



SAS Publishing



What's New in SAS[®] Enterprise Miner 5.2

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2005. What's New in SAS® Enterprise Miner 5.2. Cary, NC: SAS Institute Inc.

What's New in SAS® Enterprise Miner 5.2

Copyright © 2005, SAS Institute Inc., Cary, NC, USA

All rights reserved. Produced in the United States of America.

For a Web download or e-book: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

U.S. Government Restricted Rights Notice: Use, duplication, or disclosure of this software and related documentation by the U.S. government is subject to the Agreement with SAS Institute and the restrictions set forth in FAR 52.227-19, Commercial Computer Software-Restricted Rights (June 1987).

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, November 2005

SAS Publishing provides a complete selection of books and electronic products to help customers use SAS software to its fullest potential. For more information about our e-books, e-learning products, CDs, and hard-copy books, visit the SAS Publishing Web site at support.sas.com/pubs or call 1-800-727-3228.

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

What's New in SAS Enterprise Miner 5.2

- [Introduction](#)
- [Administration and Configuration](#)
- [New Nodes in SAS Enterprise Miner 5.2](#)
 - [Decisions Node](#)
 - [Replacement Node](#)
 - [SOM/Kohonen Node](#)
- [Enhanced Nodes in SAS Enterprise Miner 5.2](#)
 - [Decision Tree Node](#)
 - [Filter Node](#)
 - [Merge Node](#)
 - [Model Comparison Node](#)
 - [Path Node](#)
 - [Principal Components Node](#)
 - [Regression Node](#)
 - [SAS Code Node](#)
 - [Sample Node](#)
 - [Transform Variables Node](#)
- [New Model Assessment Reports in SAS Enterprise Miner 5.2](#)
- [Other Changes and Enhancements](#)
- [Credit Scoring Nodes](#)

[Introduction](#)

SAS Enterprise Miner 5.2 significantly expands both the feature set and architecture of Enterprise Miner.

The feature set is expanded by the following:

- new nodes for data preparation and statistical exploration
- new and faster two-dimensional and three-dimensional graphics for exploring relationships within data

- improved and additional reports for predictive models
- improved SAS score code
- new algorithms and graphical tools for Web path analysis
- sampling for rare case analysis

The architecture is enhanced by use of the SAS Analytics Platform. Using this platform has the following benefits:

- common client/server design and implementation that is shared with other SAS analytical software and delivery of improved thin client distribution, as well as improved server and user monitoring tools
- better integration with the SAS Business Intelligence Platform

The following sections detail the changes and enhancements.

Administration and Configuration

The following changes and enhancements improve the installation, administration, and configuration of SAS Enterprise Miner 5.2.

SAS Analytics Platform Integration

The SAS Analytics Platform provides a common client/server architecture and implementation for a family of products that includes SAS Enterprise Miner, SAS Forecast Studio, and SAS Inventory Studio. One instance of the Analytics Platform middle-tier server can serve all three applications. It is easier for SAS administrators to install and configure multiple SAS analytical applications.

In three-tier environments, administrators can monitor the status of the Analytics Platform server through both Web-based and client-based tools. Users can download and install the Enterprise Miner client directly from the Analytics Platform server. Administrators who configure multi-user environments must manually configure and start the Analytics Platform middle-tier services.

The SAS Analytics Platform cannot be licensed separately. The Analytics Platform installation is triggered by the installation of any of the existing SAS Analytics Platform products.

Access through Firewall

In previous releases when you were running the application as a thin client, you could not cross a firewall to access the server. SAS Enterprise Miner 5.2 now provides this functionality.

Log On Screen

As a result of the integration with the SAS Analytics Platform, SAS Enterprise Miner uses a new logon screen that is shared with SAS Forecast Studio and SAS Inventory Studio. You are no longer prompted for the SAS Metadata Server authentication domain when logging on to a server. Support has also been added for anonymous logons.

SAS Analytics Platform

SAS

SAS® Enterprise Miner™ 5.2

The Power to Know.

Copyright © 2002-2005 by SAS Institute Inc., Cary, NC, USA. All Rights Reserved.

User name:

Password:

☐ Remember my password

Server:

Log On Cancel

SAS Management Console Plug-In for Enterprise Miner Administration

Some of the functionality that was in the configuration wizard in SAS Enterprise Miner 5.1 has been moved to a SAS Management Console plug-in. The SAS Management Console plug-in gives administrators better control over Enterprise Miner configurations. The following information is now stored in the SAS Management Console plug-in:

- Server start-up code.
- Default project path and security.
- Maximum concurrent nodes per process execution.
- Model group definitions.
- WebDAV location is now stored in the SAS Metadata Repository and configured by SAS Management Console. As a result, users never need to know about the WebDAV location that is used by the Model Manager functions. WebDAV is Enterprise Miner's Web-based Distributed Authoring and Versioning. It is a set of extensions to the HTTP protocol that enable you to collaboratively edit and manage files on remote Web servers.

Grid Processing

Grid processing is available for enterprises that need to perform computing over multiple logical or physical systems. The execution of the process flow diagram in SAS Enterprise Miner is sent to a load balancing manager that distributes the jobs to a grid of systems. This is expected to benefit users who

run multiple, large-process flow diagrams, or users who manage a large multi-user environment.

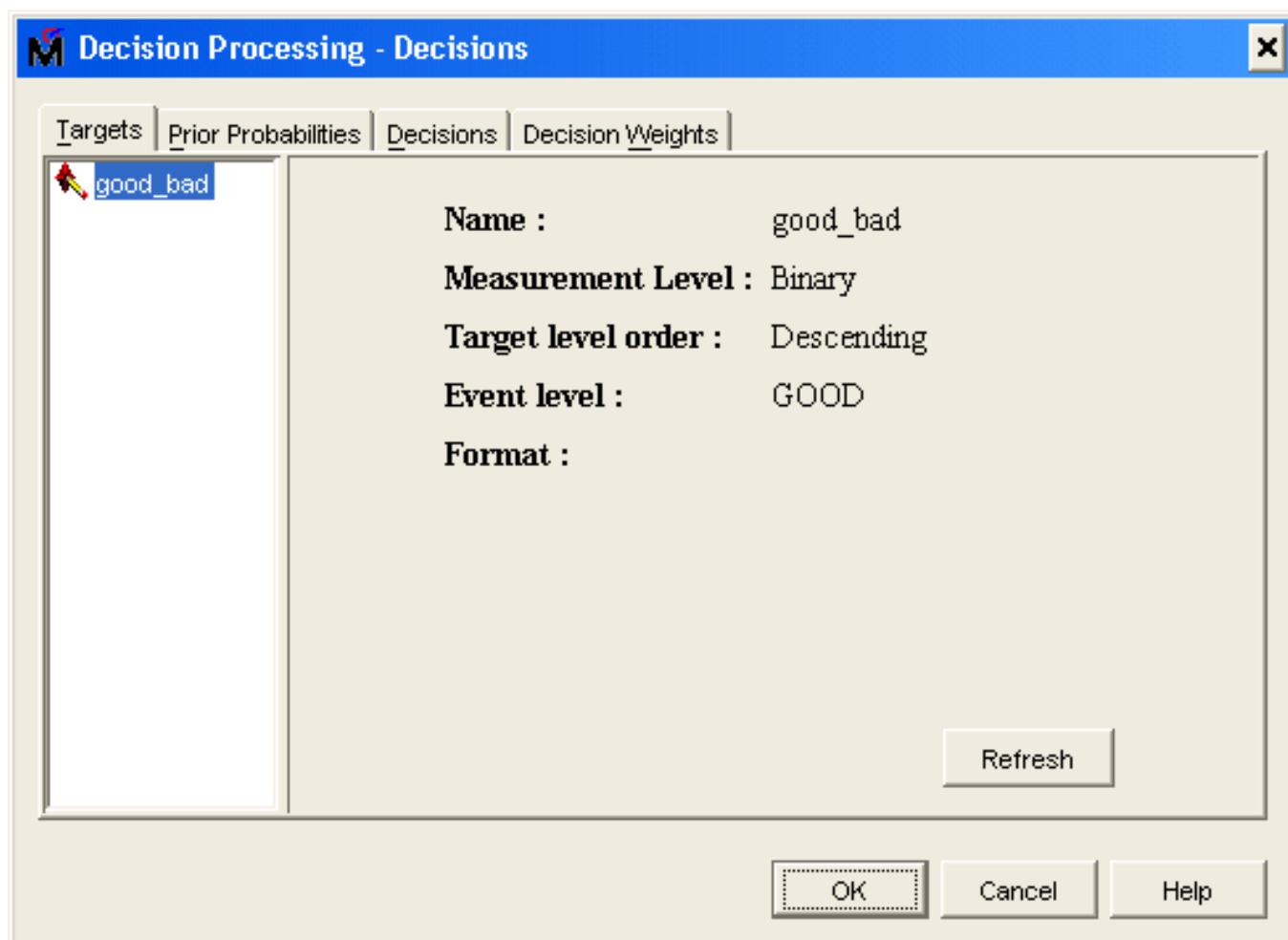
New Nodes in SAS Enterprise Miner 5.2

SAS Enterprise Miner 5.2 contains the following new nodes that add functionality to the environment.

Decisions Node

A Decisions node has been added to the SAS Enterprise Miner 5.2 tools menu. The Decisions node belongs to the Assess category of the SAS SEMMA (Sample, Explore, Modify, Model, Assess) data mining process. You use the Decisions node to define or modify target profiles for a target that produces optimal decisions. The decisions are made using a user-specified decision matrix and the output from a subsequent modeling procedure.

The Decisions node contains a tabbed Decision Processing window that you use to configure your decision matrix. The Decision Processing window shares the same layout as the Data Source Wizard — Decision Configuration window.



Decision matrices are no longer automatically built when you open the Decision Editor. Click the new **Build** button to create the decision matrix for the desired target.

After a decision matrix has been created, you can refresh the default decision by clicking the **Refresh** button. You might want to refresh the data when the target levels in the underlying data have changed or when the target order has changed.

See Layout of a Target Profile in the Help for the Enterprise Miner Target Profiler for more information about configuring your decision matrix.

Replacement Node

A Replacement node has been added to the SAS Enterprise Miner 5.2 tools menu. The Replacement node belongs to the Modify category of the SAS SEMMA (Sample, Explore, Modify, Model, Assess) data mining process. You use the SAS Enterprise Miner 5.2 Replacement node to generate score code to process unknown levels when scoring and also to interactively specify replacement values for class levels. The Replacement node must follow a node that exports a data set, such as a Data Source, Sample, or Data Partition node.

You can use the Replacement Editor to specify the replacement values. The Replacement Editor lists the class levels for the input and target variables in the training data set.

