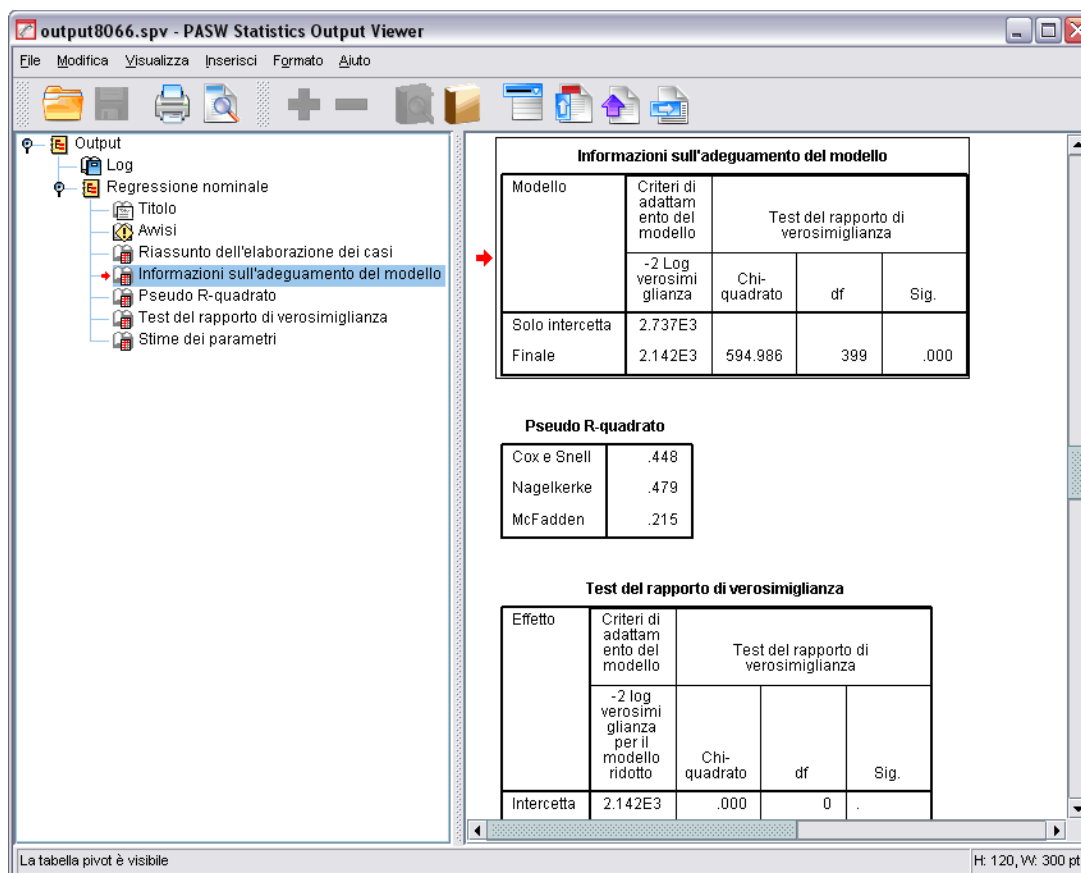


La scheda Riepilogo dell'insieme di modelli visualizza informazioni sui campi, le impostazioni di creazione e l'elaborazione della stima del modello. I risultati vengono rappresentati in una visualizzazione struttura che può essere espansa o compressa facendo clic su elementi specifici. [Per ulteriori informazioni, vedere l'argomento Scheda Riepilogo di un insieme di modelli/Informazioni in il capitolo 3 in IBM SPSS Modeler 14.2 Nodi Modelli.](#)

Il pulsante Visualizza modello visualizza i risultati in una forma modificata di SPSS Statistics Output Viewer. Per ulteriori informazioni sul visualizzatore, vedere la documentazione relativa a SPSS Statistics.

Dal menu File sono disponibili le normali opzioni di esportazione e stampa. [Per ulteriori informazioni, vedere l'argomento Visualizzazione dell'output in il capitolo 6 a pag. 348.](#)

Figura 8-7
Insieme di modelli Statistics, scheda Avanzate



Nodo Output Statistics

Il nodo Output Statistics consente di chiamare una procedura IBM® SPSS® Statistics per analizzare i dati di IBM® SPSS® Modeler. I risultati possono essere visualizzati in una finestra del browser oppure salvati nel formato del file di output SPSS Statistics. Da SPSS Modeler è possibile accedere a una vasta gamma di procedure analitiche di SPSS Statistics.

Per utilizzare questo nodo è necessario avere installata sul computer una versione compatibile di SPSS Statistics con regolare licenza. [Per ulteriori informazioni, vedere l'argomento IBM SPSS Statistics Applicazioni di supporto in il capitolo 6 a pag. 405.](#) Vedere le Note di versione per informazioni sulla compatibilità.

Se necessario, è possibile visualizzare la scheda Filtro per filtrare o rinominare i campi in modo che risultino conformi agli standard di denominazione SPSS Statistics. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics a pag. 451.](#)

Syntax Reference. Per informazioni dettagliate su specifiche procedure SPSS Statistics, consultare la guida *SPSS Statistics Command Syntax Reference*, inclusa nella copia del software di SPSS Statistics. Per visualizzare la guida dalla scheda Sintassi, scegliere l'opzione Editor di sintassi e fare clic sul pulsante Avvia Guida sintassi SPSS Statistics.

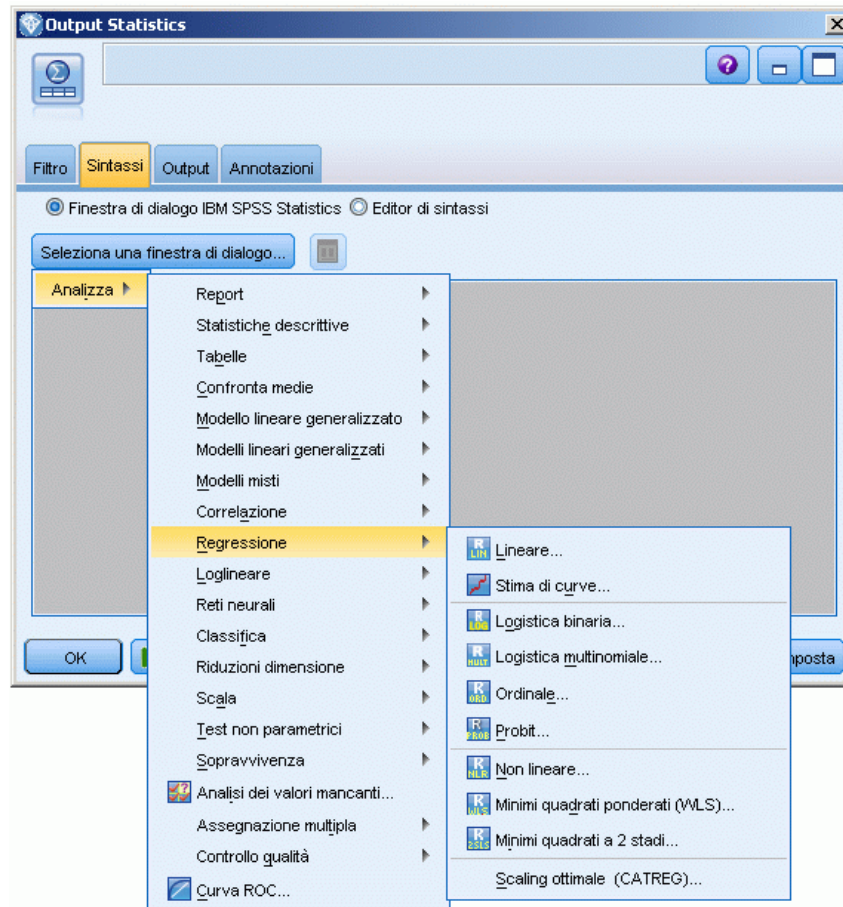
Scheda Sintassi del nodo Output Statistics

Utilizzare questa scheda per creare la sintassi per la procedura IBM® SPSS® Statistics che si desidera utilizzare per analizzare i dati. La sintassi è composta da due parti: un'**istruzione** e le **opzioni** associate a tale istruzione. L'istruzione specifica l'analisi o l'operazione da eseguire e i campi da utilizzare. Le opzioni specificano tutto il resto, comprese le statistiche da visualizzare, i campi derivati da salvare e così via.

Opzione finestra di dialogo IBM SPSS Statistics

Se non si conosce la sintassi di SPSS Statistics, il modo più semplice per creare la sintassi in IBM® SPSS® Modeler consiste nell'eseguire prima il comando in SPSS Statistics, copiare la sintassi nel nodo Output SPSS Statistics in SPSS Modeler ed eseguire infine lo stream.

Figura 8-8
 Nodo Output Statistics, selezione dalla finestra di dialogo



Per creare la sintassi:

- ▶ Fare clic sul pulsante Seleziona una finestra di dialogo.
- ▶ Scegliere una delle opzioni:
 - **Analizza.** Visualizza il contenuto del menu Analizza di SPSS Statistics; scegliere la procedura che si desidera utilizzare.
 - **Altro.** Se disponibile, visualizza le finestre di dialogo create mediante il generatore di finestre di dialogo personalizzate in SPSS Statistics e le altre finestre di SPSS Statistics non visualizzate nel menu Analizza e per le quali si dispone di una licenza. In assenza di finestre di dialogo, l'opzione non viene visualizzata.

Nota: le finestre di dialogo Preparazione automatica dati non vengono visualizzate.

Se si dispone di una finestra di dialogo personalizzata di SPSS Statistics che crea nuovi campi, tali campi non possono essere utilizzati in SPSS Modeler perché il nodo Output Statistics è un nodo terminale.

- ▶ Il resto del processo è simile a quello utilizzato per il nodo Modello Statistics. [Per ulteriori informazioni, vedere l'argomento Scheda Modello del nodo Modello Statistics a pag. 441.](#)

Opzione Editor di sintassi

Per salvare la sintassi creata per una procedura utilizzata frequentemente, procedere come descritto di seguito:

- ▶ Fare clic sul pulsante Comandi relativi ai file (il primo sulla barra degli strumenti).
- ▶ Scegliere Salva o Salva con nome dal menu.
- ▶ Salvare il file come file *.sps*.

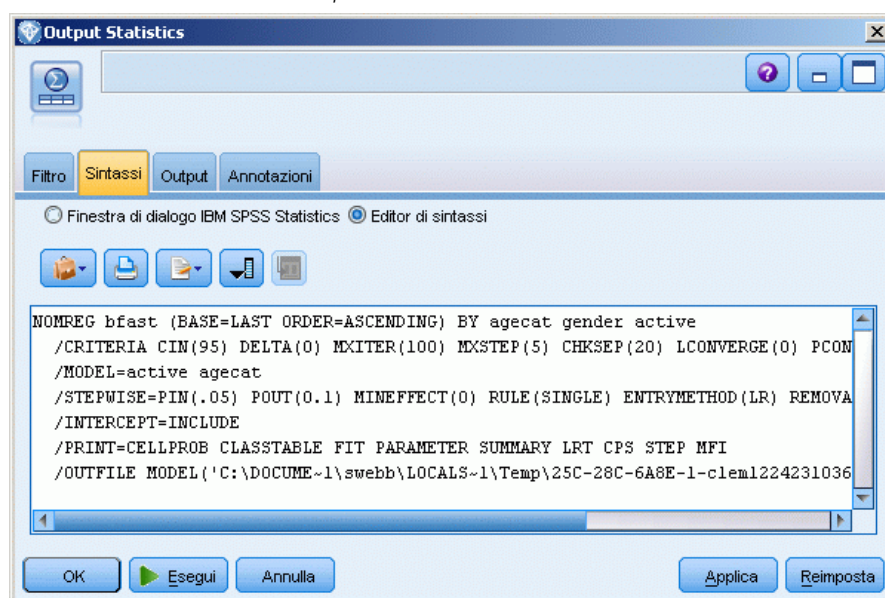
Per utilizzare i file di sintassi creati in precedenza dell'Editor di sintassi, sostituendo il contenuto corrente, se presente, procedere come descritto di seguito:

- ▶ Fare clic sul pulsante Comandi relativi ai file (il primo sulla barra degli strumenti).
- ▶ Scegliere Apri dal menu.
- ▶ Selezionare un file *.sps* e incollarne il contenuto nella scheda Sintassi del nodo Output.

Per inserire la sintassi salvata in precedenza senza sostituire il contenuto corrente, procedere come descritto di seguito:

- ▶ Fare clic sul pulsante Comandi relativi ai file (il primo sulla barra degli strumenti).
- ▶ Scegliere Inserisci dal menu
- ▶ Selezionare un file *.sps* e incollarne il contenuto nel punto del nodo Output specificato dal cursore.

Figura 8-9
Editor di sintassi nel nodo Output Statistics

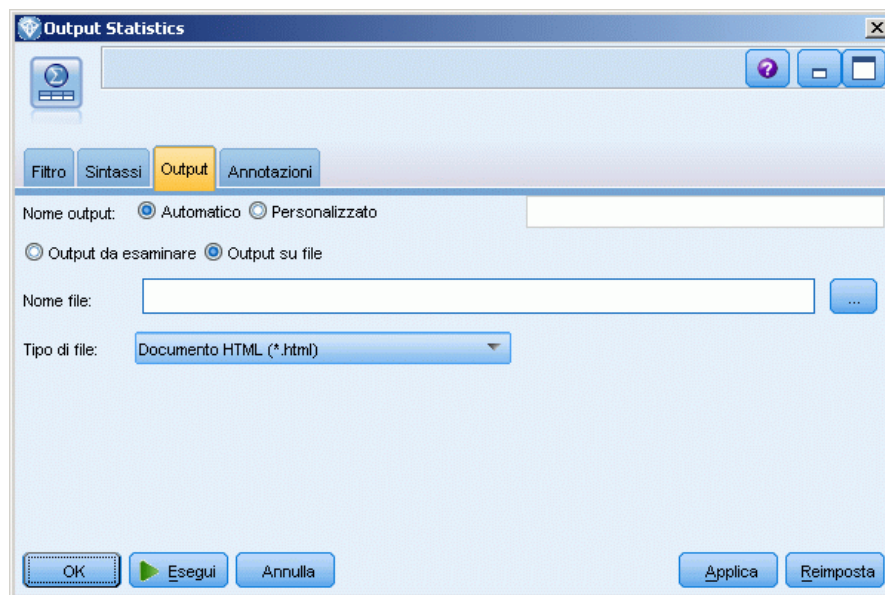


Quando si fa clic su Esegui, i risultati vengono visualizzati nel Visualizzatore output SPSS Statistics. Per ulteriori informazioni sul visualizzatore, vedere la documentazione relativa a SPSS Statistics.

Scheda Output del nodo Output Statistics

La scheda Output consente di specificare il formato e la posizione dell'output. Si può scegliere di visualizzare i risultati sullo schermo o di inviarli a uno dei tipi di file disponibili.

Figura 8-10
Scheda Output del nodo Output Statistics



Nome output. Specifica il nome dell'output generato durante l'esecuzione del nodo. Automatico sceglie un nome in base al nodo che genera l'output. Se necessario, è possibile selezionare Personalizzato per specificare un nome diverso.

Output su schermo (impostazione di default). Crea un nuovo oggetto di output da visualizzare in linea. L'oggetto di output verrà visualizzato nella scheda Output della finestra dei manager dopo l'esecuzione del nodo di output.

Output su file. Salva l'output in un file quando si esegue il nodo. Se si sceglie questa opzione, immettere un nome di file nel campo Nome file (oppure passare a una directory e specificare un nome di file utilizzando il pulsante di selezione dei file) e selezionare un tipo di file.

Tipo di file. Scegliere il tipo di file al quale si desidera inviare l'output.

- **Documento HTML (*.html).** Scrive l'output in formato HTML.