The Replacement Editor dialog box contains a table with the following data:

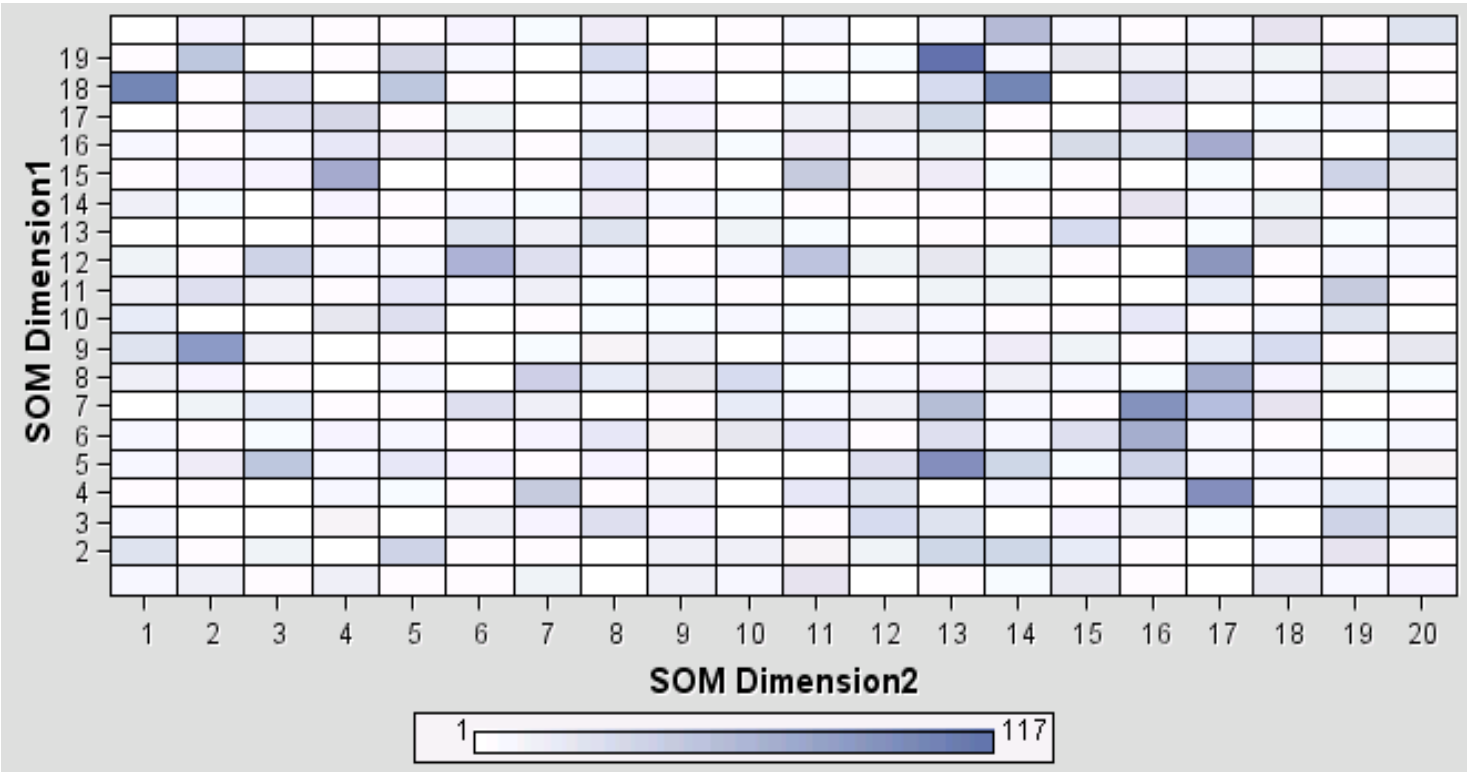
Variable	Level	Frequency	Type	Char Raw value	Num Raw Value	Replacement Value
BAD	0	4771	N		0.0	
BAD	1	1189	N		1.0	
BAD	_UNKNOWN_	.	N		.	_DEFAULT_
JOB	Other	2388	C	Other	.	
JOB	ProfExe	1276	C	ProfExe	.	
JOB	Office	948	C	Office	.	
JOB	Mgr	767	C	Mgr	.	
JOB		279	C		.	
JOB	Self	193	C	Self	.	
JOB	Sales	109	C	Sales	.	
JOB	_UNKNOWN_	.	C		.	_DEFAULT_
REASON	DebtCon	3928	C	DebtCon	.	
REASON	Homelmp	1780	C	Homelmp	.	
REASON		252	C		.	
REASON	_UNKNOWN_	.	C		.	_DEFAULT_

At the bottom of the dialog box are four buttons: **Reset**, **OK**, **Cancel**, and **Help**.

SOM/Kohonen Node

A SOM/Kohonen node has been added to the SAS Enterprise Miner 5.2 tools menu. The SOM/Kohonen node belongs to the Explore category of the SAS SEMMA (Sample, Explore, Modify, Model, Assess) data mining process. You use the SOM/Kohonen node to perform unsupervised learning by using Kohonen vector quantization (VQ), Kohonen self-organizing maps (SOMs), or batch SOMs with

Nadaraya-Watson or local-linear smoothing. Kohonen VQ is a clustering method, whereas SOMs are primarily dimension-reduction methods.

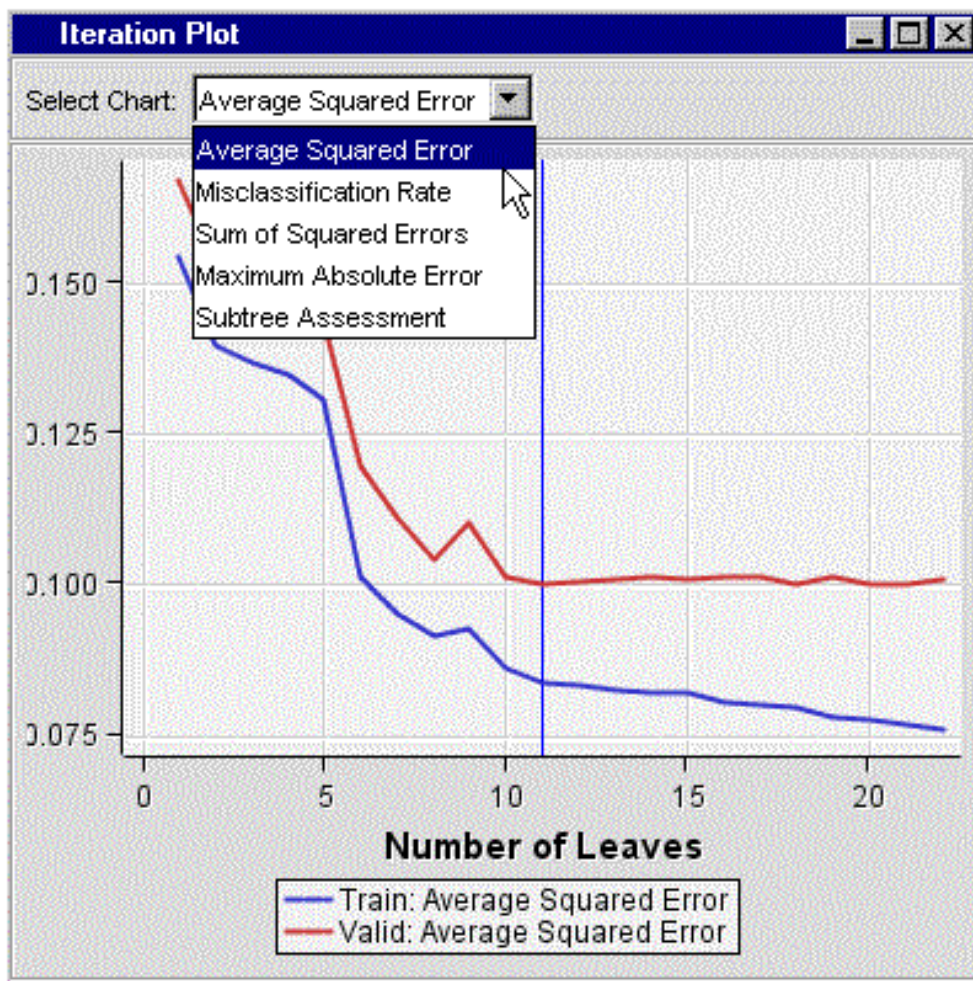


[Enhanced Nodes in SAS Enterprise Miner 5.2](#)

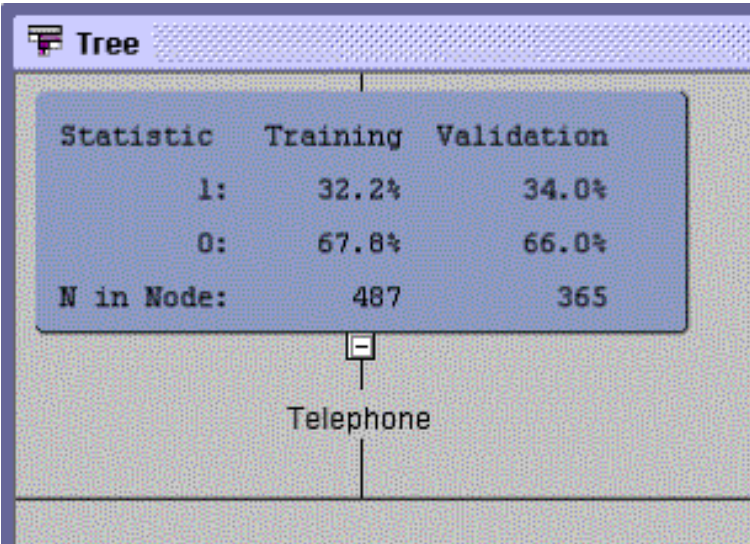
The following nodes contain specific and significant enhancements in functionality.

[Decision Tree Node](#)

Tree growth iteration plots have been added to the Decision Tree node results. A reference line in the plot indicates which subtree was selected.



Another enhancement to the Decision Tree output graphics is that Tree node plots now display information for all target levels.

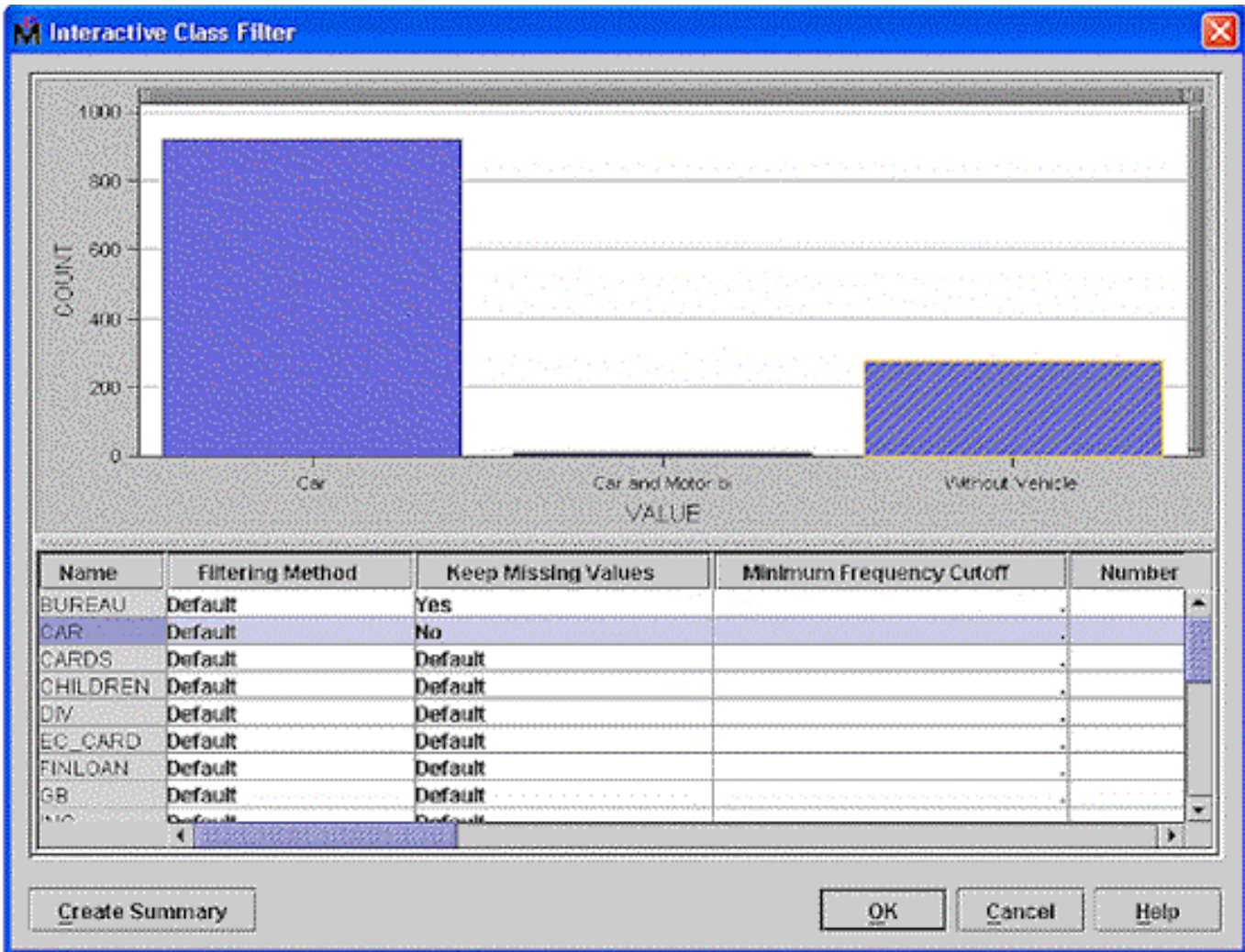


Using the graph properties for the tree diagram, you can toggle between the detailed text (which is shown) and the summarized text.

The upper limit that existed for the Split Size and Node Sample properties has been removed.

Filter Node

The Filter node is enhanced to support interactive user selection of values to filter for both continuous and categorical variables.



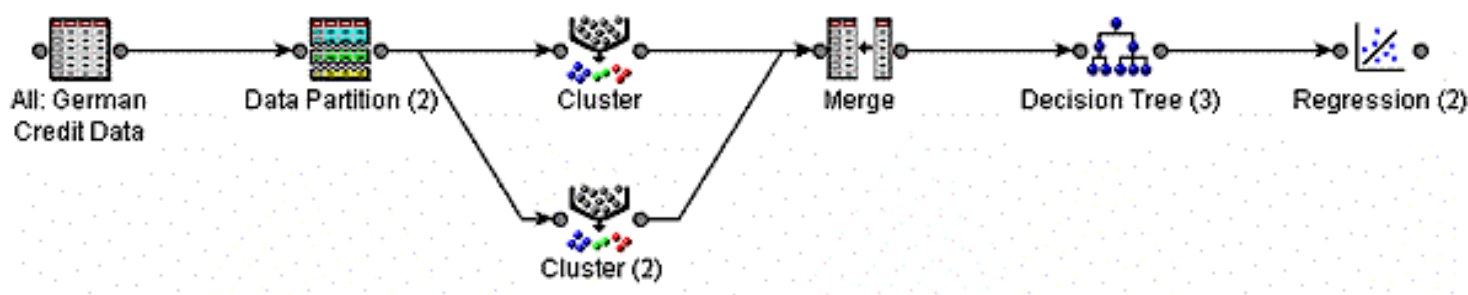
In the Filter Node General Properties sheet, you can now do the following:

- Modify the variables properties for the class and interval variables in two separate editors. To open the editors, click in the row for the Class Variables or the Interval Variables.
- Specify a new value for the Export Table property. The **All** value has been added to the Export Table property. When you select this value, no observations are removed. Instead, an indicator variable M_FILTER is created. The M_FILTER variable identifies the observations that would have been removed.

Property	Value
Node ID	Filter
Imported Data	...
Exported Data	...
Export Table	Filtered
Tables to Filter	Training Data
Create score code	Yes
<input checked="" type="checkbox"/> Class Variables	
Class Variables	...
Default Filtering Method	Rare Values (Percenta
Keep Missing Values	Yes
Normalized Values	Yes
Minimum Frequency Cu1	
Minimum Cutoff for Perc	0.01
Maximum Number of Le25	
<input checked="" type="checkbox"/> Interval Variables	
Interval Variables	...
Default Filtering Method	Standard Deviations fro
Keep Missing Values	Yes

Merge Node

The code for renaming and replacing variables has been revised to accommodate processes such as the one shown here. In this case, both segment variables that were created by the two different Cluster nodes will be used as inputs to the model nodes, and the creator fields of the prediction variables will contain the value REG2 instead of TREE3.



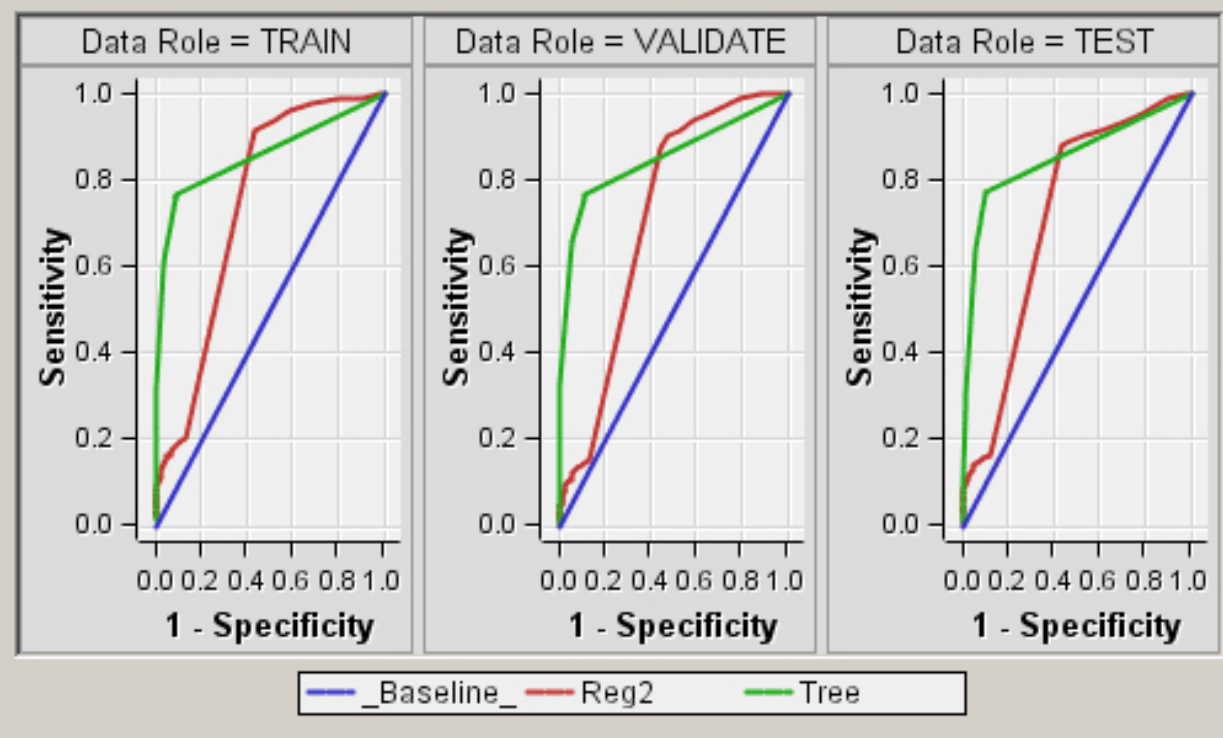
Model Comparison Node

Changes have been made to the predictive model comparison code to improve efficiency and consistency between reports. The Model Comparison node features a new Bin-based Kolmogorov-Smirnov statistic in the results tables.

The score code now creates the bin variable with a prefix of b_.

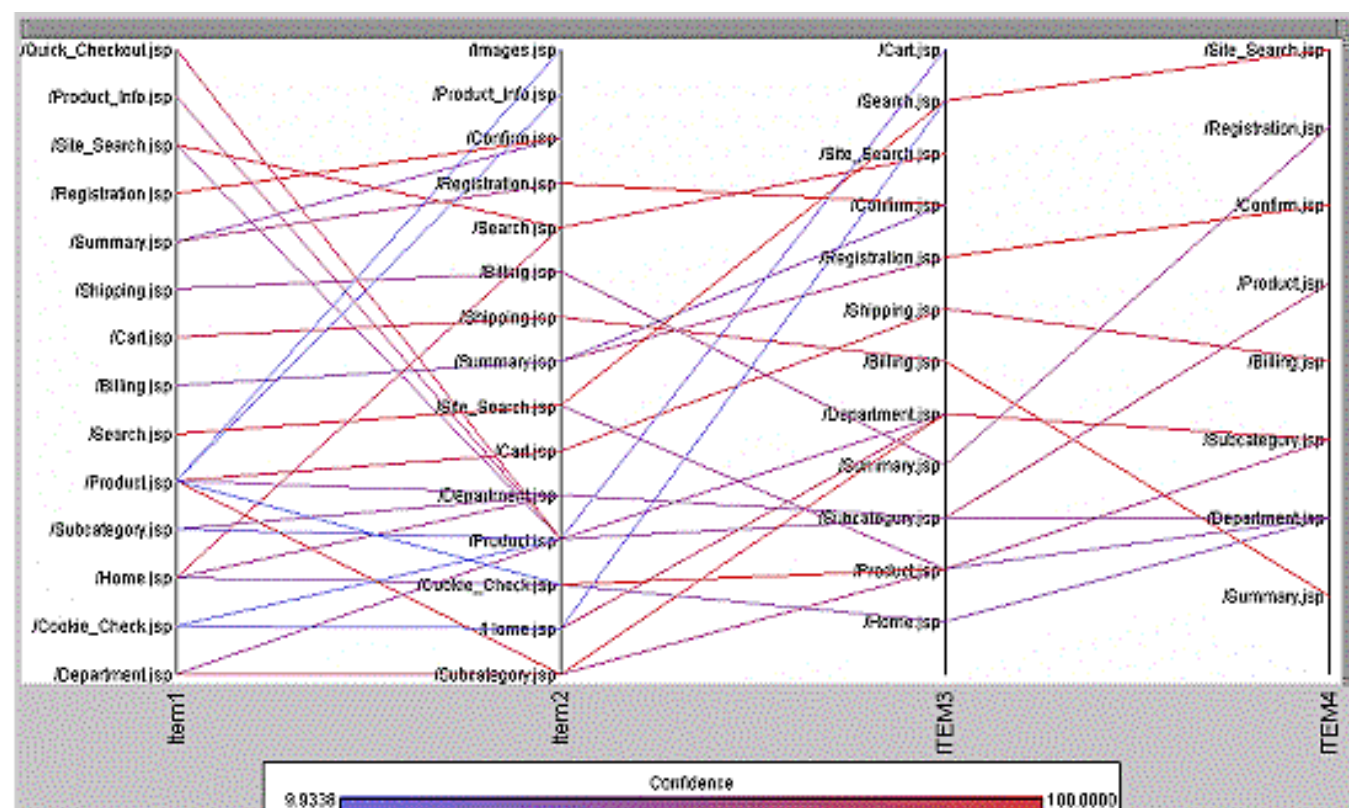
A baseline has been added to the Receiver Operating Characteristic (ROC) chart that is produced for binary targets.

ROC Chart : BAD



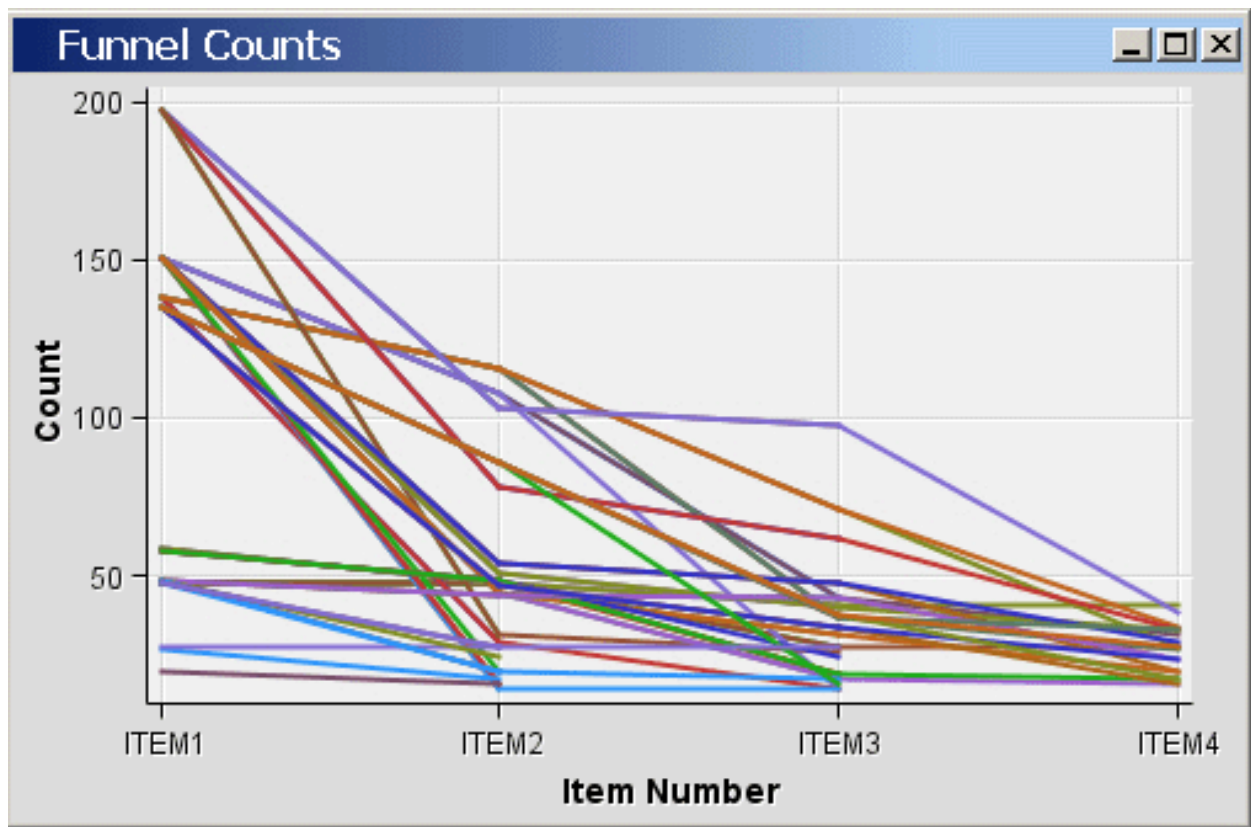
Path Node

The Path node has significant changes that reflect new capabilities in PROC PATH. Several new properties identify pattern-matching data sets. The scalability is improved. The Path node scoring has been greatly accelerated. New plots have been introduced to the Path node results, including the following view of rules sequences.