- **SPSS Statistics File Viewer (*.spv)**. Scrive l'output in un formato che può essere letto da IBM® SPSS® Statistics Output Viewer.
- **SPSS Statistics File Web Reports (*.spw)**. Scrive l'output in formato SPSS Statistics Web Reports che può essere pubblicato in un repository IBM SPSS Collaboration and Deployment Services e successivamente visualizzato in un browser Web. [Per ulteriori informazioni, vedere l'argomento Pubblica sul Web in il capitolo 6 a pag. 348.](#)

Nodo Esporta Statistics

Il nodo Esporta Statistics consente di esportare dati nel formato IBM® SPSS® Statistics *.sav*. SPSS Statistics I file *.sav* possono essere letti da SPSS Statistics Base e altri moduli. Questo formato viene inoltre utilizzato per i file cache di IBM® SPSS® Modeler.

La mappatura dei nomi di campo SPSS Modeler su nomi di variabili SPSS Statistics può a volte generare errori perché i nomi di variabili SPSS Statistics sono limitati a 64 caratteri e non possono includere alcuni caratteri, per esempio spazi, il simbolo del dollaro (\$) e i trattini (-). Esistono due modi per risolvere il problema:

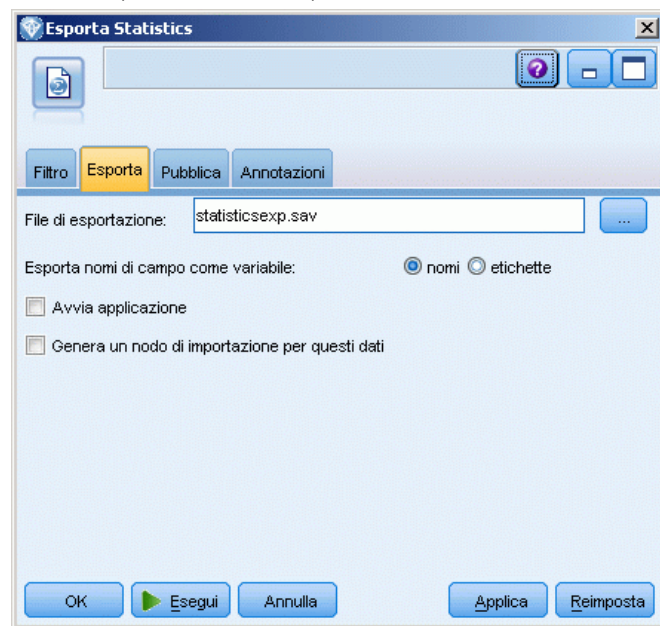
- È possibile rinominare i campi in modo da soddisfare i requisiti per i nomi di variabili SPSS Statistics facendo clic sulla scheda Filtro. [Per ulteriori informazioni, vedere l'argomento Ridenominazione o filtro dei campi per IBM SPSS Statistics a pag. 451.](#)
- Scegliere di esportare sia i nomi di campo che le etichette da SPSS Modeler.

Nota: SPSS Modeler scrive i file *.sav* in formato Unicode UTF-8. SPSS Statistics supporta i file in formato Unicode UTF-8 solo dalla versione 16.0 in avanti. Per prevenire il rischio di corruzione dei dati, i file *.sav* salvati con codifica Unicode non devono essere utilizzati nelle versioni SPSS Statistics precedenti alla 16.0. Per ulteriori informazioni, vedere l'Aiuto di SPSS Statistics.

Insiemi di risposte multiple. Tutti gli insiemi di risposte multiple definiti nello stream saranno automaticamente conservati quando il file viene esportato. Gli insiemi di risposte multiple si possono visualizzare e modificare da qualsiasi nodo mediante la scheda Filtro. [Per ulteriori informazioni, vedere l'argomento Modifica di insiemi di risposte multiple in il capitolo 4 a pag. 147.](#)

Scheda Esporta del nodo Esporta Statistics

Figura 8-11
Scheda Esporta del nodo Esporta Statistics



File di esportazione. Specificare il nome del file. Digitare un nome di file o fare clic sul pulsante di selezione dei file per individuare la posizione del file.

Esporta nomi di campo come variabile. Specificare un metodo per la gestione dei nomi e delle etichette di variabile durante l'esportazione da IBM® SPSS® Modeler in un file *.sav* di IBM® SPSS® Statistics.

- **Nomi ed etichette di variabili.** Selezionare questa opzione per esportare sia i nomi che le etichette dei campi di SPSS Modeler. I nomi vengono esportati come nomi di variabili SPSS Statistics, mentre le etichette vengono esportate come etichette di variabili SPSS Statistics.
- **etichette.** Selezionare questa opzione per utilizzare i nomi di campi SPSS Modeler come etichette di variabile in SPSS Statistics. SPSS Modeler consente di includere nei nomi dei campi caratteri non validi nei nomi di variabili SPSS Statistics. Per evitare di creare nomi SPSS Statistics non validi, selezionare etichette oppure utilizzare la scheda Filtro per modificare i nomi di campo.

Avvia applicazione. Se installato nel computer, SPSS Statistics può essere richiamato direttamente nel file di dati salvato tramite questa opzione. Le opzioni per il lancio dell'applicazione devono essere specificate nella finestra di dialogo Applicazioni di supporto. [Per ulteriori informazioni, vedere l'argomento IBM SPSS Statistics Applicazioni di supporto in il capitolo 6 a pag. 405.](#) Per creare semplicemente un file *.sav* di SPSS Statistics senza aprire un programma esterno, deselezionare questa opzione.

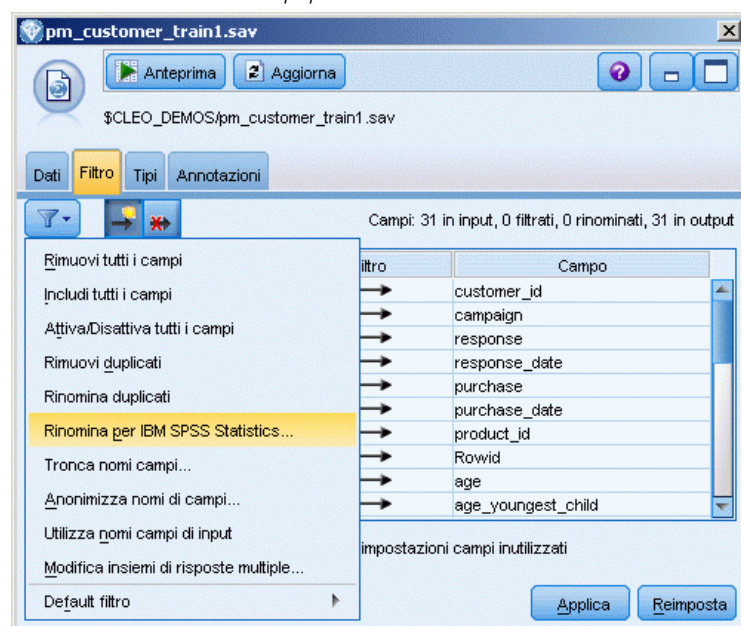
Genera un nodo di importazione per questi dati. Selezionare questa opzione per generare automaticamente un nodo di input File Statistics che leggerà il file di dati esportato. [Per ulteriori informazioni, vedere l'argomento Nodo File Statistics a pag. 434.](#)

Ridenominazione o filtro dei campi per IBM SPSS Statistics

Prima di esportare o eseguire il deployment dei dati da IBM® SPSS® Modeler ad applicazioni esterne quali IBM® SPSS® Statistics, potrebbe essere necessario rinominare o correggere i nomi dei campi. Le finestre di dialogo Trasformazioni Statistics, Output Statistics ed Esporta Statistics contengono una scheda Filtro in grado di semplificare questo processo.

Una descrizione di base della funzione della scheda Filtro viene presentata in un'altra sezione. [Per ulteriori informazioni, vedere l'argomento Impostazione delle opzioni di filtro in il capitolo 4 a pag. 144.](#) In questo argomento vengono forniti suggerimenti per la lettura dei dati in SPSS Statistics..

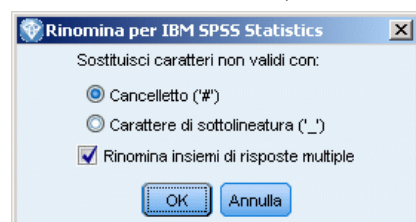
Figura 8-12
Ridenominazione dei campi per IBM SPSS Statistics nella scheda Filtro del nodo File Statistics



Per modificare i nomi dei campi in modo che corrispondano alle convenzioni di denominazione di SPSS Statistics:

- ▶ Nella scheda Filtro, fare clic sul pulsante della barra degli strumenti Comandi relativi al filtro (il primo sulla barra degli strumenti).
- ▶ Selezionare Rinomina per SPSS Statistics.

Figura 8-13
Ridenominazione dei campi



- ▶ Nella finestra di dialogo Rinomina per SPSS Statistics, è possibile scegliere di sostituire i caratteri non validi nei nomi file con un carattere di cancelletto (#) o sottolineatura (_).

Rinomina insiemi di risposte multiple. Selezionare questa opzione se si desidera modificare i nomi di insiemi di risposte multiple, che possono essere importati in SPSS Modeler utilizzando un nodo di input File Statistics. Vengono utilizzati per registrare i dati che possono avere più di un valore per ogni caso, ad esempio nelle risposte ai sondaggi.

Supernodi

Panoramica sui Supernodi

Uno dei motivi per cui l'interfaccia di programmazione visuale di IBM® SPSS® Modeler risulta semplice da utilizzare è che ogni nodo include una funzione chiaramente definita. Tuttavia, per l'elaborazione complessa può essere necessaria una lunga sequenza di nodi ed è pertanto possibile che lo spazio nell'area di disegno venga occupato e che diventi quindi difficile seguire i diagrammi dello stream. Per evitare la riduzione dello spazio in stream lunghi e complessi sono disponibili due metodi:

- È possibile suddividere una sequenza di elaborazione in più stream collegati tra di loro. Per esempio, il primo stream crea un file di dati utilizzato dal secondo stream come input. Il secondo crea un file di dati utilizzato dal terzo come input e così via. Per gestire questi stream collegati è possibile salvarli in un **progetto**. Un progetto consente di organizzare più stream con i relativi output. Tuttavia, un file di progetto contiene solo un riferimento agli oggetti inclusi e non elimina il problema della gestione di più file di stream.
- Un'alternativa più semplice per la gestione di elaborazioni di stream complesse consiste nel creare un **Supernodo**.

I Supernodi raggruppano più nodi in un unico nodo contenente sezioni di uno stream di dati. Questo metodo assicura numerosi vantaggi al data miner:

- Gli stream sono più ordinati e gestibili.
- I nodi possono essere combinati in un Supernodo specifico per un settore aziendale.
- I Supernodi possono essere esportati in librerie in modo da essere riutilizzati in più progetti di data mining.

Tipi di Supernodi

Nello stream di dati i Supernodi sono rappresentati da un'icona a forma di stella. L'icona è contraddistinta da ombreggiature diverse che indicano la direzione del flusso dello stream rispetto al Supernodo.

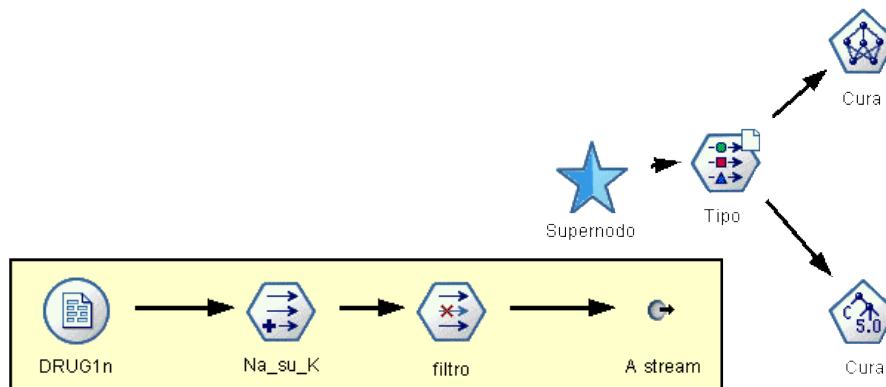
Sono disponibili tre tipi di Supernodi:

- Supernodi di input
- Supernodi di elaborazione
- Supernodo finale

Supernodi di input

I Supernodi di input contengono una sorgente dati come i normali nodi di input e possono essere utilizzati nelle stesse posizioni consentite ai normali nodi di input. L'ombreggiatura nella metà sinistra di un Supernodo di input indica che tale Supernodo è "chiuso" a sinistra e che la direzione del flusso dei dati deve essere *in entrata* nel Supernodo.

Figura 9-1
Supernodo di input con la visualizzazione Zoom avanti sopra lo stream

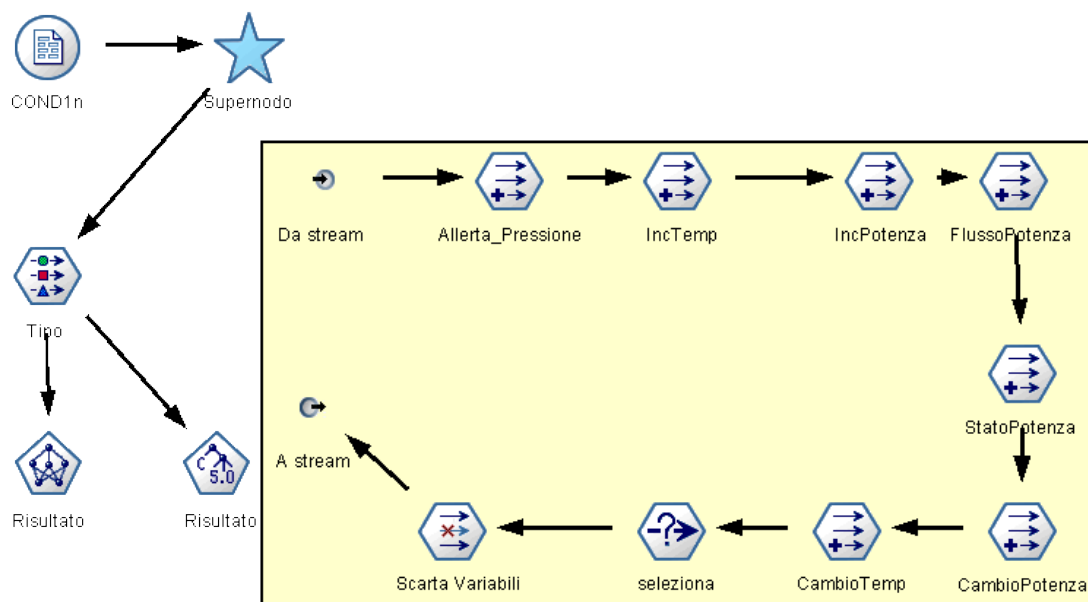


I Supernodi di input includono un solo punto di connessione a destra, che indica il flusso di dati dal Supernodo verso lo stream.

Supernodi di elaborazione

I Supernodi di elaborazione contengono solo nodi di elaborazione e non sono contrassegnati da ombreggiature in quanto il flusso dei dati è consentito sia *verso* sia *da* questo tipo di Supernodo.

Figura 9-2
Supernodo di elaborazione con la visualizzazione Zoom avanti sopra lo stream



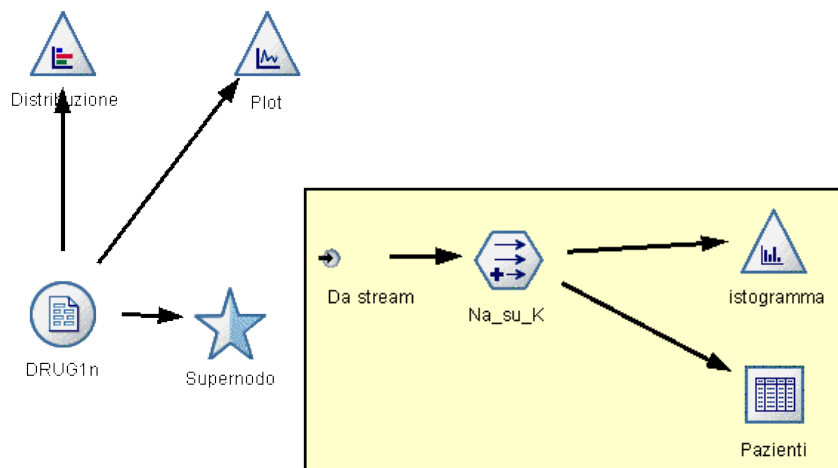
I Supernodi di elaborazione includono punti di connessione sia a destra sia a sinistra, che indicano che i dati entrano nel Supernodo e quindi escono per tornare allo stream. Sebbene i Supernodi possano contenere frammenti di stream aggiuntivi e ulteriori stream, i punti di connessione devono seguire lo stesso percorso che connette il punto *Da stream* al punto *A stream*.

Nota: I Supernodi di elaborazione sono a volte detti anche *Supernodi di manipolazione*.

Supernodo finale

I Supernodi terminali contengono uno o più nodi terminali (plot, tabelle e così via) e possono essere utilizzati nello stesso modo in cui vengono utilizzati i nodi terminali. L'ombreggiatura nella metà destra del Supernodo terminale indica che il Supernodo è "chiuso" a destra e che l'unica direzione consentita per il flusso dei dati è *in entrata* nel Supernodo terminale.

Figura 9-3
Supernodo terminale con la visualizzazione Zoom avanti sopra lo stream



I Supernodi terminali contengono un solo punto di connessione a sinistra, che indica che il flusso dei dati entra nel Supernodo dallo stream e viene interrotto all'interno del Supernodo.

I Supernodi terminali possono inoltre contenere script utilizzati per specificare l'ordine di esecuzione per tutti i nodi terminali all'interno del Supernodo. [Per ulteriori informazioni, vedere l'argomento Supernodi e script a pag. 471.](#)

Creazione di Supernodi

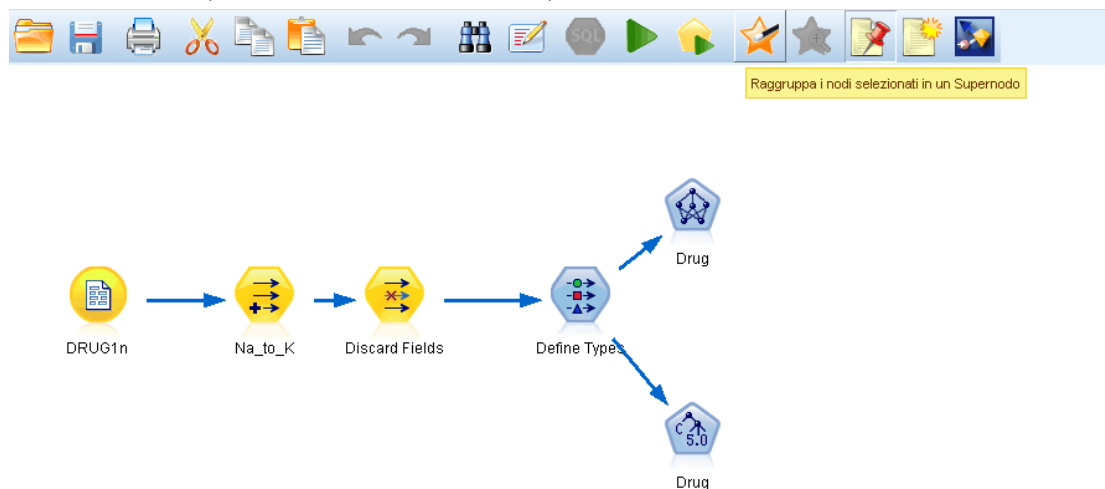
La creazione di un Supernodo consente di “contrarre” lo stream di dati tramite l'incapsulamento di più nodi in un unico nodo. Dopo aver creato o caricato uno stream nell'area di disegno, è possibile creare un Supernodo in molti modi.

Selezione multipla

La procedura più semplice per la creazione di un Supernodo consiste nel selezionare tutti i nodi che si desidera incapsulare:

- ▶ Utilizzare il mouse per selezionare più nodi nell'area di disegno dello stream. Per selezionare uno stream o una sezione di uno stream è inoltre possibile fare clic sugli elementi desiderati tenendo premuto Maiusc. *Nota:* i nodi selezionati devono appartenere a uno stream continuo o biforcuto. Non è possibile selezionare nodi che non siano adiacenti o connessi in qualche modo.
- ▶ Tramite una di queste procedure incapsulare quindi i nodi selezionati:
 - Fare clic sull'icona del Supernodo (in forma di stella) nella barra degli strumenti.
 - Fare clic con il pulsante destro del mouse sul Supernodo e dal menu di scelta rapida scegliere: Crea Supernodo > Da selezione
 - Dal menu Supernodo scegliere: Crea Supernodo > Da selezione

Figura 9-4
Creazione di un Supernodo tramite selezione multipla



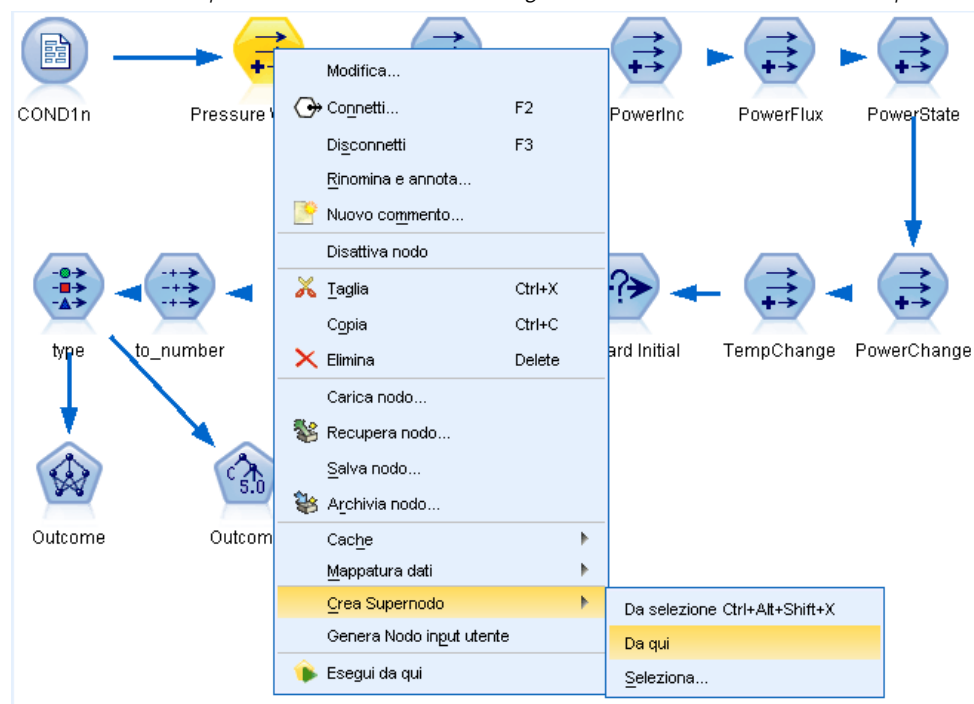
Queste tre opzioni consentono di incapsulare i nodi in un Supernodo la cui icona è ombreggiata in modo da indicare il tipo di Supernodo (di origine, di elaborazione o terminale) in base ai contenuti.

Selezione singola

È inoltre possibile creare un Supernodo selezionando un unico nodo e utilizzando le opzioni di menu per definirne inizio e fine oppure incapsulando tutti gli elementi a valle del nodo selezionato.

- ▶ Fare clic sul nodo di inizio dell'incapsulamento.
- ▶ Dal menu Supernodo scegliere:
Crea Supernodo > Da qui

Figura 9-5
Creazione di un Supernodo tramite selezione singola utilizzando il menu di scelta rapida



I Supernodi possono inoltre essere creati in modo più interattivo selezionando il nodo iniziale e quello terminale della sezione dello stream che si desidera incapsulare:

- ▶ Fare clic sul primo o sull'ultimo nodo che si desidera includere nel Supernodo.
- ▶ Dal menu Supernodo scegliere:
Crea Supernodo > Seleziona...
- ▶ In alternativa, è possibile utilizzare le opzioni del menu di scelta rapida facendo clic con il pulsante destro del mouse sul nodo desiderato.
- ▶ Il cursore assume la forma dell'icona di un Supernodo per indicare che è necessario selezionare un altro punto nello stream. Spostare il cursore a monte o a valle del frammento del Supernodo selezionato e fare clic su un nodo. Questa operazione determina la sostituzione di tutti i nodi inclusi tra i due punti selezionati con l'icona del Supernodo.

Nota: i nodi selezionati devono appartenere a uno stream continuo o biforcuto. Non è possibile selezionare nodi che non siano adiacenti o connessi in qualche modo.

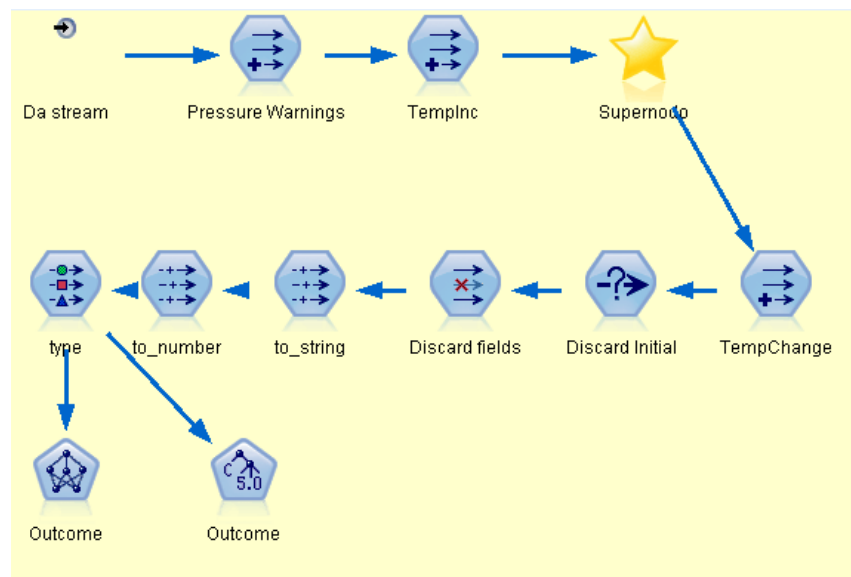
Nidificazione di Supernodi

I Supernodi possono essere nidificati all'interno di altri Supernodi. Ai Supernodi nidificati si applicano le stesse regole valide per ogni tipo di Supernodo (di input, di elaborazione e terminale). Per esempio, per un Supernodo di elaborazione contenente Supernodi nidificati è necessario che

il flusso di dati nei due sensi sia assicurato attraverso tutti i Supernodi nidificati. Se uno dei Supernodi nidificati è terminale, il flusso di dati nella gerarchia di Supernodi non viene eseguito.

Figura 9-6

Supernodo di elaborazione nidificato all'interno di un altro Supernodo di elaborazione



I Supernodi di input e terminali possono contenere altri tipi di Supernodi nidificati, ma anche in questo caso sono valide le stesse regole di base per la creazione di Supernodi.

Esempio di Supernodi validi

Quasi tutti gli elementi creati in IBM® SPSS® Modeler possono essere incapsulati in un Supernodo. Di seguito vengono presentati esempi di Supernodi validi:

Figura 9-7

Supernodo di elaborazione valido con due connessioni in un flusso di stream valido

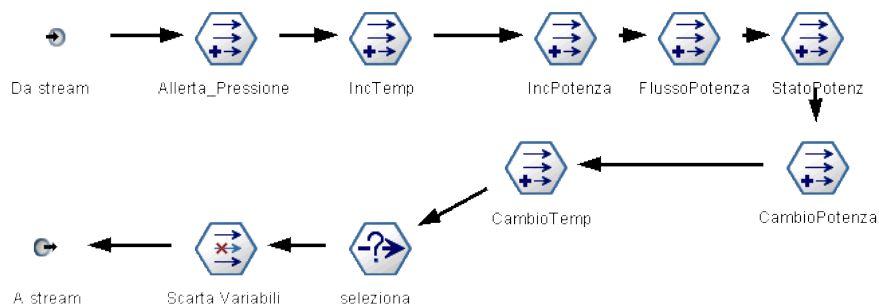


Figura 9-8
Supernodo terminale valido contenente uno stream separato utilizzato per verificare modelli generati

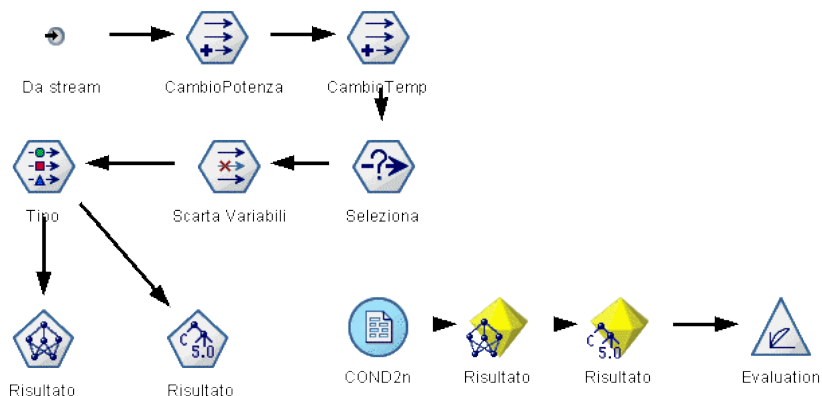
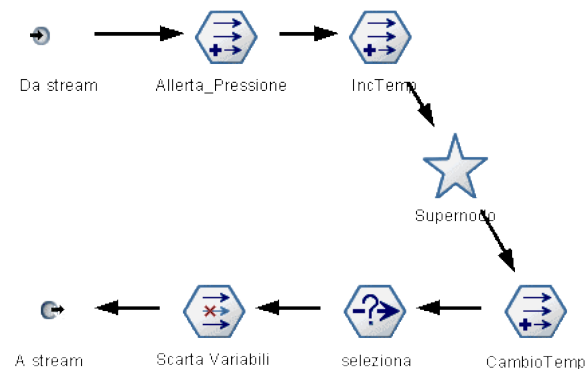


Figura 9-9
Supernodo di elaborazione valido contenente un Supernodo nidificato



Esempi di Supernodi non validi

L'aspetto più importante nella creazione di Supernodi validi consiste nell'assicurarsi che il flusso di dati possa essere eseguito in modo corretto attraverso le connessioni del Supernodo. Se sono presenti due connessioni (ovvero un Supernodo di elaborazione), il flusso di dati nello stream deve essere garantito dal connettore iniziale al connettore terminale. Analogamente, in un Supernodo di input è necessario verificare che il flusso di dati venga eseguito correttamente dal nodo di input all'unico connettore che restituisce i dati nella modalità Zoom indietro del Supernodo.

Figura 9-10

Supernodo di input non valido: nodo di input non connesso al percorso del flusso di dati

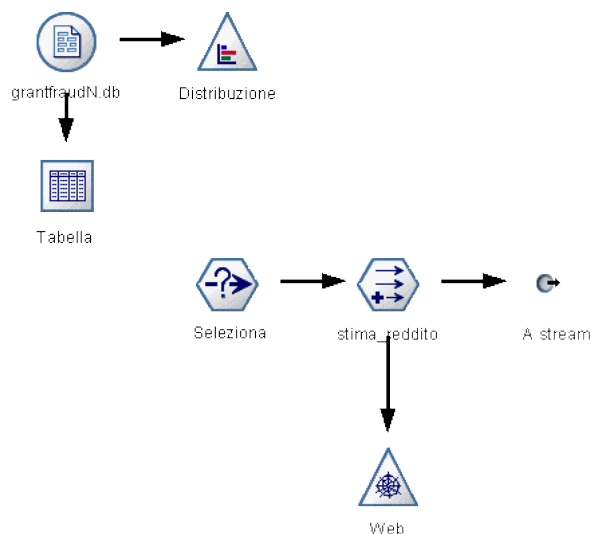
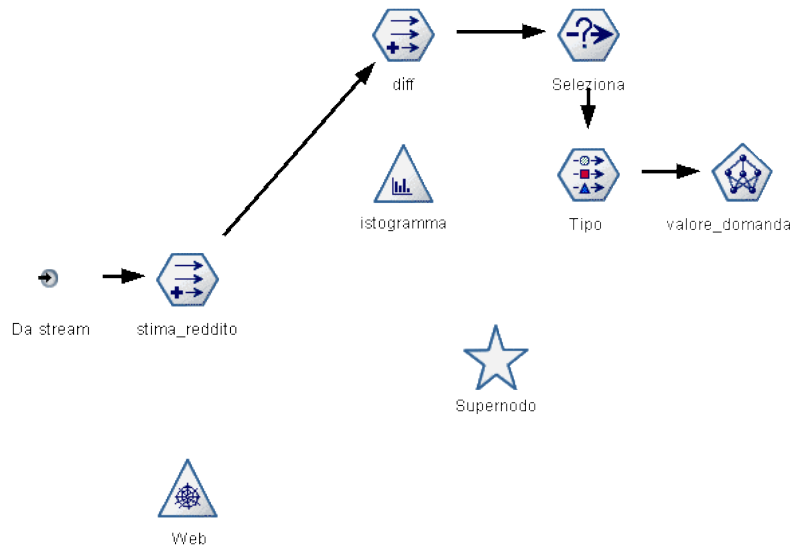


Figura 9-11

Supernodo terminale non valido: Supernodo nidificato non connesso al percorso del flusso di dati



Blocco di Supernodi

Una volta creato un Supernodo, è possibile bloccarlo con una password per impedire che venga modificato. Questa protezione può risultare per esempio necessaria nella creazione di stream, o parti di stream, come modelli a valore fisso che devono essere utilizzati nell'organizzazione da addetti con meno esperienza nell'impostazione delle interrogazioni di IBM® SPSS® Modeler.