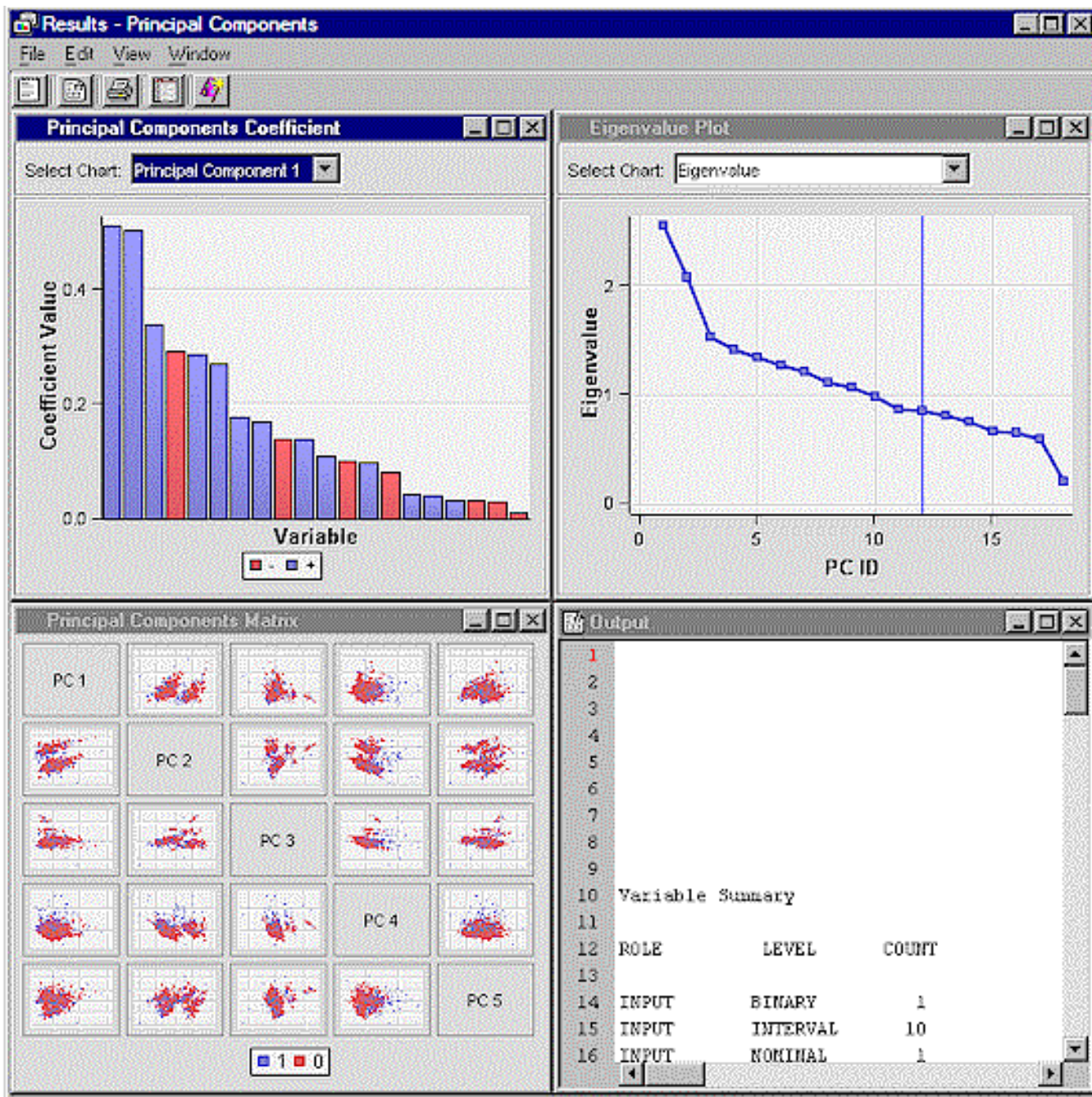
In the Path Node General Properties sheet, the new **Training Mode** property enables you to choose between Train and Score modes when the Path Analysis node runs. When the mode is set to Train, the node discovers a set of rules. When the mode is set to score, a previously found set of rules is applied to the transaction data set. You must specify a scoring rules data set when the Training Mode property is set to Score. The default setting for the Training Mode property is Train.

The new Funnel Counts plot shows the drop-off in the number of visitors along a particular path of interest. It can be useful to see how visitor attrition occurs along a path, indicating points of interest such as where the biggest drop-off points are.

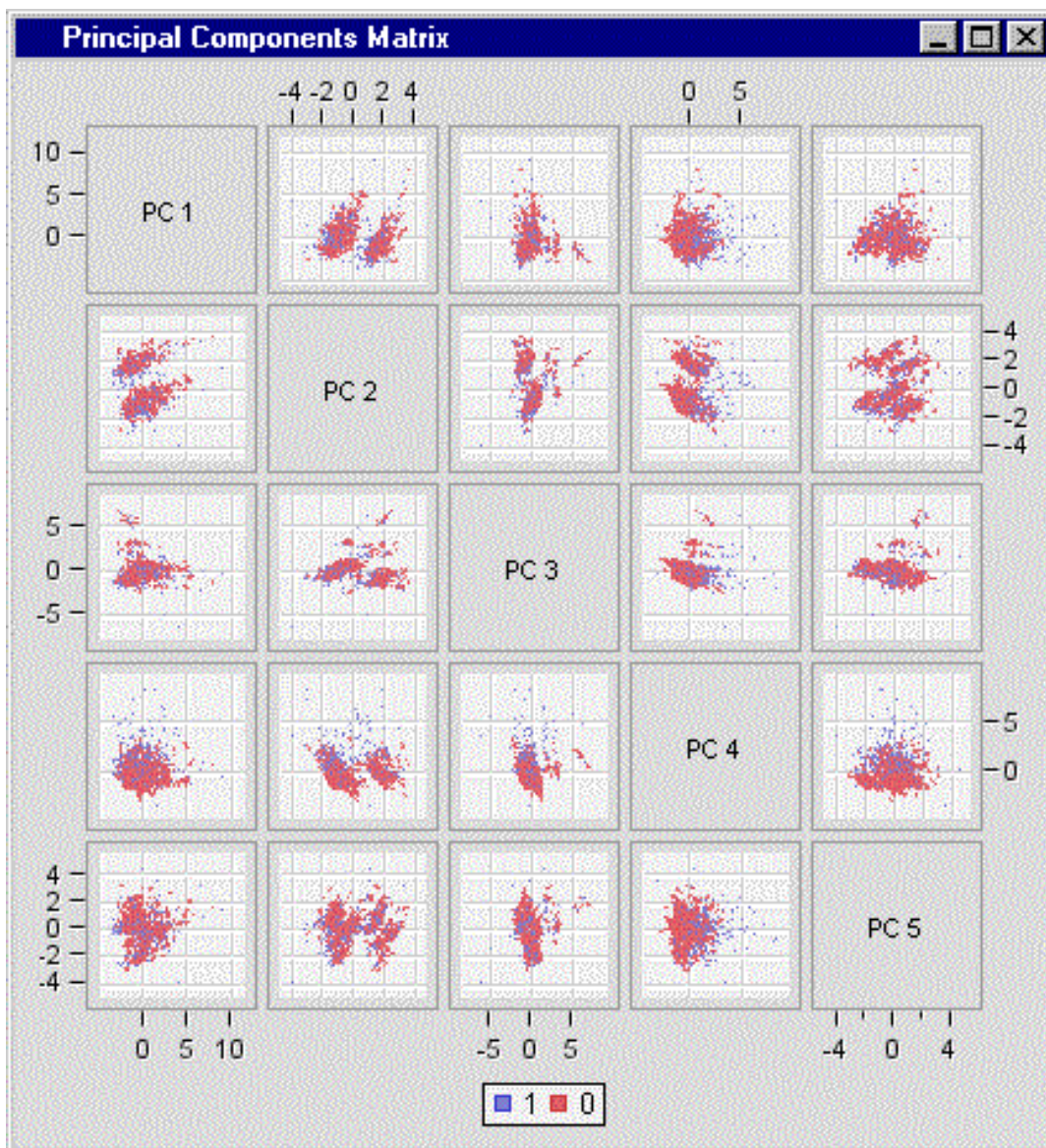


Principal Components Node

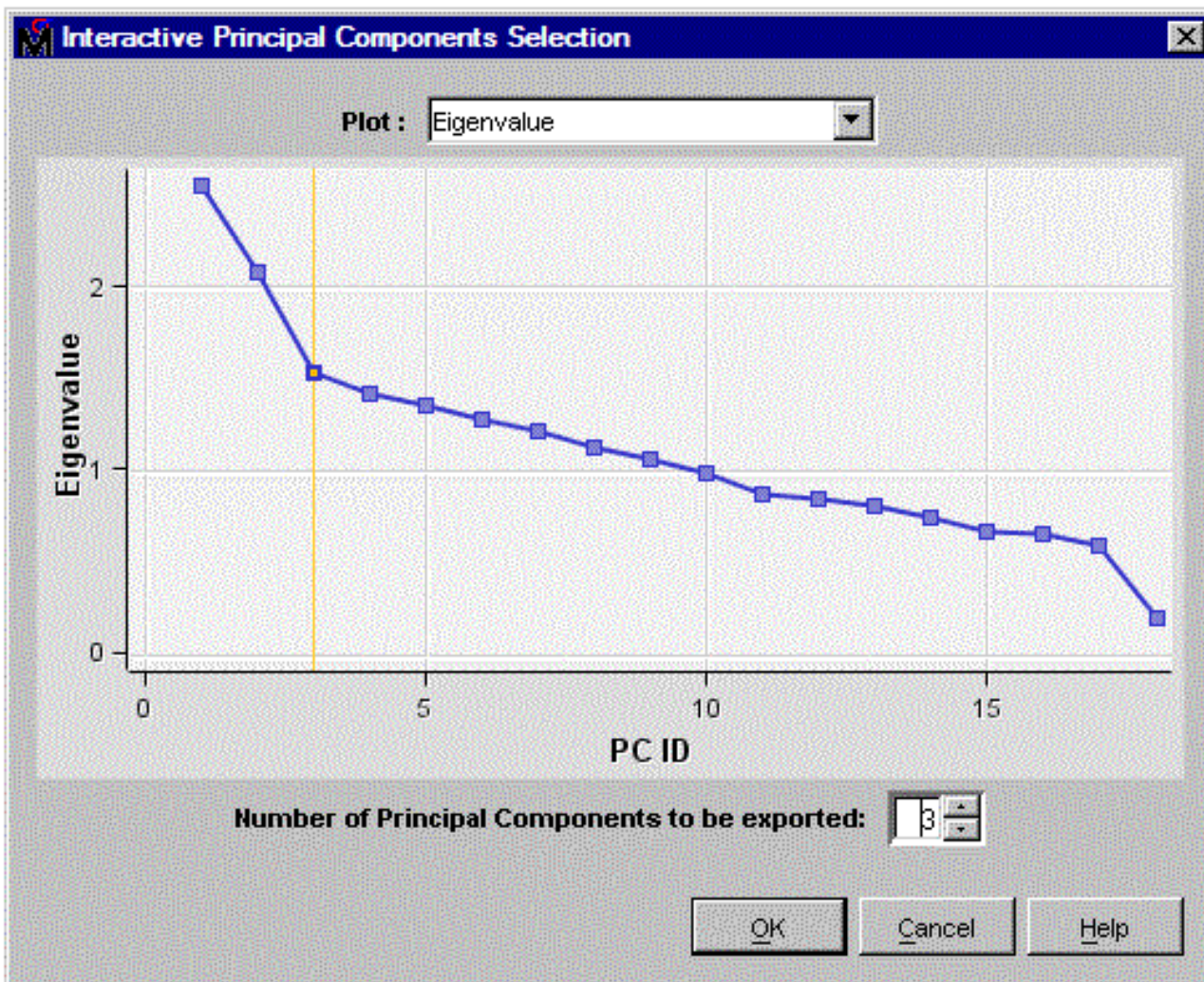
Major improvements have been made to the graphics output of the Principal Components node.