Il blocco di un Supernodo non impedisce agli utenti di immettere nella scheda Parametri valori relativi agli eventuali parametri definiti; un Supernodo bloccato può essere inoltre eseguito senza digitare la password.

Nota: il blocco e lo sblocco non possono essere eseguiti utilizzando script.

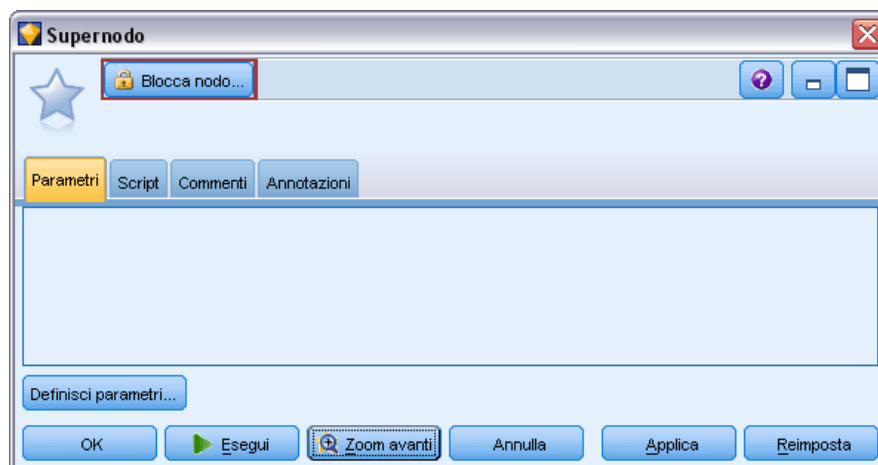
Blocco e sblocco di Supernodi

Avviso: una volta perse, le password non possono essere recuperate.

È possibile bloccare o sbloccare un Supernodo da una delle tre schede.

Figura 9-12

Blocco di un Supernodo

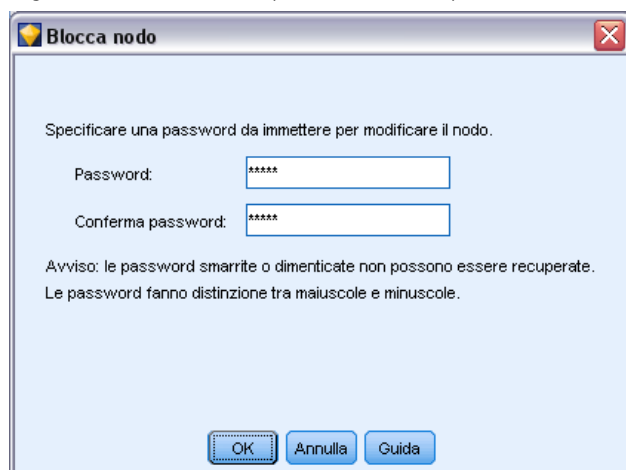


Fare clic su Blocca nodo.

Digitare e confermare la password.

Figura 9-13

Digitare e confermare la password del Supernodo.



- Fare clic su OK.

Un Supernodo protetto da password è riconoscibile nell'area di disegno dello stream dal simbolo del lucchetto che compare nella parte superiore sinistra dell'icona Supernodo.

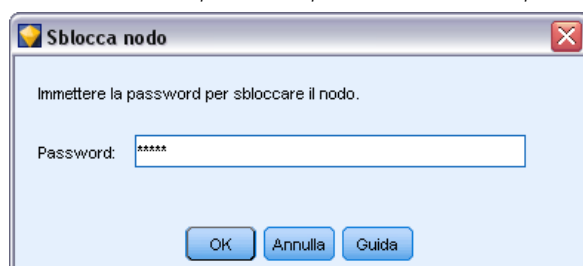
Figura 9-14
Supernodo di input bloccato come parte di uno stream



Sblocco di un Supernodo

- Per eliminare in modo definitivo la protezione password, fare clic su Sblocca nodo; viene richiesto di digitare la password.

Figura 9-15
Inserimento della password per sbloccare un Supernodo

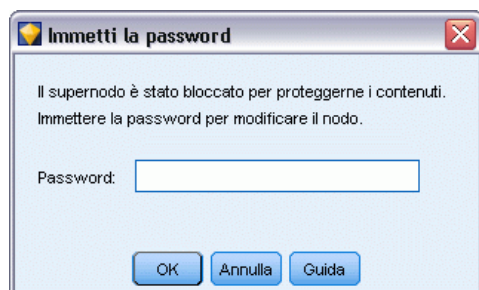


- Digitare la password e fare clic su OK; il Supernodo non è più protetto da password e nello stream scompare il simbolo del lucchetto accanto all'icona.

Modifica di un Supernodo bloccato

Se l'utente tenta di definire parametri o di effettuare lo zoom in avanti per visualizzare un Supernodo bloccato, viene richiesto di immettere la password.

Figura 9-16
Inserimento della password per effettuare lo zoom in avanti o modificare un Supernodo.



- Immettere la password e fare clic su OK.

L'utente è ora in grado di modificare le definizioni dei parametri ed eseguire lo zoom in avanti o indietro tutte le volte che lo desidera fino a quando non chiude lo stream in cui si trova il Supernodo.

Tenere presente che questo non elimina la protezione password, ma consente solo l'accesso all'utente per l'utilizzo del Supernodo. [Per ulteriori informazioni, vedere l'argomento Blocco e sblocco di Supernodi a pag. 462.](#)

Modifica Supernodo

Dopo aver creato un Supernodo è possibile analizzarlo più da vicino ingrandendolo; se il Supernodo è bloccato, verrà richiesta l'immissione della password. [Per ulteriori informazioni, vedere l'argomento Modifica di un Supernodo bloccato a pag. 463.](#)

Per visualizzare i contenuti di un Supernodo è possibile utilizzare l'icona della modalità Zoom avanti del Supernodo disponibile nella barra degli strumenti di IBM® SPSS® Modeler oppure il metodo seguente:

- ▶ Fare clic con il pulsante destro del mouse su un Supernodo.
- ▶ Dal menu di scelta rapida scegliere Zoom avanti.

I contenuti del Supernodo selezionato verranno visualizzati in un ambiente di SPSS Modeler leggermente diverso, contenente i connettori che indicano il flusso di dati nello stream o nel frammento di stream. In questo livello dell'area di disegno dello stream sono disponibili numerose operazioni che è possibile eseguire:

- Modificare il tipo di Supernodo, ovvero di input, di elaborazione o terminale.
- Creare parametri o modificare i valori di un parametro. I parametri vengono utilizzati negli script e nelle espressioni CLEM.
- Specificare opzioni di cache per il Supernodo e i relativi sottonodi.
- Creare o modificare lo script di un Supernodo (solo per i Supernodi terminali).

Modifica dei tipi di Supernodo

In alcuni casi può essere utile modificare il tipo di un Supernodo. Questa opzione è disponibile e si applica solo al livello della modalità Zoom avanti di un Supernodo. I tre tipi di Supernodi disponibili sono:

Supernodo di input	Una sola connessione in uscita
Supernodo di elaborazione	Due connessioni: una in entrata e una in uscita.
Supernodo finale	Una sola connessione in entrata

Modifica del tipo di un Supernodo

- ▶ Assicurarsi di essere nella modalità Zoom avanti del Supernodo.
- ▶ Dal menu Supernodo, selezionare Tipo Supernodo e quindi scegliere il tipo.

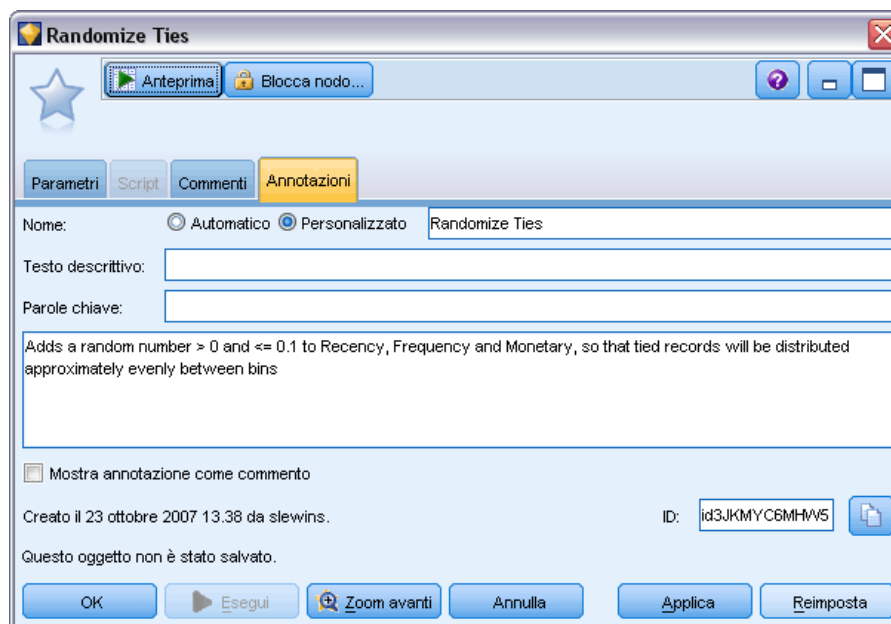
Annotazione e ridenominazione dei Supernodi

È possibile rinominare un Supernodo visualizzato nello stream nonché scrivere annotazioni per progetti o report. Per accedere a queste proprietà:

- ▶ Fare clic con il pulsante destro del mouse su un Supernodo (nella modalità Zoom indietro del Supernodo) e quindi scegliere Rinomina e annota.
- ▶ In alternativa, scegliere Rinomina e annota dal menu Supernodo. Questa opzione è disponibile in entrambe le modalità Zoom avanti e Zoom indietro.

In entrambi i casi verrà aperta una finestra di dialogo con la scheda Annotazioni selezionata. Utilizzare le opzioni disponibili per personalizzare il nome visualizzato nell'area di disegno dello stream e fornire informazioni sulle operazioni eseguite per il Supernodo.

Figura 9-17
Annotazione di un Supernodo



Utilizzo di commenti con i Supernodi

Se si crea un Supernodo da un nodo o da un insieme per il quale sono stati creati dei commenti, è necessario includere tali commenti nella selezione per la creazione del Supernodo se si desidera che i commenti appaiano nello stesso. Se non viene incluso nella selezione, il commento rimarrà scollegato nello stream nel momento in cui viene creato il Supernodo.

Quando si espande un Supernodo che include commenti, i commenti vengono reintegrati nella posizione in cui si trovavano prima della creazione del Supernodo.

Quando si espande un Supernodo che include oggetti commentati senza che i commenti siano inclusi nel Supernodo, gli oggetti vengono reintegrati nella loro posizione originale senza che i commenti vengano nuovamente associati.

Parametri dei Supernodi

In IBM® SPSS® Modeler è possibile impostare variabili definite dall'utente, per esempio *Valminimo*, i cui valori possono essere specificati per l'utilizzo in script o espressioni CLEM. Queste variabili vengono denominate **parametri**. È possibile impostare parametri per stream, sessioni e Supernodi. Tutti i parametri impostati per un Supernodo sono disponibili durante la creazione di espressioni CLEM per il Supernodo specifico o qualsiasi nodo nidificato. Le impostazioni di parametri relativi a Supernodi nidificati non sono disponibili per il Supernodo genitore.

Per la creazione e l'impostazione di parametri relativi a Supernodi è necessario eseguire due passaggi:

- Definire i parametri per il Supernodo.
- Specificare quindi il valore per ogni parametro del Supernodo.

Questi parametri possono essere utilizzati in espressioni CLEM per tutti i nodi incapsulati.

Definizione dei parametri di un Supernodo

È possibile definire i parametri per un Supernodo in entrambe le modalità Zoom avanti e Zoom indietro. I parametri definiti si applicano a tutti i nodi incapsulati. Per definire i parametri di un Supernodo è innanzitutto necessario accedere alla scheda Parametri della finestra di dialogo del Supernodo. Utilizzare uno dei metodi seguenti per aprire la finestra di dialogo:

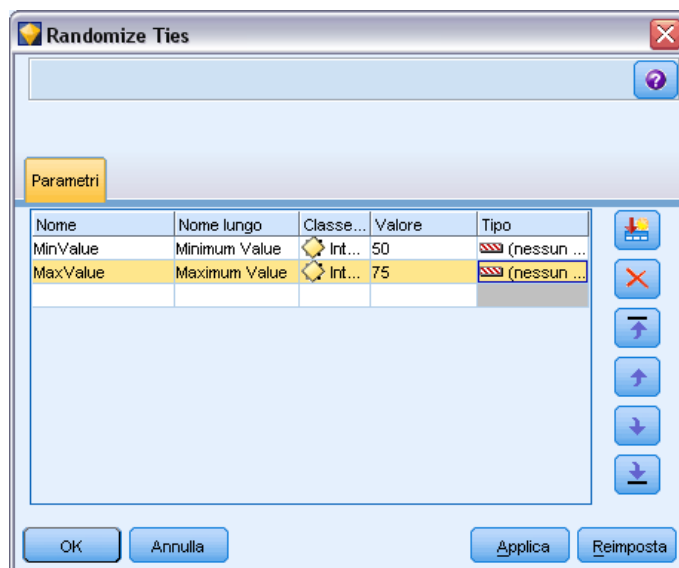
- Fare doppio clic su un Supernodo nello stream.
- Dal menu Supernodo scegliere Imposta parametri.
- In alternativa, nella modalità Zoom avanti di un Supernodo scegliere Imposta parametri dal menu di scelta rapida.

Dopo aver aperto la finestra di dialogo, la scheda Parametri è visibile per qualsiasi parametro definito in precedenza.

Per definire un nuovo parametro

- ▶ Fare clic sul pulsante Definisci parametri per aprire la finestra di dialogo.

Figura 9-18
Definizione di parametri per un Supernodo



Nome. In questa colonna sono visualizzati i nomi dei parametri. È possibile creare un nuovo parametro inserendo nel campo un nome, Per esempio, per creare un parametro per la temperatura minima, è possibile digitare *valminimo*. Non includere il prefisso *\$P-* che indica un parametro nelle espressioni CLEM. Tale nome è inoltre quello visualizzato nel Generatore espressioni CLEM.

Nome lungo. Elenca il nome descrittivo di ogni parametro creato.

Classe archiviazione. Selezionare un tipo di archiviazione dall'elenco. La classe di archiviazione indica la modalità di memorizzazione dei valori dei dati nel parametro. Per esempio, quando si utilizzano valori contenenti zero iniziali che si desidera mantenere (per esempio 008), è necessario selezionare Stringa come tipo di archiviazione. In caso contrario, gli zeri verranno rimossi dal valore. I tipi di archiviazione disponibili sono stringa, intero, reale, ora, data e timestamp. Per i parametri data, tenere presente che è necessario specificare i valori con la notazione standard ISO, come illustrato di seguito.