Of special interest is the new Principal Components Matrix plot. In the following display, the plots are color coded by target event. Two clusters are clearly seen in the PC1 and PC2 dimensions. The initial matrix contains five rows and five columns. You can determine the number of subplots that appear.



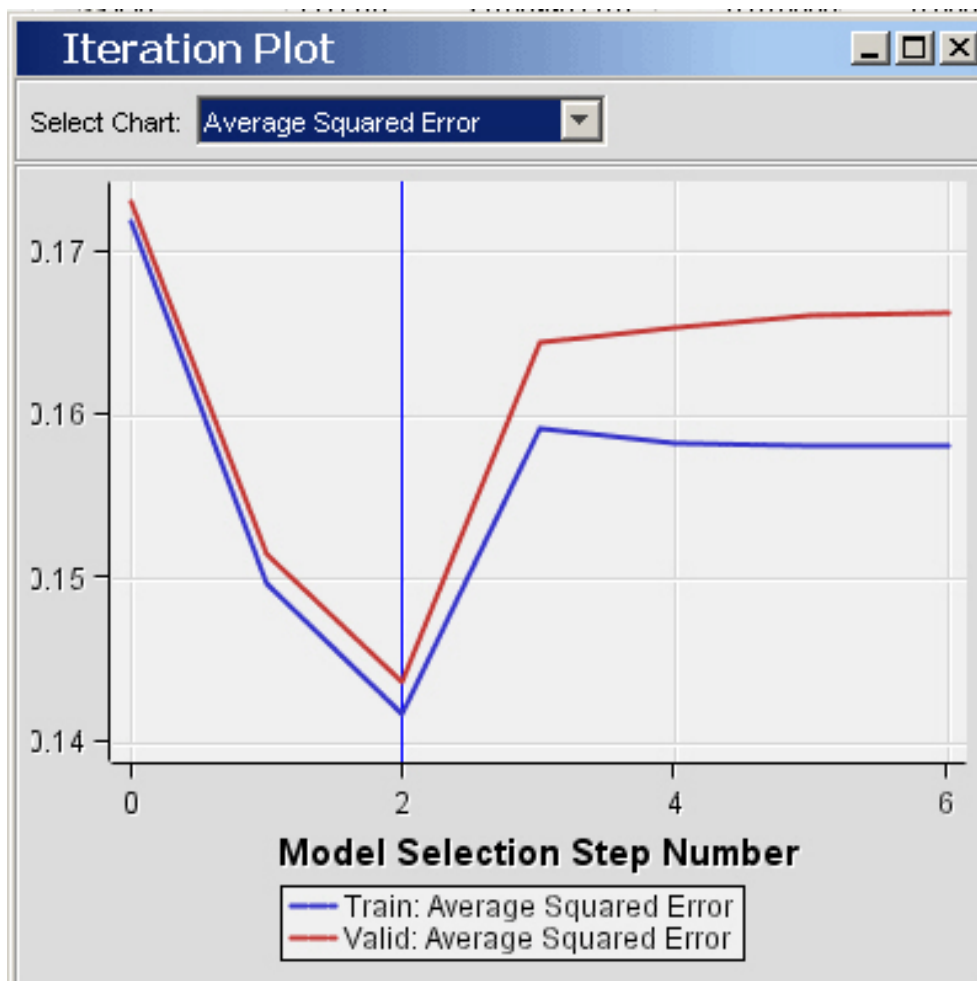
Using the Interactive Principal Components Selection window, you can now select how many principal components should be retained for analysis.



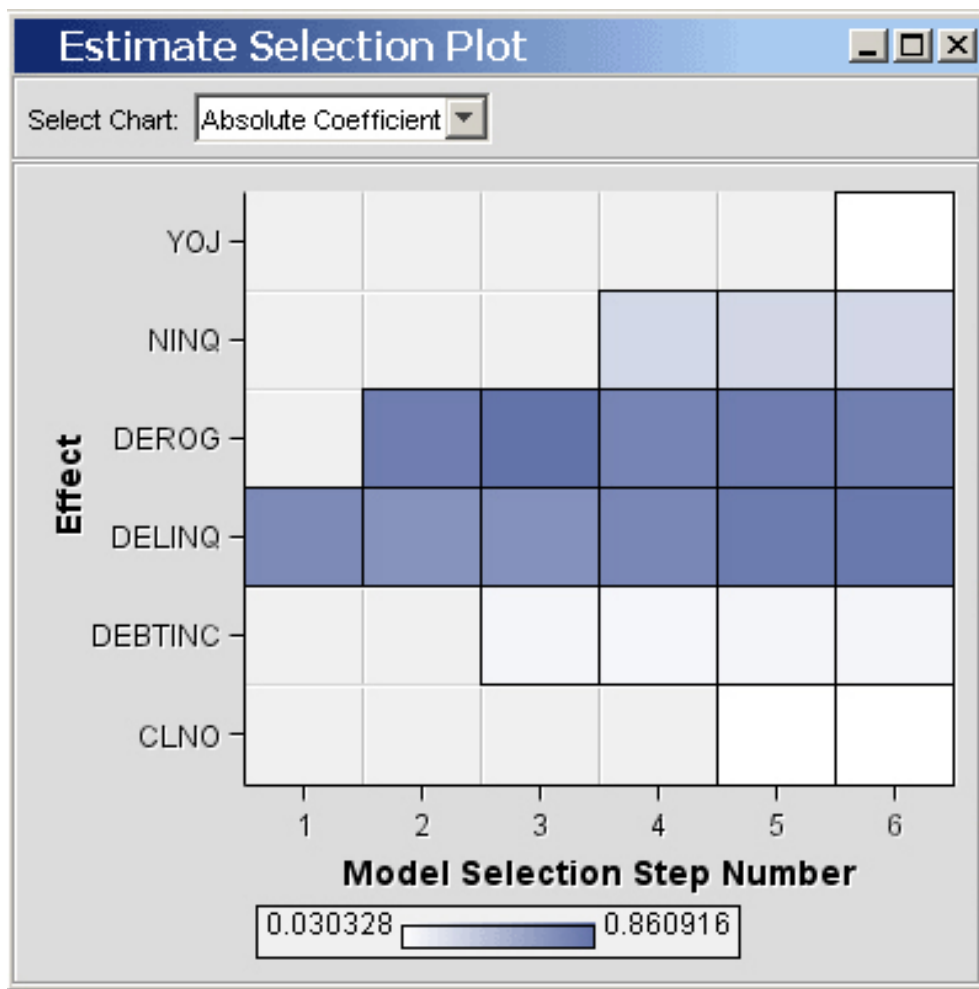
Regression Node

The following changes have been made to the plots that are generated when you specify a model selection method:

- The Iteration plot has been enhanced with a reference line for the selected step in the regression method. You can also now select the statistic to display in the plot.



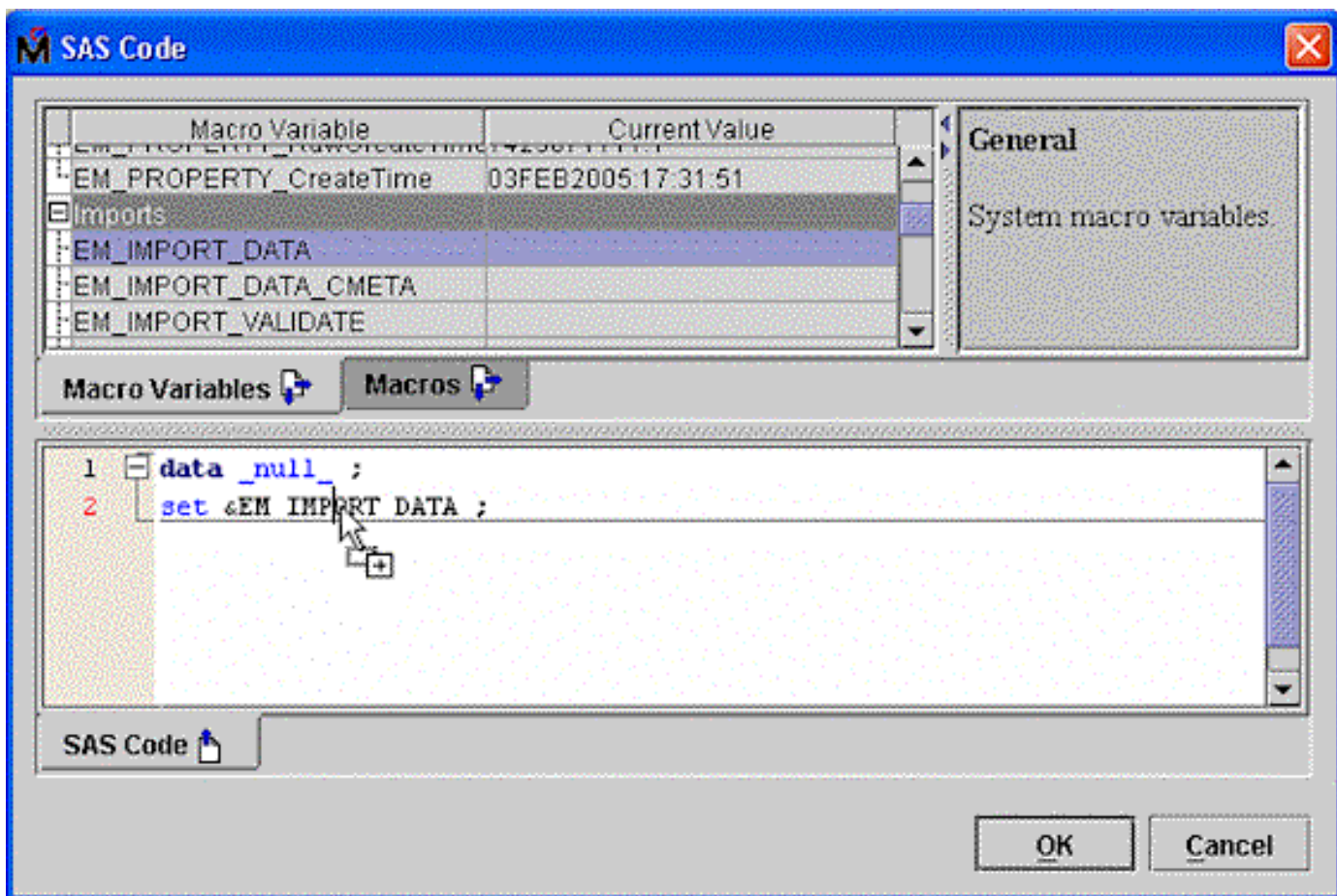
- The new Estimate Selection Plot displays the parameter estimates across various model selection steps.