Valore. Elenca il valore corrente di ogni parametro. Modificare il parametro in base alle esigenze specifiche. Si noti che per specificare i parametri data è necessario utilizzare la notazione standard ISO (ovvero, YYYY-MM-DD). Le date specificate in altri formati non sono accettate.

Tipo (facoltativo). Se si pianifica il deployment dello stream in un'applicazione esterna, selezionare un livello di misurazione dall'elenco. In caso contrario, si consiglia di non modificare la colonna *Tipo*. Se si desidera specificare dei vincoli per il valore del parametro quale un limite superiore e inferiore per un intervallo numerico, selezionare Specifica dall'elenco.

Si noti che è possibile impostare le opzioni nome lungo, classe archiviazione e tipo per i parametri solo tramite l'interfaccia utente e non tramite script.

Fare clic sulle frecce a destra per spostare il parametro selezionato più in alto o in basso nell'elenco dei parametri. Per rimuovere il parametro selezionato, utilizzare il pulsante di eliminazione contrassegnato da una *X*.

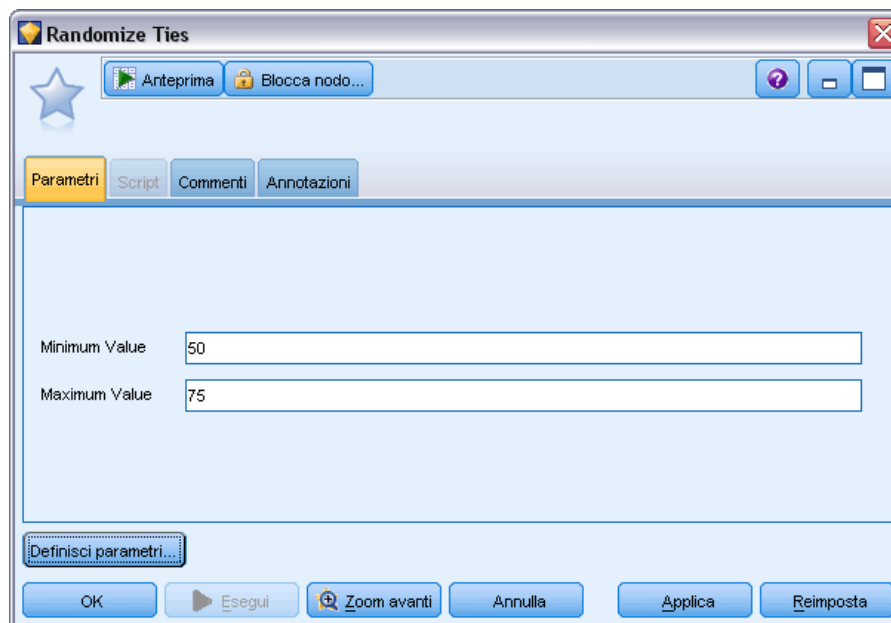
Impostazione dei parametri di un Supernodo

Dopo aver definito i parametri di un Supernodo, è possibile specificare i valori utilizzando i parametri in script o espressioni CLEM.

Per specificare i parametri di un Supernodo

- ▶ Fare doppio clic sull'icona del Supernodo per aprire la finestra di dialogo del Supernodo.
- ▶ In alternativa, scegliere Imposta parametri dal menu Supernodo.
- ▶ Fare clic sulla scheda Parametri. *Nota:* i campi inclusi in questa finestra di dialogo sono gli stessi campi definiti nella finestra di dialogo che viene aperta facendo clic sul pulsante Definisci parametri in questa scheda.
- ▶ Immettere un valore nella casella di testo per ogni parametro creato. Per esempio, è possibile impostare il valore *valminimo* su una particolare soglia significativa. Questo parametro può quindi essere utilizzato in numerose operazioni, quali la selezione a scopo di ulteriori analisi dei record superiori o inferiori alla soglia specificata.

Figura 9-19
Definizione dei parametri per un Supernodo



Utilizzo dei parametri di un Supernodo per l'accesso alle proprietà del nodo proprietà

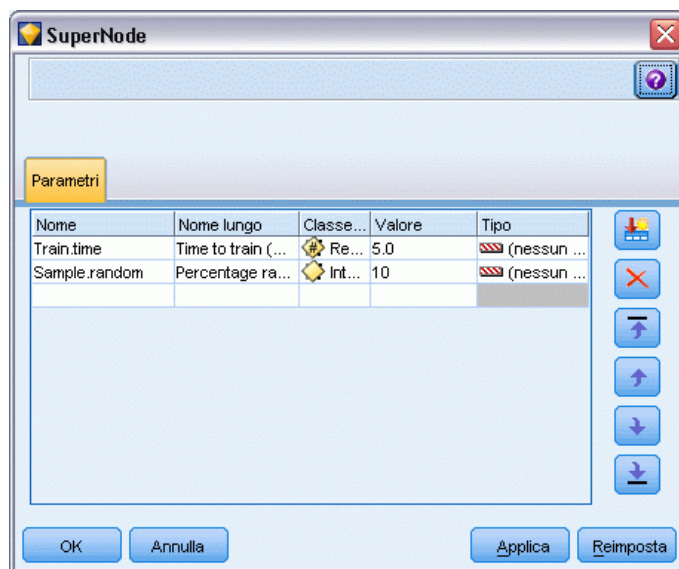
I parametri del Supernodo possono inoltre essere utilizzati per definire le proprietà del nodo (denominate anche **parametri di configurazione**) per i nodi incapsulati. Per esempio, si supponga di voler specificare che un Supernodo esegua l'addestramento di un nodo Rete neurale incapsulato per un determinato periodo di tempo tramite un campione casuale dei dati disponibili. Se si utilizzano i parametri, sarà possibile specificare valori relativi al periodo di tempo e alla percentuale del campione.

Figura 9-20
Frammento di stream incapsulato in un Supernodo



Il Supernodo di esempio contiene un nodo Campione denominato *Campione* e un nodo Rete neurale denominato *Addestra*. È possibile utilizzare le finestre di dialogo del nodo per impostare **Campione** del nodo Campione su % casuale e **Condizione arresto** del nodo Rete neurale su Ora. Dopo aver specificato queste opzioni, sarà possibile accedere alle proprietà del nodo tramite i parametri e indicare valori specifici per il Supernodo. Nella finestra di dialogo del Supernodo fare clic su Definisci parametri e creare i seguenti parametri:

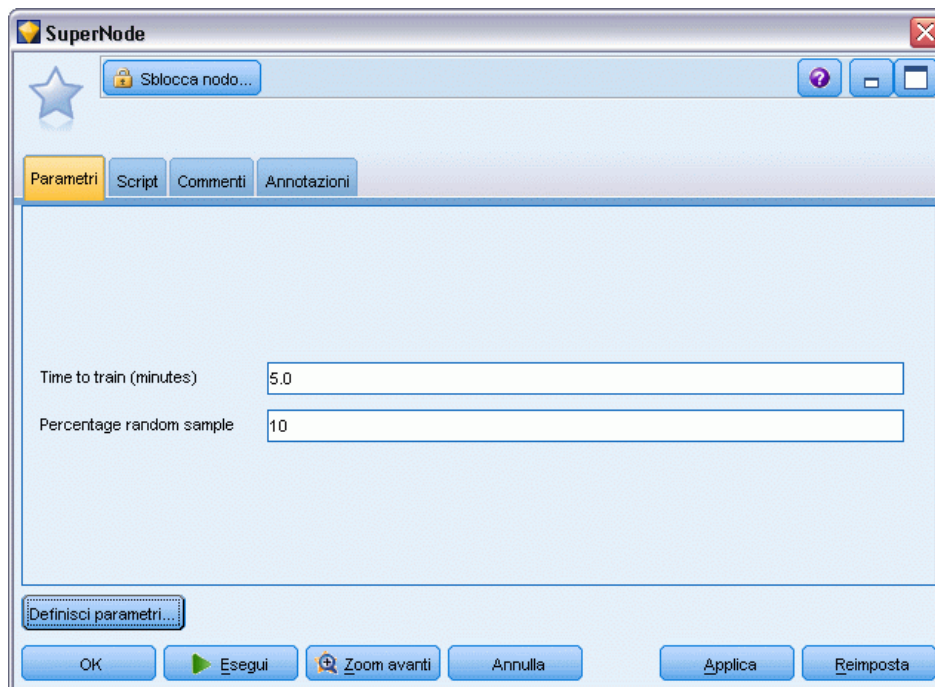
Figura 9-21
Definizione di parametri per l'accesso alle proprietà del nodo



Nota: i nomi di parametri, quali *Campione.random*, utilizzano la sintassi corretta per il riferimento alle proprietà del nodo, dove *Campione* rappresenta il nome del nodo e *random* una proprietà del nodo. Per ulteriori informazioni, vedere l'argomento [Panoramica sui riferimenti alle proprietà in il capitolo 9 in IBM SPSS Modeler 14.2 Guida per script e automazione](#).

Dopo aver definito questi parametri, è possibile modificare valori per le proprietà dei nodi Campione e Rete neurale senza riaprire le relative finestre di dialogo. È infatti sufficiente scegliere Imposta parametri dal menu Supernodo per accedere alla scheda Parametri della finestra di dialogo del Supernodo, in cui è possibile specificare nuovi valori per % casuale e Ora. Questo risulta particolarmente utile quando si analizzano i dati durante numerose iterazioni di generazione di modelli.

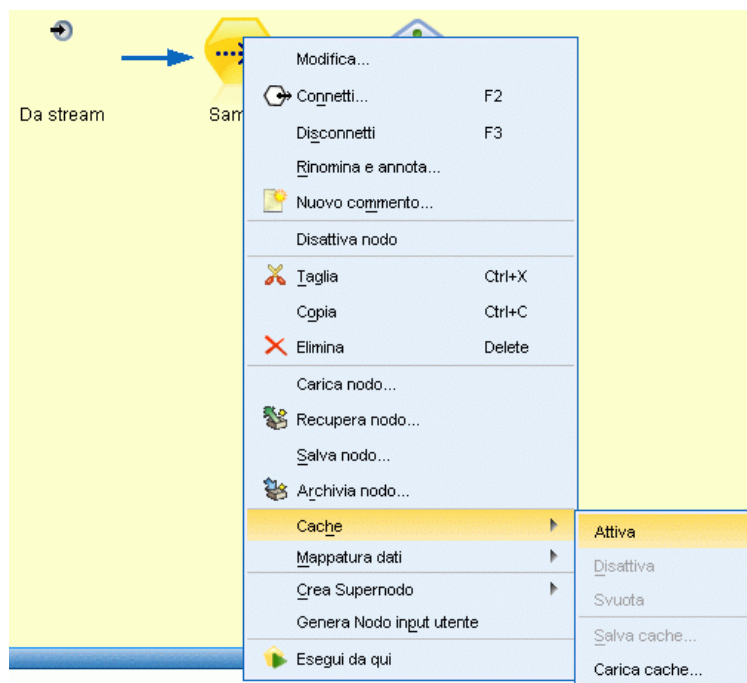
Figura 9-22
Definizione dei valori per le proprietà di nodi nella scheda Parametri della finestra di dialogo del Supernodo



Supernodi e cache

Dall'interno di un Supernodo è possibile inserire nella cache tutti i nodi eccetto i nodi terminali. Per controllare l'esecuzione della cache, fare clic con il pulsante destro del mouse su un nodo e scegliere una delle numerose opzioni dal menu di scelta rapida Cache. Questa opzione di menu è disponibile sia all'interno sia all'esterno di un Supernodo, nonché per i nodi incapsulati in un Supernodo.

Figura 9-23
Selezione di opzioni di cache per un Supernodo



Per le cache dei Supernodi esistono diverse regole:

- Se per uno dei nodi incapsulati in un Supernodo è attiva la cache, anche per il Supernodo sarà attiva la cache.
- La disattivazione della cache di un Supernodo determina la disattivazione della cache per *tutti* i nodi incapsulati.
- L'attivazione della cache in un Supernodo determina in realtà l'attivazione della cache per l'ultimo sottonodo che supporta la cache. In altre parole, se l'ultimo sottonodo è un nodo Selezione, la cache verrà attivata per tale nodo. Se l'ultimo sottonodo è un nodo finale (che non consente l'esecuzione della cache) la cache verrà attivata per il primo nodo a monte in grado di supportarla.
- Dopo aver impostato cache per i sottonodi di un Supernodo, qualsiasi attività a monte del nodo per cui è attivata la cache, per esempio l'aggiunta o la modifica di nodi, determinerà lo svuotamento delle cache.

Supernodi e script

È possibile utilizzare il linguaggio di script di IBM® SPSS® Modeler per scrivere programmi semplici che consentono di manipolare ed eseguire i contenuti di un Supernodo terminale. Per esempio, è possibile che si desideri specificare l'ordine di esecuzione per uno stream complesso. Se, per esempio, un Supernodo contiene un nodo Globali che deve essere eseguito prima di un nodo Plot, è possibile creare uno script che esegua il nodo Globali per primo. I valori calcolati da questo nodo, quali la media o la deviazione standard, possono quindi essere utilizzati quando viene eseguito il nodo Plot.

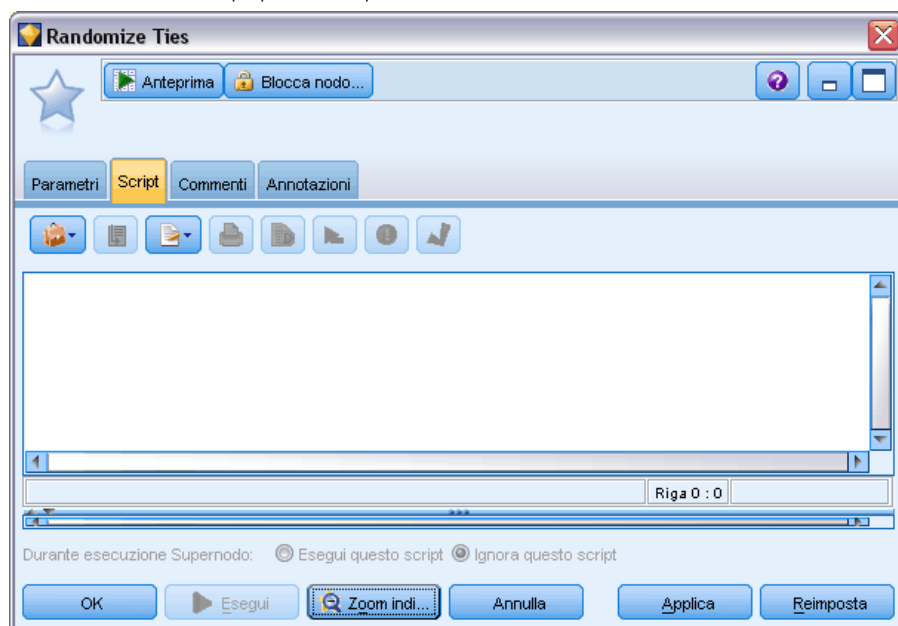
La scheda Script della finestra di dialogo del Supernodo è disponibile solo per i Supernodi terminali.

Per aprire la finestra di dialogo di script per un Supernodo terminale

- ▶ Fare clic con il pulsante destro del mouse nell'area di disegno del Supernodo e scegliere Script Supernodo.
- ▶ In alternativa, sia in modalità Zoom avanti che Zoom indietro è possibile scegliere Script del Supernodo dal menu Supernodo.

Nota: Gli script dei Supernodi vengono eseguiti per stream e Supernodi solo se nella finestra di dialogo è stata selezionata l'opzione Esegui questo script.

Figura 9-24
Creazione di uno script per un Supernodo



Le opzioni specifiche per lo script e il relativo utilizzo in SPSS Modeler sono descritti nel documento *Scripting and Automation Guide*, disponibile nel DVD di SPSS Modeler. [Per ulteriori informazioni, vedere l'argomento Panoramica sugli script in il capitolo 2 in IBM SPSS Modeler 14.2 Guida per script e automazione.](#)

Salvataggio e caricamento dei Supernodi

Uno dei vantaggi dei Supernodi è la possibilità di salvarli e riutilizzarli in altri stream. Per il salvataggio e il caricamento di Supernodi, viene utilizzata un'estensione *slb*.