The default for the Min Resource Use property is No when modeling a nominal or ordinal target and when specifying a model selection method. To improve performance but disable the model selection method, change this value to Yes.

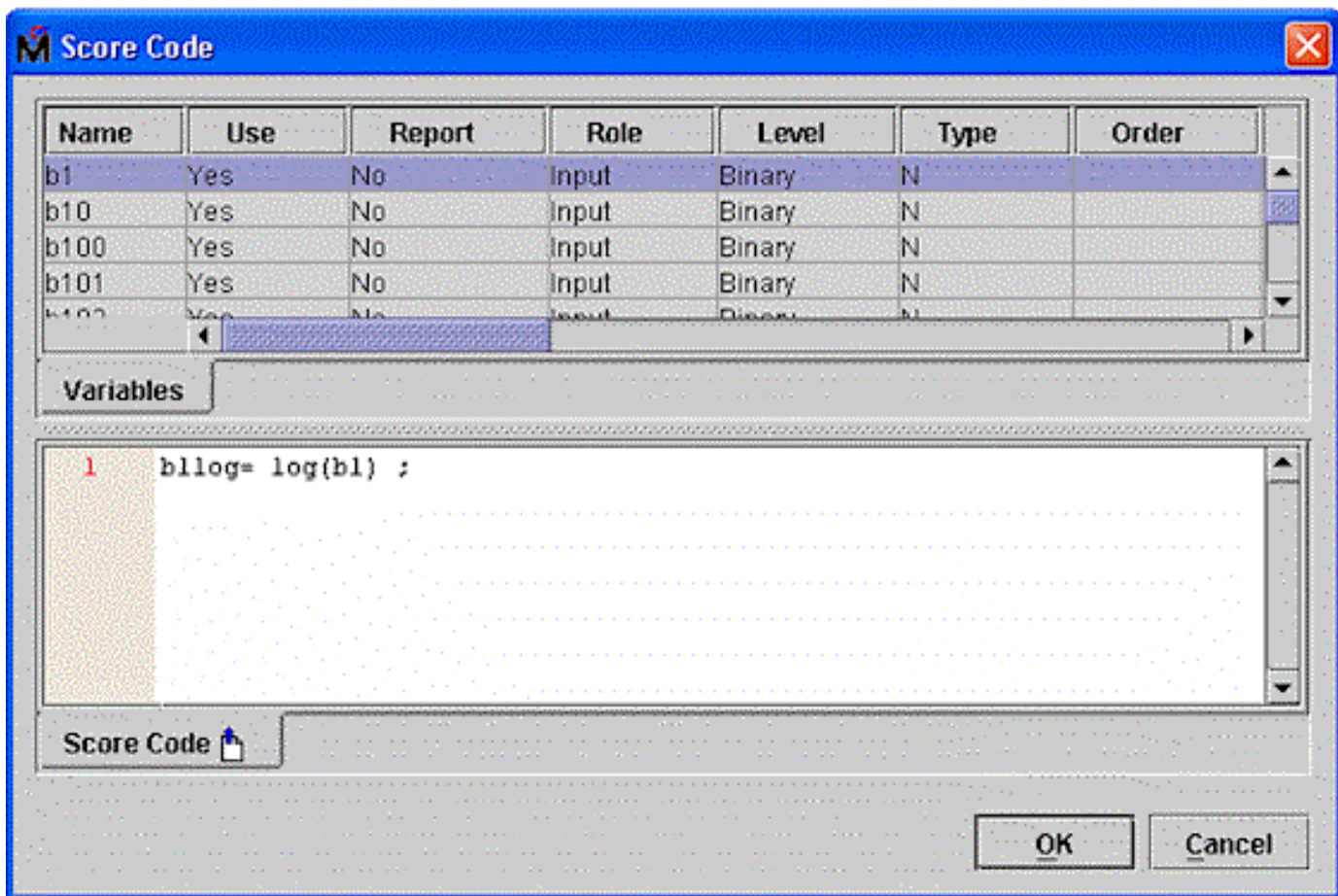
[SAS Code Node](#)

The SAS Code node features new training and scoring code editors.



The SAS Code editor window has Macro Variables and Macros tabs that contain lists of macro variables and macros that are used in SAS training code. You can use your mouse to drag items from the Macro Variables and Macros lists and drop them in the SAS Code editor to enhance and simplify SAS code creation.

The Score Code editor window has a Variables tab above the score code editor. The Variables tab enables you to view data set variables at the same time as the score code under composition, which makes creating new scoring code easier.



The SAS Code node is also able to use new macros:

- **%EM_MODEL()** — Use the %EM_MODEL() macro to specify to Enterprise Miner which target or targets your code has modeled. You can also control whether Enterprise Miner should assess individual targets, compute fit statistics, and generate score code to produce a decision, classification, or residuals variables.
- **%EM_METACHANGE()** — Use the %EM_METACHANGE() macro to simplify the process of changing column meta out of a node.
- **%EM_REPORT()** — Use the %EM_REPORT() macro to generate process flow diagram summary reports and plots. Many new graphics types are supported, and you have more control over the appearance of the plots.
- **%EM_REGISTER()** — Use the %EM_REGISTER() macro to specify a name and register a SAS data set for use with Enterprise Miner. The new PROPERTY argument indicates that the data set is a property of the node. If you specify this argument, then the data is saved when the node is copied or the flow is exported.

Sample Node

You now have better control of the sample makeup. The Level Based value has been added as an option to the Criterion property in the Sample Node Stratified Properties. If Level Based is selected, then the sample is based on the proportion captured and sample proportion of a specific level.


This criterion is applicable only when there is one stratification variable. Use the Level Based Options

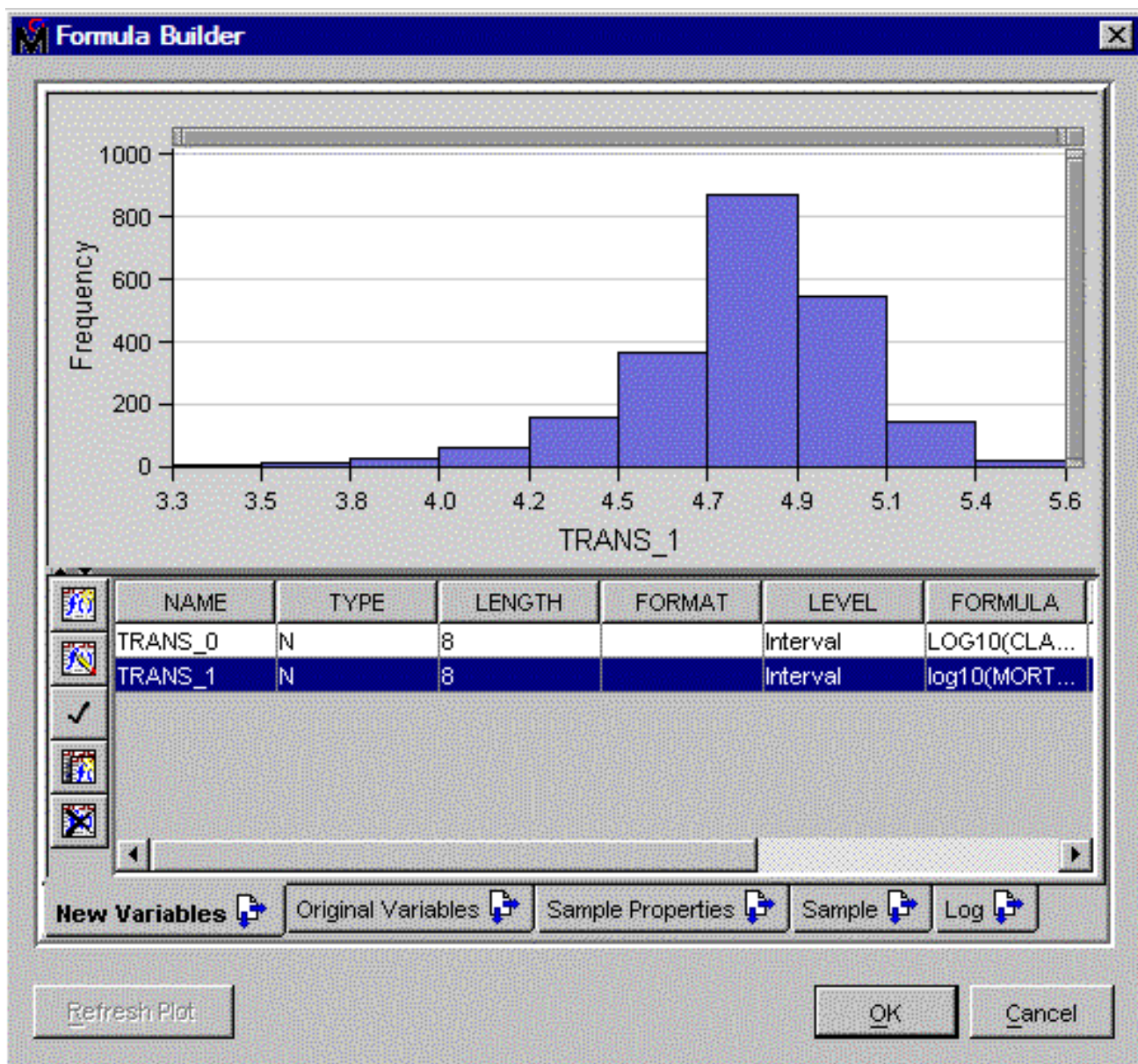
properties to specify parameters for the proportion captured, sample proportion, and level of interest.

Property	Value
Stratified	
Criterion	Level Based
Ignore Small Strata	No
Minimum Strata Size	5
Level Based Options	
Level Selection	Rarest Level
Level Proportion	100.0
Sample Proportion	50.0

Transform Variables Node

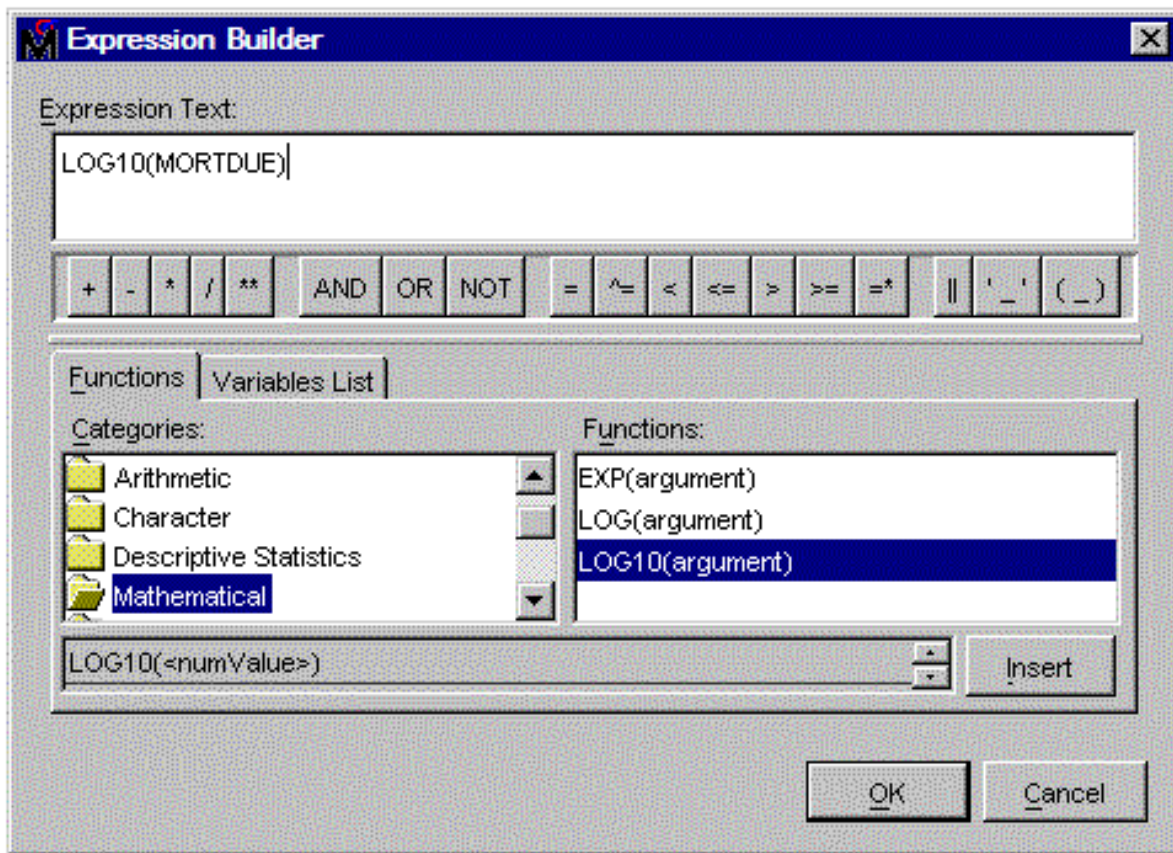
You can now create data transformations by using the push button interfaces that are found in the Formula Builder and Expression Builder windows.