Per salvare un Supernodo

- ▶ Utilizzare la modalità Zoom avanti per il Supernodo.

- ▶ Dal menu Supernodo scegliere Salva Supernodo.
- ▶ Specificare un nome di file e una directory nella finestra di dialogo.
- ▶ Specificare se si desidera aggiungere il Supernodo salvato al progetto corrente.
- ▶ Scegliere Salva.

Per caricare un Supernodo

- ▶ Dal menu Inserisci nella finestra di IBM® SPSS® Modeler, scegliere Supernodo.
- ▶ Selezionare un file di Supernodo (*slb*) dalla directory corrente o passare a una directory diversa.
- ▶ Fare clic su Carica.

Nota: i Supernodi importati includono valori di default per tutti i parametri. Per modificare i parametri, fare doppio clic su un Supernodo nell'area di disegno dello stream.

Note

This information was developed for products and services offered worldwide.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785, U.S.A.

For license inquiries regarding double-byte character set (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing, Legal and Intellectual Property Law, IBM Japan Ltd., 1623-14, Shimotsuruma, Yamato-shi, Kanagawa 242-8502 Japan.

Il seguente paragrafo non si applica per il Regno Unito o altri paesi in cui le presenti disposizioni non sono conformi alle leggi locali: SPSS INC., UNA SOCIETÀ IBM, FORNISCE IL PRESENTE DOCUMENTO “COSÌ COM'È” SENZA GARANZIA DI ALCUN TIPO, ESPRESSA O IMPLICITA, COMPRESE, TRA LE ALTRE, LE GARANZIE IMPLICITE DI NON VIOLAZIONE DEI DIRITTI ALTRUI, COMMERCIALIZZABILITÀ O IDONEITÀ PER UNO SCOPO SPECIFICO. Alcuni stati non consentono limitazioni di garanzie espresse o implicite in determinate transazioni, pertanto quanto sopra potrebbe non essere applicabile.

Le presenti informazioni possono includere imprecisioni tecniche o errori tipografici. Le modifiche periodiche apportate alle informazioni contenute in questa pubblicazione verranno inserite nelle nuove edizioni della pubblicazione. SPSS Inc. può apportare miglioramenti e/o modifiche al/ai prodotto/i e/o al/ai programma/i descritti nella presente pubblicazione in qualsiasi momento senza preavviso.

Eventuali riferimenti contenuti nelle presenti informazioni a siti Web non SPSS e non IBM vengono forniti per comodità dell'utente e non implicano in alcun modo l'approvazione di tali siti Web. I materiali presenti in tali siti Web non sono parte dei materiali di questo prodotto SPSS Inc. e l'utilizzo di tali siti Web è a rischio esclusivo dell'utente.

Al momento dell'invio di informazioni a IBM o SPSS, l'utente concede a IBM e SPSS un diritto non esclusivo a utilizzare o distribuire tali informazioni nel modo che ritiene appropriato senza incorrere in alcun obbligo nei confronti dell'utente.

Le informazioni relative ai prodotti non SPSS provengono dai fornitori di tali prodotti, dagli annunci pubblicati o da altre fonti pubbliche disponibili. SPSS non ha testato tali prodotti e non può confermare l'accuratezza delle dichiarazioni relative alle prestazioni, alla compatibilità o ad altre caratteristiche relative ai prodotti non SPSS. Le domande relative alle funzionalità dei prodotti non SPSS devono essere indirizzate ai fornitori di tali prodotti.

Licensees of this program who wish to have information about it for the purpose of enabling: (i) the exchange of information between independently created programs and other programs (including this one) and (ii) the mutual use of the information which has been exchanged, should contact:

IBM Software Group, Attention: Licensing, 233 S. Wacker Dr., Chicago, IL 60606, USA.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this document and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement or any equivalent agreement between us.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

All statements regarding IBM's future direction or intent are subject to change or withdrawal without notice, and represent goals and objectives only.

Le presenti informazioni includono esempi di dati e report utilizzati in operazioni aziendali quotidiane. Per fornire una descrizione il più possibile esaustiva, gli esempi includono nomi di persone, società, marchi e prodotti. Tutti questi nomi sono fittizi e ogni somiglianza a nomi e indirizzi utilizzati da aziende reali è puramente casuale.

If you are viewing this information softcopy, the photographs and color illustrations may not appear.

Marchi

IBM, il logo IBM e [ibm.com](http://www.ibm.com) sono marchi di IBM Corporation, registrati in numerose giurisdizioni nel mondo. Un elenco aggiornato dei marchi IBM è disponibile sul Web all'indirizzo <http://www.ibm.com/legal/copytrade.shtml>.

SPSS è un marchio di SPSS Inc., una società IBM, registrati in numerose giurisdizioni nel mondo.

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.

IT Infrastructure Library è un marchio registrato della Central Computer and Telecommunications Agency ora parte dell'Office of Government Commerce.

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.

ITIL è un marchio registrato e un marchio comunitario registrato dell'Office of Government Commerce ed è registrato presso l'Ufficio Brevetti degli Stati Uniti.

UNIX è un marchio registrato di The Open Group negli Stati Uniti e in altri paesi.

Cell Broadband Engine è un marchio di Sony Computer Entertainment, Inc. negli Stati Uniti e/o in altri paesi ed è utilizzato su licenza.

Java e tutti i marchi e i logo basati su Java sono marchi di Sun Microsystems, Inc. negli Stati Uniti e/o negli altri paesi.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

I nomi di altri prodotti o servizi possono essere marchi di IBM, SPSS o altre società.

- 1-ogni-*n*, 64
 - accorpamento dei dati di serie storica, 207
 - aggiunta
 - record, 73
 - aggregazione dei dati di serie storica, 207
 - aggregazione di record, 196
 - alterna, 338
 - animazione
 - nelle visualizzazioni, 235
 - animazione nei grafici, 232, 234
 - anno fiscale
 - nodo Intervalli di tempo, 215
 - anonimizzazione dei nomi di campi, 147
 - ANOVA
 - nodo Medie, 394
 - ANOVA monodirezionale
 - nodo Medie, 394
 - anti-join, 81
 - apertura
 - oggetti di output, 347
 - applicazioni di supporto, 405
 - archiviazione di campi
 - conversione, 165
 - aree nei grafici, 314
 - assegnazione di tipi di dati, 56, 96, 125
 - associazione per colonna, 417
 - associazione per riga, 417
 - attributi del campo, 140
 - attributi del tipo, 140

 - bacchetta magica nei grafici, 317
 - bin decile, 181
 - bin percentile, 181
 - bin quartile, 181
 - bin quintile, 181
 - bin ventile, 181
 - blocchi
 - nodo Discretizza, 181
 - blocco di Supernodi, 461–462
 - browser del nodo Analisi
 - interpretazione, 367
 - browser del nodo Esplora
 - generazione di grafici, 383
 - generazione di nodi, 383
 - Menu File, 374
 - Menu Modifica, 374
 - browser del nodo Qualità
 - generazione di nodi Filtro, 381
 - generazione di nodi Seleziona, 382
 - browser del nodo Report, 403
 - browser del nodo Statistiche
 - generazione di nodi Filtro, 393
 - interpretazione, 391
 - Menu Genera, 391

 - browser del nodo Tabella
 - Menu Genera, 357
 - ricerca, 357
 - riordinamento delle colonne, 352, 357
 - selezione di celle, 352, 357

 - cache
 - Supernodi, 470
 - calcola durate
 - preparazione automatica dati, 103
 - calcolo durate
 - preparazione automatica dati, 103
 - campi
 - anonimizzazione dei dati, 169
 - delimitatori, 23
 - derivazione di più campi, 156
 - etichette di campi e valori, 56, 125, 134
 - riordinamento, 227
 - selezione multipla, 158
 - trasposizione, 200
 - campi chiave, 74, 196
 - campi chiavi principali
 - nodo Database, 414
 - campi delle etichette
 - assegnazione di etichette ai record in output, 138
 - campi di partizione, 56, 125, 138, 193–194
 - campionamento di dati contigui, 64
 - campioni di addestramento
 - bilanciamento, 72
 - partizionamento dei dati, 193–194
 - campioni di test
 - partizionamento dei dati, 193–194
 - campioni di validazione
 - partizionamento dei dati, 193–194
 - campioni non casuali, 63–64
 - campioni ponderati, 68
 - campioni raggruppati, 63–64, 68
 - campioni sistematici, 63–64
 - campioni stratificati, 63–64, 68–69
- campo Conteggio
 - integrazione o aggregazione di serie storiche, 209
 - nodo Intervalli di tempo, 209
- campo EtichettaTempo
 - nodo Intervalli di tempo, 207
- campo IndiceTempo
 - nodo Intervalli di tempo, 207
- cancelare valori, 56
- caratteri di commento
 - in nodi Testo variabile, 22
- Caratteri EOL, 22
- caricamento di massa, 417
- categorizzazione con supervisione, 186
- chi-quadrato
 - nodo Matrice, 362
- Chi-quadrato di Pearson
 - nodo Matrice, 362

- chiavi contigue, 74
- cluster, 338
- codifica, 23, 26, 420
- codifica fittizia, 196
- codifica UTF-8, 23, 26, 420
- Cognos, vedere IBM Cognos BI, 41
- collegamenti
 - nodo Web, 283
- colore
 - nelle visualizzazioni, 233
- comando CREATE INDEX, 414
- combinazione di dati, 90
 - da più file, 80
- comma, 23, 141
- commenti
 - utilizzo con i Supernodi, 465
- concatenazione di record, 90
- condizioni
 - definizione di una serie, 164
- connessioni
 - a IBM SPSS Collaboration and Deployment Services Repository, 7
- connessioni database
 - definizione, 16
 - valori preimpostati, 17
- conteggi
 - nodo Discretizza, 181
 - output del nodo Statistiche, 391
- conteggio uguale
 - nodo Discretizza, 181
- contrassegno di elementi, 314, 317
- controllo dei tipi, 137
- conversione dei livelli di misurazione, 129
- conversione di insiemi in flag, 196–197
- coordinate polari, 333
- copia degli attributi del tipo, 140
- copia delle visualizzazioni, 338
- correlazioni, 389
 - etichette descrittive, 390
 - in valore assoluto, 390
 - nodo Medie, 399
 - output del nodo Statistiche, 391
 - probability, 390
 - significatività, 390
- correlazioni di Pearson
 - nodo Medie, 399
 - output del nodo Statistiche, 391
- costi
 - grafici di valutazione, 302
- creazione
 - nuovi campi, 154–155
- creazione di flag, 196, 199
- CRISP-DM
 - Data Understanding, 6
- data, formato di archiviazione, 28, 54
- data/ora, 127
- database
 - caricamento di massa, 417
 - livelli di supporto, 13
- database ADO
 - importazione, 32
- database di grandi dimensioni, 60
 - esecuzione di un'esplorazione dei dati, 369
- database In2data
 - importazione, 32
- database Quanvert
 - importazione, 32
- date
 - impostazione di formati, 141, 143
- dati
 - aggregazione, 73
 - analisi, 60
 - anonimizzazione, 169
 - esplorazione, 369
 - preparazione, 60
 - storage, 28, 54, 166, 168
 - tipo di archiviazione, 132
- dati annuali
 - nodo Intervalli di tempo, 214
- dati caso
 - nodo Input Data Collection, 31
- dati categoriali, 127, 129
- dati continui, 127, 129, 134
- dati CSV
 - importazione, 32
- dati di campionamento, 69
- dati di riepilogo, 73
- dati di serie storiche
 - aggregazione, 204, 207
 - assegnazione di etichette, 204–205, 207, 209
 - creazione dai dati, 207
 - definizione, 204–205, 207, 209
 - intervalli, 205
 - periodo di stima, 209
 - record di controllo, 209
 - riempimento, 204, 207
- dati di sondaggi
 - importazione, 31, 37–38
 - nodo Input Data Collection, 31
- dati di sondaggi Data Collection
 - importazione, 31
- dati di testo a campi fissi, 24
- dati di testo a campi liberi, 20
- dati di testo delimitati, 20
- dati distorti, 71
- dati flag, 127
- dati mensili
 - nodo Intervalli di tempo, 216
- dati nominali, 127, 135
- dati non distorti, 71
- dati ordinali, 128, 135
- dati Quancept
 - importazione, 32

- dati Quantum
 - importazione, 32
- dati ricerche di mercato
 - importazione, 31, 38
 - nodo Input Data Collection, 31
 - nodo Input IBM SPSS Data Collection, 37
- dati ritardati, 225
- dati sbilanciati, 71
- dati senza tipo, 128
- dati settimanali
 - nodo Intervalli di tempo, 217
- dati sintetici
 - nodo Input utente, 50
- dati Surveycraft
 - importazione, 32
- dati trimestrali
 - nodo Intervalli di tempo, 215
- dati Triple-S
 - importazione, 32
- decimali
 - formati di visualizzazione, 143
- Definizione dei provider di dati, 7
- definizione dei riquadri, 232, 234
- delimitatori, 22–23, 417
- derivazione multipla, 156
- deviazione standard
 - nodo Discretizza, 185
 - nodo Globali, 404
 - output del nodo Statistiche, 391
- deviazione standard per aggregazione, 74
- deviazione standard/media
 - utilizzato per la discretizzazione dei campi, 185
- dimensione commit, 417
- dimensione.
 - nelle visualizzazioni, 233
- diminuzione di dati, 61, 63
- direzione dei campi, 56, 125, 138
- discretizzazione ottimale, 186
- distribution, 266
- diversi input, 80
- documentazione, 2
- documenti MDD
 - importazione, 32
- DPD (Definizione dei provider di dati), 7
- duplicati
 - campi, 80, 144
 - record, 92

- editor query
 - nodo di input Database, 19
- elaborazione parallela
 - nodo Aggregazione, 75
 - ordinamento, 80
 - unione, 88
- elementi di tempo ciclico
 - preparazione automatica dati, 103
- elementi estetici
 - nelle visualizzazioni, 232
- elementi grafici
 - conversione, 335
 - modifica, 335
 - modificatori di collisione, 337
 - tipi, 335
- eliminazione
 - file delle mappe, 251
 - fogli di stile di visualizzazione, 251
 - modelli di visualizzazione, 251
 - oggetti di output, 347
- entrate
 - grafici di valutazione, 302
- esclusione dei campi inutilizzati
 - preparazione automatica dati, 102
- esecuzione
 - definizione dell'ordine di, 471
- esempi
 - cenni generali, 4
 - Guida alle applicazioni, 2
- esempi di applicazioni, 2
- esplorazione
 - esplorazione dei dati iniziale, 369
 - nodo Esplora, 369
- esplorazione dei dati
 - nodo Esplora, 369
- esplorazione di grafici, 309
 - aree, 314
 - bacchetta magica, 317
 - contrassegno di elementi, 317
 - sezioni di grafici, 310
- Esporta numero di decimali, 143
- esportazione
 - file delle mappe, 251
 - fogli di stile di visualizzazione, 251
 - modelli di visualizzazione, 251
 - output, 351
 - Supernodi, 472
- esportazione di dati
 - a Excel, 427
 - a IBM SPSS Statistics, 449
 - file DAT, 427
 - formato file piatto, 419
 - formato SAS, 426
 - formato XML, 428
 - in un database, 408
 - Nodo di esportazione IBM Cognos BI, 41, 422, 424
 - text, 427
- espressioni CLEM, 60
- etichette, 136
 - esportazione, 426, 450
 - importazione, 44, 434
 - nelle visualizzazioni, 233
 - specificata, 56, 125, 132, 134–136
- etichette dei dati
 - nelle visualizzazioni, 233

- etichette dei valori
 - nodo File Statistics, 434
- etichette di variabili
 - nodo Esporta Statistics, 449
 - nodo File Statistics, 434
- eventi
 - creazione, 291
- Excel
 - avvio da IBM SPSS Modeler, 427
- extension
 - campo derivato, 156