To open the Formula Builder window, select the Transform node in your process flow diagram, and then in the Transform Variables node properties panel, select the ellipsis button  to the right of the Formula Builder entry.



The Formula Builder window is designed to respond to both keyboard and point-and-click mouse entries when you are building expressions. You can view distributions of both input variables and created (transformed) variables. The Formula Builder window enables you to see which variables need to be transformed as well as to see the distributions of the transformed variables that were created. A tab is available to examine and manipulate the Sample Properties.

New formulas are tested by executing the code against the preview sample and testing for errors. The preceding display shows a view of two newly created variables, including a plot of the second variable, TRANS_1.

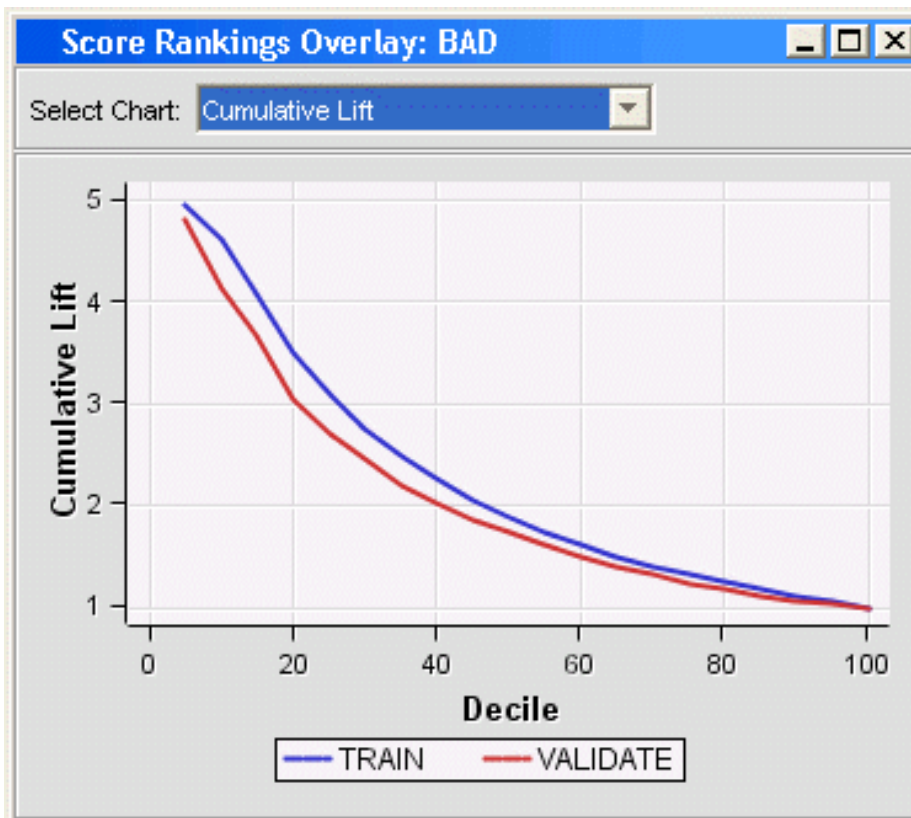


The Expression Builder window contains a grouped list of functions, a list of variables, and an Expression Text box that has Boolean operator shortcut keys.

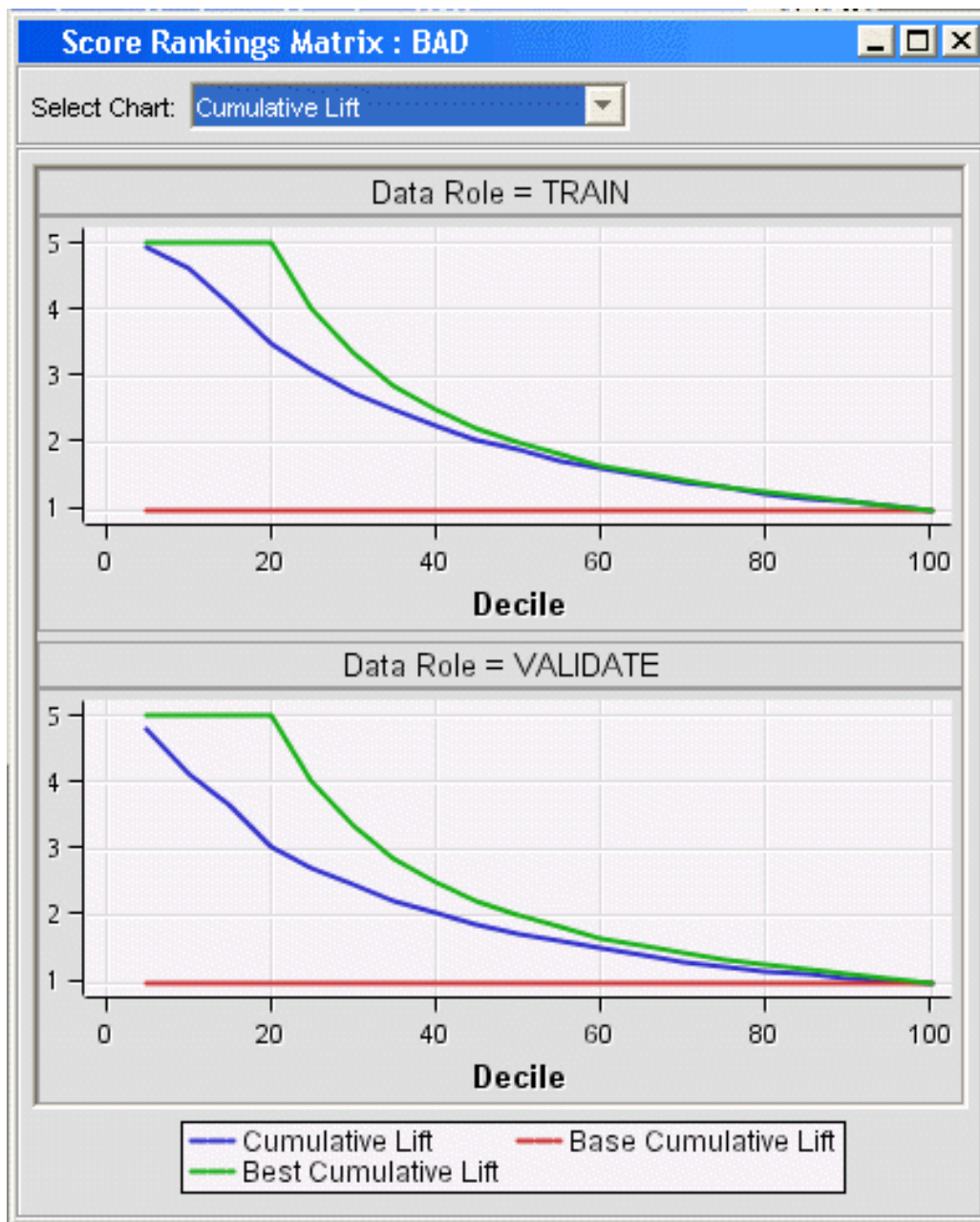
[New Model Assessment Reports in Enterprise Miner 5.2](#)

The assessment plots that are displayed in modeling nodes and the Model Comparison node have been enhanced to improve usability. The Score Rankings and Score Distribution plots now provide a selection control for conveniently changing the displayed statistics.

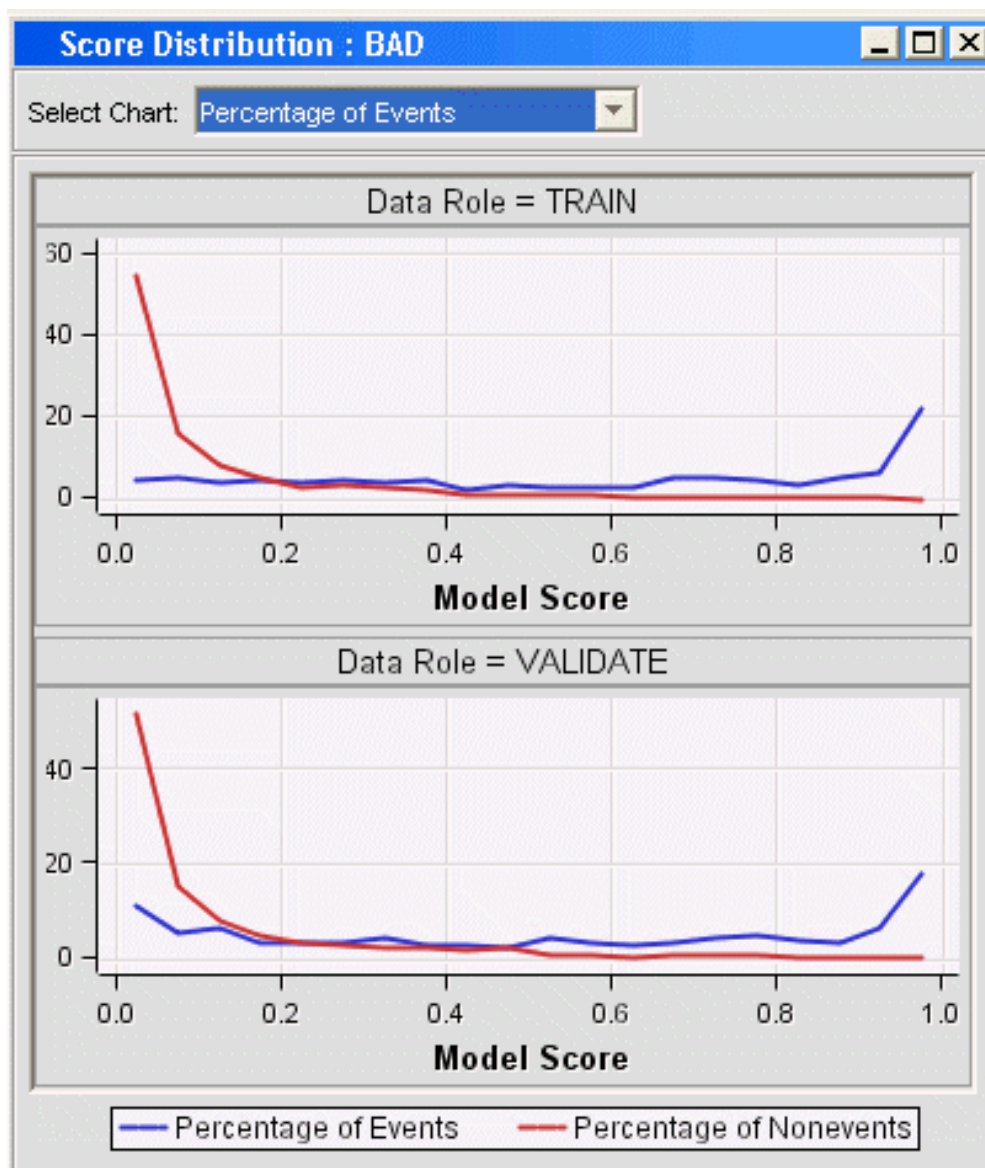
- **Score ranking overlay plots** use the same set of axes to simultaneously display selected statistics across deciles for both training and validation data sets. Report variables are not displayed in score ranking overlay plots.