- fattori di bilanciamento, 72
- fattori di scala, 72
- file *.sav*, 434
- file *.sd2* (SAS), 42
- file *.slb*, 472
- file *.ssd* (SAS), 42
- file *.tpt* (SAS), 42
- file DAT
 - esportazione, 351, 427
 - salvataggio, 355
- file delimitati da virgole
 - esportazione, 351, 427
 - salvataggio, 355
- file delle mappe
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - posizione, 249
 - ridenominazione, 251
 - selezione in Selezionatore modelli per lavagna grafica, 246
- File di dati di IBM SPSS Statistics
 - importazione dati di sondaggi, 32
- file di dati *employee_data.sav*, 435
- file di Excel
 - esportazione, 427
- file di formato, 43
- file di output
 - salvataggio, 355
- file di testo, 20
 - esportazione, 427
- file di trasporto
 - nodo di input SAS, 42
- file piatti, 20
- file SMZ
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - ridenominazione, 251
- file XLS
 - esportazione, 427
- filtro di campi, 85, 143
 - per IBM SPSS Statistics, 451
- finestra del browser
 - Menu Genera, 362

- fogli di lavoro
 - importazione da Excel, 44
- fogli di stile
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - ridenominazione, 251
- fogli di stile di visualizzazione
 - applicazione, 342
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - posizione, 249
 - ridenominazione, 251
- forma
 - nelle visualizzazioni, 233
- formati
 - dati, 27, 141
- formati di archiviazione, 27
- formati di ora, 143
- formati di output, 355
- formati di visualizzazione
 - decimali, 143
 - numeri, 143
 - scientifico, 143
 - simbolo di raggruppamento, 143
 - valuta, 143
- formati di visualizzazione numeri, 143
- formato di visualizzazione scientifico, 143
- formato di visualizzazione valute, 143
- formato HDATA
 - nodo Input Data Collection, 31
- formato VDATA
 - nodo Input Data Collection, 31
- formula di derivazione dei campi, 158
- forzatura dei valori, 137
- frequenze
 - nodo Discretizza, 181
- funzione Max
 - aggregazione serie storica, 208
- funzione Media
 - aggregazione serie storica, 208
- funzione Min
 - aggregazione serie storica, 208
- funzione Moda
 - aggregazione serie storica, 208
- Funzione più recente
 - integrazione di serie storiche, 209
- funzione Primi
 - aggregazione serie storica, 208
- funzione Somma
 - aggregazione serie storica, 208
- funzione Ultimo
 - aggregazione serie storica, 208
- funzione Vero se uno qualsiasi è vero
 - aggregazione serie storica, 208

- funzione Vuoto
 integrazione di serie storiche, 209
- Generatore espressioni, 60
- generazioni di nodi da grafici, 319
 nodi Bilanciamento, 320
 nodi Filtro, 320
 nodi Nuovo campo, 320
 nodi Ricodifica, 321
 nodi Selezione, 320
- gestione dei vuoti, 56, 125, 132
 nodo Discretizza, 179
 riempimento di valori, 166
- gestione valori mancanti, 96
- giustificazione
 per campi, 141
- gradi di libertà
 nodo Matrice, 362
 nodo Medie, 398–399
- grafici
 3-D, 238
 aree, 314
 copia, 344
 da Lavagna grafica, 239
 dimensioni degli elementi grafici, 326
 distribuzioni, 261
 eliminazione di aree, 316
 esplorazione, 309
 esportazione, 344
 etichette dell'asse, 339
 foglio di stile, 341
 generazione da un'esplorazione dati, 383
 generazione di nodi, 319
 grafici, 252
 grafici di valutazione, 295
 istogrammi, 266
 nota a piè di pagina, 339
 plot multiplo, 275
 raccolte, 270
 rotazione di un'immagine 3-D, 238
 salvataggio, 344
 salvataggio dei layout modificati, 341
 salvataggio delle modifiche al layout, 341
 salvataggio dell'output, 355
 scheda Annotazioni, 237
 schede output, 237
 schema di colori di default, 341
 serie storiche, 291
 sezioni, 310
 stampa, 344
 title, 339
 Web, 280
- grafici 3-D, 238
 grafici a dispersione, 252, 275
 grafici a punti, 252, 275
 Grafici dei guadagni, 295, 306
 grafici dei profitti, 295, 306
- grafici delle risposte, 295, 306
 grafici di associazione, 280
 grafici lift, 295, 306
 grafici lineari, 252, 275
- hassubstring, funzione, 160
- HTML
 salvataggio dell'output, 356
- IBM SPSS Collaboration and Deployment Services
 Repository
 connessione a, 7
 utilizzo come posizione per modelli, fogli di stile di
 visualizzazione e mappe, 250
- IBM SPSS Modeler, 1
 documentazione, 2
- IBM SPSS Statistics
 avvio da IBM SPSS Modeler, 405, 444, 450
 nomi di campi validi, 451
 posizione della licenza, 405
- IBM SPSS Text Analytics, 2
- importanza
 confronto di medie, 396
 nodo Medie, 398–399
- importazione
 file delle mappe, 251
 fogli di stile di visualizzazione, 251
 modelli di visualizzazione, 251
 Supernodi, 472
- Imposta seme aleatorio
 campionamento di record, 67, 195
- impostazione automatica, 128, 131
- impostazioni automatiche, 323
- incrementi di minuti
 nodo Intervalli di tempo, 221–222
- incrementi di secondi
 nodo Intervalli di tempo, 223–224
- indici BITMAP
 tabelle di database, 415
- indicizzazione delle tabelle del database, 414
- insieme di dati principale, 91
- insiemi
 conversione in flag, 196–197
 trasformazione, 174, 177
- insiemi a categorie multiple, 147
- insiemi a risposta multipla
 definizione, 147
 eliminazione, 147
 insiemi a categorie multiple, 147
 insiemi di dicotomie multiple, 147
 nelle visualizzazioni, 243
 nodo Input Data Collection, 31
 nodo Input IBM SPSS Data Collection, 37–38
 nodo Input IBM SPSS Statistics, 435
- insiemi di dicotomie multiple, 147
- integrazione dei dati di serie storica, 207

- intervalli, 127
 - dati di serie storiche, 204
 - valori mancanti, 132
- intervalli di celle
 - file di Excel, 44
- intervalli di confidenza
 - nodo Medie, 398–399
- intervalli di numeri interi, 134
- intervalli di numeri reali, 134
- intervallo
 - output del nodo Statistiche, 391
- interventi
 - creazione, 291
- istanziamento, 56, 125, 127, 130–131
 - nodi Input, 58
- istruzioni Se, Allora, Altrimenti, 165

- jitter, 258, 338
- join, 80–81, 83
 - esterni parziali, 84
- join di insiemi di dati, 90
- join esterno, 81
- join interno, 81
- join parziali, 81, 84

- language
 - nodo Input IBM SPSS Data Collection, 35
- larghezza di colonna
 - per campi, 141
- layout diretto per grafici Web , 284
- layout reticolare per grafici Web, 284
- legenda
 - posizione, 338
- linea base
 - opzioni relative al nodo Valutazione, 301
- linea migliore
 - opzioni relative al nodo Valutazione, 301
- lista di campionamento, 63
- livellamento LOESS
 - nodo Plot, 256
- livellamento lowess *Vedere* livellamento LOESS
 - nodo Plot, 256
- livelli, supporto database, 13
- livello di misurazione, 56, 125
 - definizione, 127
 - modifiche in visualizzazioni, 241
 - nelle visualizzazioni, 243

- manager
 - scheda output, 347
- mappatura di campi, 410
- marchi, 475
- mascheramento dei dati da utilizzare in un modello, 169
- massimo
 - nodo Globali, 404
 - output del nodo Statistiche, 391
- matrice di coincidenza
 - nodo Analisi, 364
- means
 - confronto, 393–395, 397
- media
 - nodo Discretizza, 185
 - nodo Globali, 404
 - output del nodo Statistiche, 391
- Media della funzione più recente
 - integrazione di serie storiche, 209
- mediana
 - output del nodo Statistiche, 391
- membro (importazione SAS)
 - impostazioni, 43
- metadati, 56, 125, 132
 - nodo Input Data Collection, 31
- metodo chiave, 80
- minimo
 - nodo Globali, 404
 - output del nodo Statistiche, 391
- misurazioni orarie
 - nodo Intervalli di tempo, 219–220
- misurazioni quotidiane
 - nodo Intervalli di tempo, 218–219
- mode
 - output del nodo Statistiche, 391
- modelli
 - anonimizzazione dei dati per, 169
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - nodo Report, 401
 - ridenominazione, 251
- modelli di visualizzazione
 - eliminazione, 251
 - esportazione, 251
 - importazione, 251
 - posizione, 249
 - ridenominazione, 251
- modelli IBM SPSS Statistics, 440
 - dettagli insieme avanzati, 442
 - informazioni su, 440
 - insieme di modelli, 442
 - opzioni modello, 441
- modello di elaborazione CRISP-DM
 - Data Preparation, 96
- modifica dei grafici
 - dimensioni degli elementi grafici, 326
- modifica dei valori dei dati, 154
- modifica delle visualizzazioni, 321
 - aggiunta di effetti 3-D, 333
 - assi, 329
 - categorie, 331
 - colori e motivi, 324
 - combinazione categorie, 331
 - compressione delle categorie, 331
 - esclusione di categorie, 331

- forma dei punti, 326
- formati dei numeri, 328
- impostazioni automatiche, 322
- margini, 327
- ordinamento di categorie, 331
- posizione della legenda, 338
- proporzioni dei punti, 326
- regole, 322
- riempimento, 327
- riquadri, 333
- rotazione dei punti, 326
- scale, 329
- selezione, 322
- text, 323
- trasformazione dei sistemi di coordinate, 333
- trasparenza, 324
- trasposizione, 333
- tratteggio, 324
- modificatori di collisione, 335

- nodi di esportazione, 407
- nodi di input
 - cenni generali, 6
 - istanziamento di tipi, 58
 - nodo di input Database, 13
 - nodo di input Excel, 44
 - Nodo di input IBM Cognos BI, 39, 41
 - nodo di input SAS, 42
 - Nodo di input XML, 45
 - nodo Enterprise View, 7
 - nodo File Statistics, 434
 - nodo Input utente, 50–51
 - nodo Testo fisso, 24
 - nodo Testo variabile, 20
- nodi Grafici, 231
 - animazione, 232, 234
 - Distribuzione, 261
 - Grafico, 252
 - Histogram, 266
 - Lavagna grafica, 239
 - Plot multiplo, 275
 - Plot tempo, 291
 - Raccolta, 270
 - riquadri, 232, 234
 - sovrapposizioni, 232
 - Valutazione, 295
 - Web, 280
- nodi IBM SPSS Statistics, 433
- Nodi incapsulati, 456
- nodi Operazioni su campi, 96
 - generazione da un'esplorazione dati, 383
- nodi Operazioni su record, 60
 - nodo Intervalli di tempo, 204
- nodi output, 346, 353–354, 359, 364, 369, 389, 400, 403, 444
 - pubblica su Web, 348
 - scheda output, 355
- nodo Accodamento
 - cenni generali, 90
 - corrispondenza di campi, 91
 - impostazione delle opzioni, 91
 - impostazione di tag per i campi, 86
- nodo Aggregazione
 - cenni generali, 73
 - elaborazione parallela, 75
 - impostazione delle opzioni, 74
 - performance, 75
- nodo Aggregazione RFM
 - cenni generali, 75
 - discretizzazione indipendente, 76, 189
 - discretizzazione nidificata, 76, 189
 - impostazione delle opzioni, 77
- nodo Analisi, 364
 - scheda Analisi, 364
 - scheda output, 355
- nodo Analisi RFM
 - cenni generali, 189
 - discretizzazione dei valori, 192
 - discretizzazione indipendente, 76, 189
 - discretizzazione nidificata, 76, 189
 - impostazioni, 190
- nodo Anonimizza
 - cenni generali, 169
 - creazione di valori anonimizzati, 172
 - impostazione delle opzioni, 170
- nodo Bilanciamento
 - cenni generali, 71
 - generazione da grafici, 319
 - impostazione delle opzioni, 72
- nodo Campione
 - campioni casuali, 63–64
 - campioni non casuali, 63–64
 - campioni ponderati, 68
 - campioni raggruppati, 63–64, 68
 - campioni sistematici, 63–64
 - campioni stratificati, 63–64, 68–69
 - Dimensioni dei campioni per gli strati, 69
 - lista di campionamento, 63
- nodo Crea flag, 196
- nodo Cronologia, 226
 - cenni generali, 225
- nodo Database, 408
 - mappatura dei campi dati di origine alle colonne del database, 410
 - nome di tabella, 408
 - opzioni di unione, 410
 - scheda Esporta, 408
 - schema, 412
 - sorgente dati, 408
 - tabelle di indicizzazione, 414
- nodo di esportazione File piatto, 419
 - scheda Esporta, 419
- Nodo di esportazione IBM Cognos BI, 41, 422, 424

- nodo di input Database, 13
 - editor query, 19
 - query SQL, 14
 - selezione di tabelle e visualizzazioni, 18
- nodo di input Excel, 44
- Nodo di input IBM Cognos BI, 39, 41
- nodo di input Microsoft Excel, 44
- nodo di input SAS
 - file *.sd2* (SAS), 42
 - file *.ssd* (SAS), 42
 - file *.tpt* (SAS), 42
 - file di trasporto, 42
- Nodo di input XML, 45
- nodo Discretizza
 - anteprima dei bin, 188
 - bin a larghezza fissa, 180
 - bin con deviazione standard/media, 185
 - cenni generali, 178
 - conteggio uguale, 181
 - impostazione delle opzioni, 179
 - ottimale, 186
 - ranghi, 184
 - somme uguali, 181
- nodo Distribuzione , 261
 - scheda Aspetto, 263
 - scheda Plot, 262
 - utilizzo del grafico, 264
 - utilizzo della tabella, 264
- nodo Elimina duplicati
 - cenni generali, 92
 - impostazioni di ottimizzazione, 94
 - ordinamento record, 93
- nodo Enterprise View, 7
- nodo Esplora, 369
 - scheda impostazioni, 370
 - scheda output, 355
- nodo Esporta IBM SPSS Data Collection, 420
- nodo Esporta SAS, 426
- nodo Esporta Statistics, 449
 - scheda Esporta, 450
- nodo Esporta XML, 428
- nodo Esportazione da Excel, 427
- nodo File cache, 434
- nodo File Statistics, 434
- nodo Filtro
 - cenni generali, 143
 - impostazione delle opzioni, 144
 - insiemi a risposta multipla, 147
- nodo Globali, 403
 - scheda impostazioni, 404
- nodo Importazione da Excel
 - generazione da output, 427
- nodo Input Data Collection, 31
 - file di metadati, 32
 - file di registro, 32
- nodo Input IBM SPSS Data Collection, 38
 - impostazioni delle connessioni al database, 36–37
 - insiemi a risposta multipla, 37
 - language, 35
 - tipi di etichette, 35
- nodo Input utente
 - cenni generali, 50
 - impostazione delle opzioni, 51
- nodo Intervalli di tempo, 205, 207, 209
 - cenni generali, 204
- nodo Istogramma , 266
 - scheda Aspetto, 269
 - scheda Plot, 267–268
 - utilizzo del grafico, 269
- nodo Lavagna grafica , 239
 - scheda Aspetto, 247
- nodo Matrice, 359
 - browser di output, 362
 - evidenziazione, 360
 - ordinamento di righe e colonne, 360
 - percentuali di colonna, 360
 - percentuali di riga, 360
 - scheda Aspetto, 360
 - scheda impostazioni, 359
 - scheda output, 355
 - tavole di contingenza, 360
- nodo Medie, 393
 - browser di output, 397–398
 - campi accoppiati, 395
 - gruppi indipendenti, 394
 - importanza, 396
 - scheda output, 355
- nodo Nuovo campo
 - cenni generali, 154
 - condizionali, 165
 - conteggio, 164
 - conversione della classe di archiviazione dei campi, 165
 - derivazione multipla, 156
 - flag, 159
 - formula, 158
 - generazione da grafici, 319
 - generazione da un nodo Discretizza, 188
 - generazione dai bin, 178
 - generazione dai collegamenti di un grafico Web, 287
 - generazione mediante la preparazione automatica dati, 123
 - impostazione delle opzioni, 155
 - ricodifica di valori, 165
 - set, 161
 - stato, 162
- nodo Ordina
 - cenni generali, 78
 - impostazioni di ottimizzazione, 79
- nodo Output IBM SPSS Statistics
 - scheda Output, 448
- nodo Output Statistics, 444
 - scheda Sintassi, 445
- nodo Partizione, 193–194