- **Score ranking matrix plots** overlay the selected statistics for standard, baseline, and best models in a lattice that is defined by the training and validation data sets. The following example plots model cumulative lift, base model cumulative lift, and the best cumulative lift across deciles. Plots can also be created for report variables.



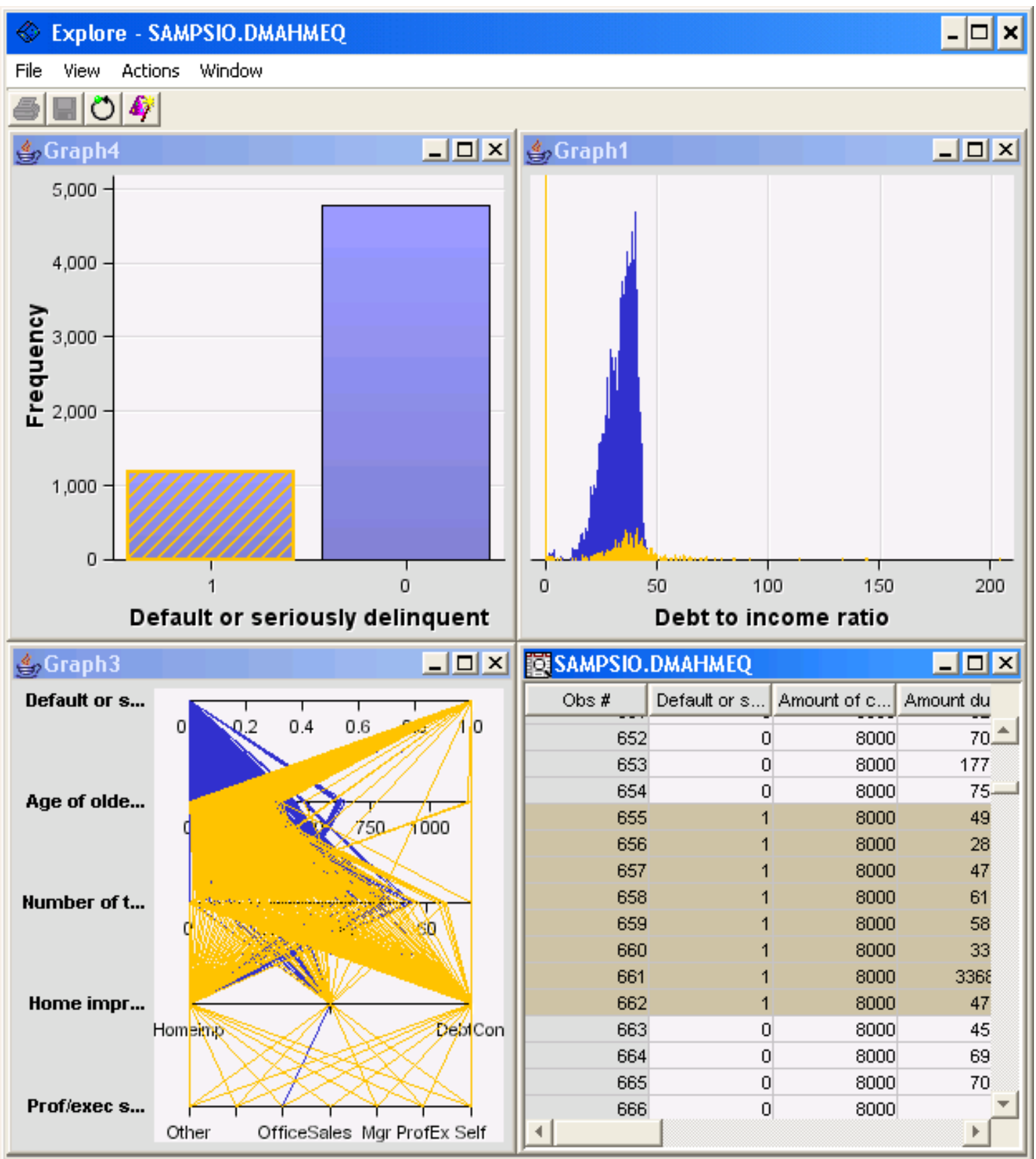
- **Score distribution plots** display statistics across score bins in a lattice that is defined by the training and validation data sets. Plots can also be created for report variables.



Other Changes and Enhancements

Graphics

The graphics libraries in the SAS Enterprise Miner client have been significantly enhanced with improved performance and many new plot types including two-dimensional and three-dimensional graphics. The following plot shows the new parallel axis and density plots.



Data Source Wizard

The Enterprise Miner wizard for adding data sources has been revised. The Data Source Wizard Decision Editor has been changed to support multiple decisions for categorical targets. The wizard leads you more clearly through the process of creating decisions and weights.

Data Source Wizard – Step 7 of 8 Decision Configuration

Targets | Prior Probabilities | **Decisions** | Decision Weights

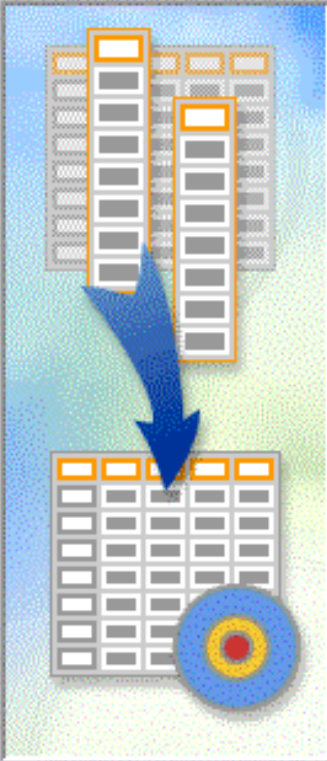
Do you want to use the decisions?

☒ Yes ☐ No

Decision Name	Label	Cost Variable	Constant
AMEX	American Ex...	< Constant >	1.2
VISA	Visa	< Constant >	2.4
MC	MasterCard	< Constant >	2.2
DISC	Discover	< Constant >	0.9

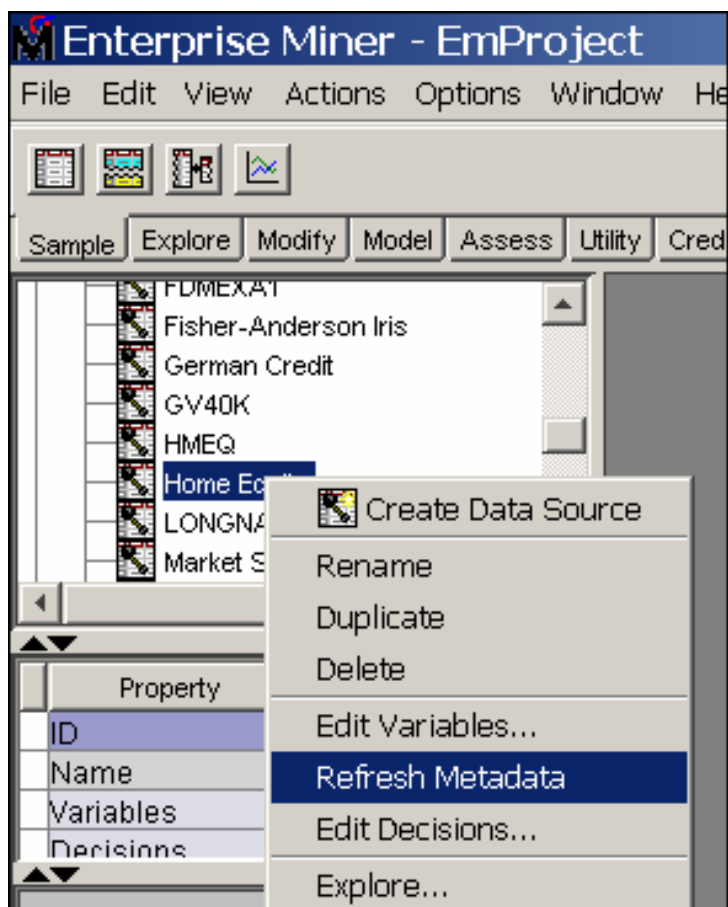
Add
Delete
Delete All
Reset
Default

< Back Next > Cancel Help







Refresh Metadata for a Defined Data Source

Changes to the underlying data (such as adding variables to the data source or removing variables from the data source) can affect the metadata of a data source that is defined in your project. SAS Enterprise Miner 5.2 now enables you to refresh the metadata for a defined data source rather than create a new data source. Right-click the data source in your project and select **Refresh Metadata**.



Quick Access to Node Data

All SAS Enterprise Miner 5.2 nodes now have an Exported Data property. The Exported Data property is located near the top of each node's properties list, which is displayed in the SAS Enterprise Miner 5.2 properties panel. The Exported Data property makes it easier for you to find SAS data sets that were created by a node. Simply select a node in your process flow diagram, then click the ellipsis button  to the right of the Exported Data property to open the Exported Data window for the selected node.

Property	Value
Node ID	Tree
Imported Data	
Exported Data	
Variables	

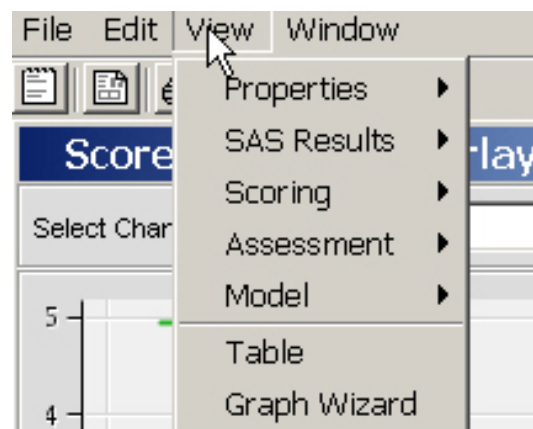
Port	Table	Role	Data Exists
TRAIN	EMWS5.Tree_TRAIN	Train	Yes
VALIDATE	EMWS5.Tree_VALIDATE	Validate	Yes
TEST	EMWS5.Tree_TEST	Test	Yes
RANK	EMWS5.Tree_EMRANK	RANK	Yes
SCOREDIST	EMWS5.Tree_EMSCOREDIST	SCOREDIST	Yes
REPORTFIT	EMWS5.Tree_EMREPORTFIT	REPORTFIT	Yes
TREE	EMWS5.Tree_ENTREE	TREE	Yes

Appearance Settings Are Now Persisted

Changes that you make to the size of the Variables Editor, Data Explorer, Table Browser, Results Viewer, and Help window or to the layout of the main window are now persisted between sessions. This enables you to customize the layout of the main window to meet your needs.

Reorganized View Menu for Results

The View menu in the Results Viewer has been reorganized for usability.



Extension Nodes

You can write your own XML files to create additional nodes to use in SAS Enterprise Miner. Starting in SAS Enterprise Miner 5.2, the deployment of custom nodes to clients has been simplified, so that all extension files (XML files and their icons) are managed from a central location.

Credit Scoring Nodes

SAS Credit Scoring is a new solution in SAS Enterprise Miner 5.2. This solution offers the ability to

rapidly generate automated credit scoring models that rely on statistical models.

If your site has licensed SAS Credit Scoring, the credit scoring nodes will appear on the Credit Scoring tab in your Enterprise Miner session. SAS Enterprise Miner includes the following credit scoring nodes:

- Interactive Grouping
- Scorecard
- Reject Inference
- Credit Exchange

You can use these nodes to perform the following tasks:

- build scorecard models that assign score points to customer attributes
- group and select characteristics, automatically or interactively, by using Weights of Evidence and Information Value measures
- normalize score points to conform with company or industry standards

The business value of statistical models can be assessed by using strategy curves, profit charts, and a reject inference process in order to produce models for scoring through-the-door populations. A credit exchange tool provides additional reporting of the credit scoring results and can exchange information with the SAS Credit Risk solution.

The Interactive Grouping and Scorecard nodes have been enhanced for this release. See SAS Credit Scoring Overview for more information.

[Back to Top of What's New in SAS Enterprise Miner 5.2](#)