- nodo Plot, 252
 - scheda Aspetto, 259
 - scheda Opzioni, 258
 - scheda Plot, 255
 - utilizzo del grafico, 260
- nodo Plot multiplo , 275
 - scheda Aspetto, 278
 - scheda Plot, 276
 - utilizzo del grafico, 279
- nodo Plot tempo, 291
 - scheda Aspetto, 293
 - scheda Plot, 292
 - utilizzo del grafico, 294
- Nodo Preparazione automatica dei dati, 98
- nodo Raccolta , 270
 - scheda Aspetto, 273
 - scheda Opzioni, 271–272
 - utilizzo del grafico, 274
- nodo Report, 400
 - scheda Modello, 401
 - scheda output, 355
- nodo Ricodifica, 174, 177
 - cenni generali, 174, 178
 - generazione da un nodo Distribuzione, 264
- nodo Riempimento
 - cenni generali, 166
- nodo Riordina campi, 227
 - impostazione delle opzioni, 227
 - ordinamento automatico, 229
 - ordinamento personalizzato, 227
- nodo Riorganizza, 197, 199
 - con il nodo Aggregazione, 199
- nodo Risultati classificatore binario
 - campi di output, 150
 - combinazione di punteggi, 150
- nodo Seleziona
 - cenni generali, 61
 - generazione da grafici, 319
 - generazione dai collegamenti di un grafico Web, 287
- nodo Statistiche, 389
 - correlazioni, 389
 - etichette di correlazione, 390
 - scheda impostazioni, 389
 - scheda output, 355
 - statistics, 389
- nodo Tabella, 353
 - giustificazione di colonne, 141
 - impostazioni di output, 354
 - larghezza di colonna, 141
 - scheda Formato, 141
 - scheda impostazioni, 354
 - scheda output, 355
- nodo Testo fisso
 - cenni generali, 24
 - impostazione delle opzioni, 24
 - riconoscimento automatico data, 27
- nodo Testo variabile, 20
 - impostazione delle opzioni, 22
 - riconoscimento automatico data, 23
- nodo Tipo
 - cancellare valori, 56
 - cenni generali, 125
 - copia dei tipi, 140
 - dati continui, 134
 - dati nominali, 135
 - dati ordinali, 135
 - gestione dei vuoti, 132
 - giustificazione di colonne, 141
 - impostazione del ruolo di modellazione, 138
 - impostazione delle opzioni, 127, 129
 - larghezza di colonna, 141
 - scheda Formato, 141
 - tipo di campo flag, 136
- nodo Trasformazioni, 383
- nodo Trasformazioni Statistics, 435
 - impostazione delle opzioni, 436
 - scheda Sintassi, 436
 - sintassi consentita, 438
- nodo Trasponi, 200
 - campi numerici, 200
 - campi stringa, 200
 - nomi di campo, 200
- nodo Unione, 81
 - cenni generali, 80
 - filtro di campi, 85
 - impostazione delle opzioni, 83
 - impostazione di tag per i campi, 86
 - impostazioni di ottimizzazione, 88
- nodo Valutazione , 295
 - condizione risultato, 302
 - espressione punteggio, 302
 - lettura dei risultati, 306
 - regola aziendale, 302
 - scheda Aspetto, 304
 - scheda Opzioni, 302
 - scheda Plot, 301
 - utilizzo del grafico, 308
- nodo Web , 280
 - definizione di collegamenti, 283
 - dispositivo di scorrimento, 287
 - dispositivo di scorrimento dei collegamenti, 287
 - modifica del layout, 287
 - modifica delle soglie, 288
 - modifica di punti, 287
 - riepilogo Web, 290
 - scheda Aspetto, 285
 - scheda Opzioni, 283
 - scheda Plot, 281
 - utilizzo del grafico, 286
- nomi di campo, 146
 - anonimizzazione, 147
 - esportazione di dati, 408, 419, 426, 450

- nomi di variabili
 - esportazione di dati, 408, 419, 426, 450
- normalizza obiettivo continuo, 106, 123
- normalizzazione di valori
 - nodi Grafici, 276, 293
- note legali, 474
- nulli, 132, 379
 - nelle tabelle Matrice, 360
- numero intero, formato di archiviazione, 28, 54
- numero reale, formato di archiviazione, 28, 54

- ODBC
 - caricamento di massa tramite, 417
 - connessione per il nodo di esportazione IBM Cognos BI, 424
 - nodo di input Database, 13
- ODBC, nodo di esportazione. *Vedere* nodo di esportazione Database, 408
- oggetti di output, 347
- opzioni
 - IBM SPSS Statistics, 405
- opzioni dei ranghi, 184
- opzioni di unione, esportazione database, 410
- opzioni modello
 - nodo Modello Statistics, 441
- Oracle, 13
- ordinamento
 - campi, 227
 - campi preordinati, 79, 94
 - nodo Elimina duplicati, 93
 - record, 78
- ordinamento di dati, 78, 227
- ordine crescente, 78
- ordine decrescente, 78
- ordine dei dati di input, 86
- ordine di esecuzione
 - specifico, 471
- ordine naturale
 - modifica, 227
- output
 - esportazione, 351
 - generazione di nuovi nodi da, 348
 - HTML, 351
 - salvataggio, 348
 - stampa, 348
- output del nodo Matrice
 - salvataggio come testo, 355
- output HTML
 - nodo Report, 402
 - visualizzazione in un browser, 351
- output tabulare
 - riordinamento delle colonne, 352
 - selezione di celle, 352
- output XML
 - nodo Report, 402

- palette
 - nascondere, 323
 - spostamento, 323
 - visualizzazione, 323
- parametri
 - impostazione per Supernodi, 466
 - proprietà dei nodi, 468
 - Supernodi, 466, 468
- parametri Supernodo, 466, 468
- pari merito
 - nodo Discretizza, 181
- parola chiave FILLFACTOR
 - indicizzazione delle tabelle del database, 416
- parola chiave UNIQUE
 - indicizzazione delle tabelle del database, 415
- partizionamento dei dati, 193–194
 - grafici di valutazione, 302
 - nodo Analisi, 364
- performance
 - dati di campionamento, 63
 - nodi Discretizza, 188
 - nodi Nuovo campo, 188
 - nodo Aggregazione, 75
 - ordinamento, 80
 - unione, 88
- period, 141
- periodi
 - nodo Intervalli di tempo, 213
- periodi ciclici
 - nodo Intervalli di tempo, 214
- periodicità
 - dati di serie storiche, 204
- periodo di stima, 209
- pesi
 - grafici di valutazione, 302
- più campi
 - selezione, 158
- preparazione automatica dati
 - analisi dei campi, 113
 - collegamenti tra visualizzazioni, 111
 - creazione, 107
 - dettagli campo, 118
 - dettagli dell'azione, 120
 - escludi campi, 104
 - esclusione dei campi inutilizzati, 102
 - fields, 101
 - generazione di un nodo Nuovo campo, 123
 - impostazioni campo, 102
 - input, preparazione, 105
 - nomina campi, 109
 - normalizza obiettivo continuo, 106, 123
 - obiettivi, 98
 - obiettivi, preparazione, 105
 - potere predittivo, 116
 - prepara date e ore, 103
 - preparazione degli input, 105
 - preparazione degli obiettivi, 105

- reimposta visualizzazioni, 111
- riepilogo delle azioni, 115
- riepilogo elaborazione campi, 112
- selezione delle funzioni, 107
- tabella campi, 117
- vista del modello, 110
- proprietà
 - nodo, 468
 - per campi, 141
- proprietà dei nodi, 468
- pubblica su Web, 348
- punteggi di propensione
 - bilanciamento dei dati, 72
- punteggi di propensione regolata
 - bilanciamento dei dati, 72
- punteggio
 - opzioni relative al nodo Valutazione, 302
- punti di divisione
 - nodo Discretizza, 178
- Python
 - script di caricamento di massa, 417
- qualità dei dati
 - Browser del nodo Esplora, 378
- query
 - nodo di input Database, 13–14
- query SQL
 - nodo di input Database, 13–14, 19
- raggruppamento dei valori, 264
- ranghi frazionari, 184
- rappresentazione delle associazioni, 280
- recency
 - impostazione della data relativa, 77
- record
 - conteggi, 74
 - etichette, 138
 - length, 24
 - trasposizione, 200
 - unione, 80
- record di controllo
 - modelli di serie storica, 209
- record incompleti, 83
- record univoci, 92
- regola aziendale
 - opzioni relative al nodo Valutazione, 302
- regressione di minimi quadrati ponderati localmente
 - nodo Plot, 256
- report
 - salvataggio dell'output, 355
- report sulla qualità
 - Browser del nodo Esplora, 378
- residui
 - nodo Matrice, 360
- ricerca
 - browser del nodo Tabella, 357
- ricodifica, 174, 178
- ricodifica automatica, 174
- riconoscimento automatico data, 23, 27
- riconoscimento date, 23, 27
- ridenominazione
 - campi per l'esportazione, 451
 - file delle mappe, 251
 - fogli di stile di visualizzazione, 251
 - modelli di visualizzazione, 251
- ridenominazione degli oggetti di output, 347
- righe vuote
 - file di Excel, 44
- riordinamento delle colonne
 - browser del nodo Tabella, 352, 357
- ristrutturazione dei dati, 197
- risultati
 - opzioni relative al nodo Valutazione, 302
- ROI
 - grafici, 295, 306
- rotazione di grafici 3-D, 238
- ruoli
 - specifici di campi, 56, 125, 138
- ruoli di modellazione
 - specifici di campi, 56, 125, 138
- salvataggio
 - oggetti di output, 347, 355
 - output, 348
- SAS
 - impostazione delle opzioni di importazione, 43
- sblocco di Supernodi, 462
- scarto
 - campi, 143
- scenario, 7
- scheda della sintassi
 - nodo Output Statistics, 445
- schema
 - nodo Database, 412
- script
 - Supernodi, 471
- selezione di righe (casi), 61
- selezione di valori, 310, 314, 317
- serie storiche, 225
- sezioni nei grafici, 310
- sfaccettatura
 - nelle visualizzazioni, 234
- significatività
 - intensità delle correlazioni, 390
- simbolo decimale, 22–23, 141
 - formati di visualizzazione numeri, 143
 - nodo di esportazione File piatto, 419
- simbolo di raggruppamento
 - formati di visualizzazione numeri, 143
- Sintassi XPath, 45
- sistemi di coordinate
 - trasformazione, 333
- smoother
 - nodo Plot, 256

- soglie
 - visualizzazione di soglie di bin, 188
- somma
 - nodo Globali, 404
 - output del nodo Statistiche, 391
- sorgenti dati
 - connessioni database, 16
- sostituzione di valori di campo, 166
- sovrapposti, 338
- sovrapposizione di colori nei grafici, 232
- sovrapposizione di dimensioni nei grafici, 232
- sovrapposizione di forme nei grafici, 232
- sovrapposizione di riquadri nei grafici, 232, 234
- sovrapposizioni per i grafici, 232
- sovrascrittura di tabelle di database, 408
- SPSS Modeler Server, 1
- stampa dell'output, 348
- statistica di valutazione delle prestazioni, 364
- statistica *F*
 - nodo Medie, 398
- statistiche riassuntive
 - nodo Esplora, 369
- statistics
 - descrizioni, 241, 336
 - modifiche in visualizzazioni, 335
 - nodo Esplora, 369
 - nodo Matrice, 359
- storage, 132
 - conversione, 165–166, 168
- stringa, formato di archiviazione, 28, 54
- Supernodi, 453
 - blocco, 461–462
 - caricamento, 472
 - creazione, 456
 - creazione di cache per, 470
 - impostazione dei parametri, 466
 - modifica, 464
 - nidificazione, 458
 - protezione password, 461–463
 - salvataggio, 472
 - sblocco, 462
 - script, 471
 - Supernodi di elaborazione, 454
 - Supernodi di input, 454
 - Supernodi terminali, 455
 - tipi di, 453
 - uso di commenti con, 465
 - zoom avanti, 464
- tabelle
 - join, 81
 - salvataggio come testo, 355
 - salvataggio dell'output, 355
- tag, 80, 86
- tavole di contingenza
 - nodo Matrice, 359–360
- tempo, formato di archiviazione, 28, 54
- test *t*
 - campioni accoppiati, 395
 - campioni indipendenti, 394
 - nodo Medie, 394–395, 399
- testo
 - codifica, 23, 26, 420
- text
 - dati, 20, 24
 - delimitati, 20
- time
 - impostazione di formati, 141
- timestamp, 127
- timestamp, formato di archiviazione, 28, 54
- tipi di campo, 56, 125
 - nelle visualizzazioni, 243
- tipi di dati, 24, 56, 96, 125, 127
 - istanziamento, 130
- tipi di etichette
 - nodo Input IBM SPSS Data Collection, 35
- tipi di variabili
 - nelle visualizzazioni, 243
- tipo di utilizzo, 27, 127
- tipo flag, 127, 136
- tipo insieme, 127
- trasformazioni
 - reclassify, 174, 178
 - ricodifica, 174, 178
- trasparenza
 - nelle visualizzazioni, 233
- trasparenza nei grafici, 232
- trasposizione di dati, 200
- troncamento dei nomi di campo, 144, 146
- type, 27
- un errore standard della media
 - output del nodo Statistiche, 391
- unione in base all'ordine, 80
- utilizzo di riquadri
 - nelle visualizzazioni, 234
- valore chiave per aggregazione, 74
- valore di seme casuale
 - campionamento di record, 67, 195
- valore massimo per aggregazione, 74
- valore medio per aggregazione, 74
- valore medio per i record, 73
- valore minimo per aggregazione, 74
- valore *p*
 - importanza, 396
- valore seme
 - campionamento e record, 67, 195
- valori
 - etichette di campi e valori, 56, 125, 132
 - lettura, 131
 - specificata, 132
- valori attesi
 - nodo Matrice, 360

- valori falsi, 136
- valori globali, 403
- valori mancanti, 96, 132, 137
 - gestione, 379
 - in nodi Aggregazione, 73
 - nelle tabelle Matrice, 360
 - riempimento, 379
- valori mancanti definiti dall'utente, 379
 - nelle tabelle Matrice, 360
- valori mancanti di sistema, 379
 - nelle tabelle Matrice, 360
- valori non definiti, 83
- valori nulli
 - dati misti, 29, 54
 - nelle tabelle Matrice, 360
- valori preimpostati, connessione a database, 17
- valori sommati, 74
- valori veri, 136
- valori vuoti
 - nelle tabelle Matrice, 360
- valutazione dei modelli, 364
- valutazione del modello, 295
- Variabili codici
 - nodo Input IBM SPSS Data Collection, 35
- Variabili di sistema
 - nodo Input IBM SPSS Data Collection, 35
- variabili SourceFile
 - nodo Input IBM SPSS Data Collection, 35
- varianza
 - output del nodo Statistiche, 391
- virgolette
 - importazione di file di testo, 24
 - per esportazione in database, 408
- vista del modello
 - nella preparazione automatica dati, 110
- visualizzazione
 - grafici e diagrammi, 231
 - output HTML in un browser, 351
- visualizzazioni
 - assi, 329
 - categorie, 331
 - colori e motivi, 324
 - copia, 338
 - forma dei punti, 326
 - formati dei numeri, 328
 - margini, 327
 - modalità di modifica, 321
 - modifica, 321
 - posizione della legenda, 338
 - proporzioni dei punti, 326
 - riempimento, 327
 - riquadri, 331, 333
 - rotazione dei punti, 326
 - scale, 329
 - text, 323
 - trasformazione dei sistemi di coordinate, 333
 - trasparenza, 324
 - trasposizione, 331, 333
 - tratteggi, 324
 - vuoti, 379
 - nelle tabelle Matrice, 360
- Zoom indietro da un Supernodo, 464