- FILE_AVG_GIFT
- LAST_GIFT_AMT
- LIFETIME_AVG_GIFT_AMT
- LIFETIME_GIFT_AMOUNT

TIP You can hold down the Ctrl key to select multiple rows. Then, when you select a new **Method** for one of the selected variables, the new method will apply to all of the selected variables.

b. Select the transformation **Method** for the following interval variables and select **Optimal Binning** from the drop-down menu that appears:

- LIFETIME_CARD_PROM
- LIFETIME_GIFT_COUNT
- MEDIAN_HOME_VALUE
- MEDIAN_HOUSEHOLD_INCOME
- PER_CAPITA_INCOME
- RECENT_RESPONSE_PROP
- RECENT_STAR_STATUS

The optimal binning transformation is useful when there is a nonlinear relationship between an input variable and the target. For more information about this transformation, see the SAS Enterprise Miner Help.

c. Click **OK**.

5. In the Diagram Workspace, right-click the Transform Variables node, and select **Run** from the resulting menu. Click **Yes** in the Confirmation window that opens.
6. In the window that appears when processing completes, click **OK**.

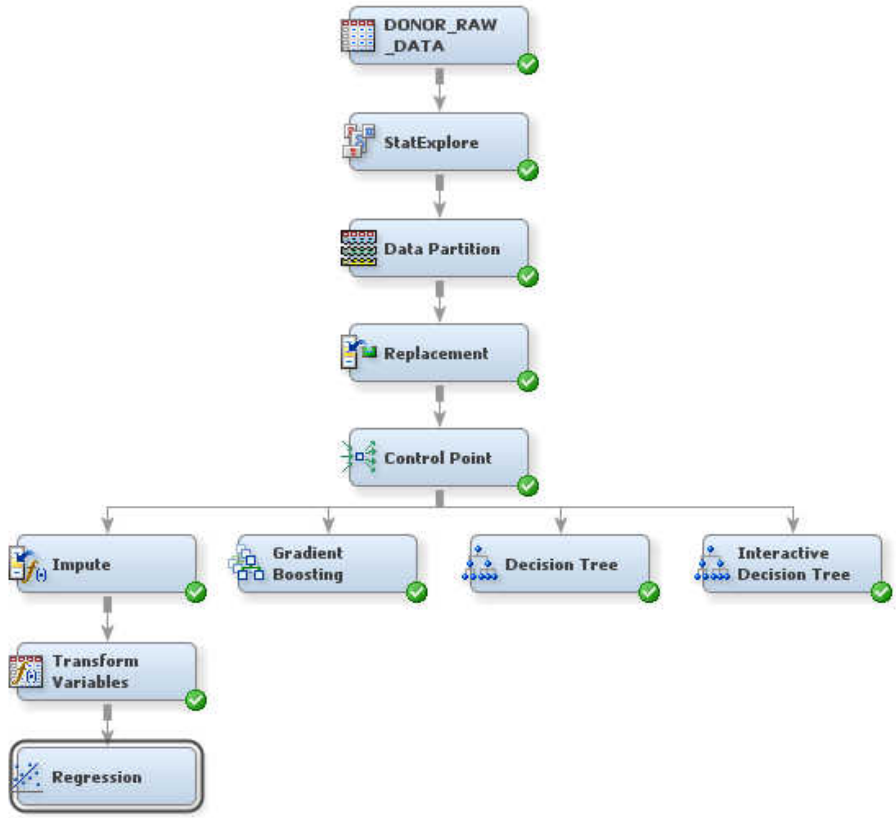
Note: In the data that is exported from the Transform Variables node, a new variable is created for each variable that is transformed. The original variable is not overwritten. Instead, the new variable has the same name as the original variable but is prefaced with an identifier of the transformation. For example, variables to which the log transformation have been applied are prefaced with LOG_, and variables to which the optimal binning transformation have been applied are prefaced with OPT_. The original version of each variable also exists in the exported data and has the role `Rejected`.

Analyze with a Logistic Regression Model

As part of your analysis, you want to include some parametric models for comparison with the decision trees that you built in Chapter 5, “Build Decision Trees,” on page 21. Because it is familiar to the management of your organization, you have decided to include a logistic regression as one of the parametric models.

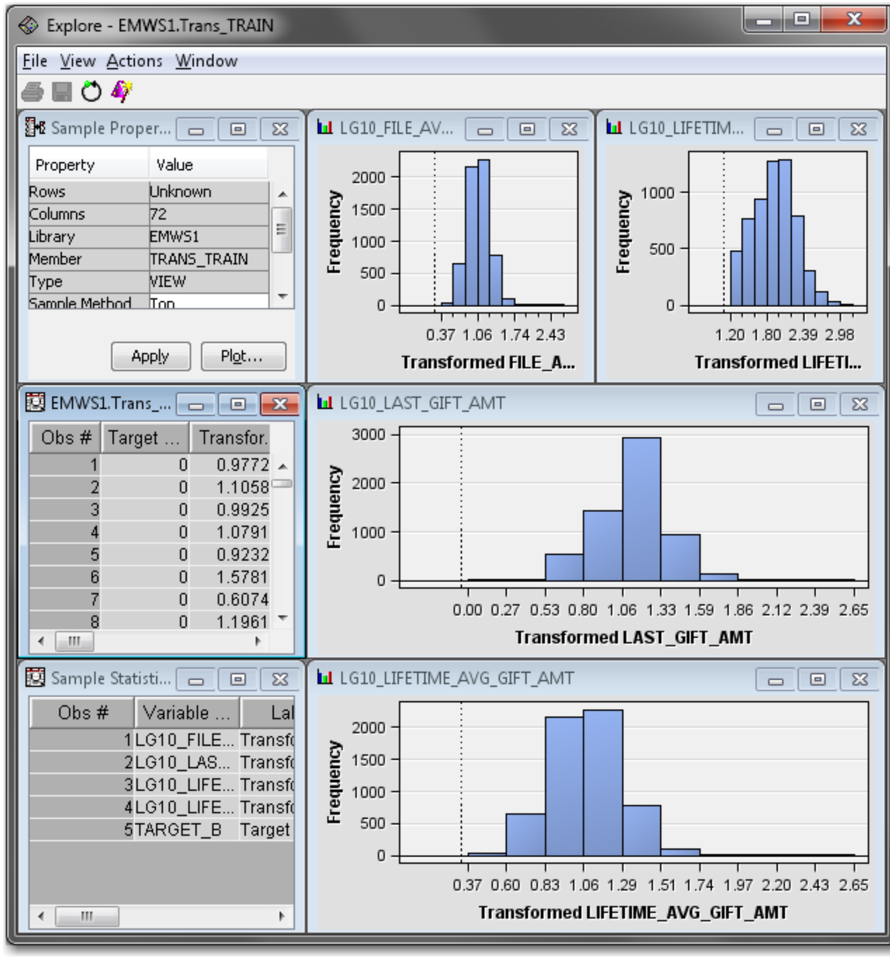
To use the Regression node to fit a logistic regression model:

1. Select the **Model** tab on the Toolbar.
2. Select the **Regression** node icon. Drag the node into the Diagram Workspace.
3. Connect the **Transform Variables** node to the **Regression** node.



4. To examine histograms of the imputed and transformed input variables, right-click the Regression node and select **Update**. In the diagram workspace, select the Regression node. In the Properties Panel, scroll down to view the Train properties, and click on the ellipses that represent the value of **Variables**. The Variables — Reg window appears.

   a. Select all variables that have the prefix LG10_. Click **Explore**. The Explore window appears.



You can select a bar in any histogram, and the observations that are in that bucket are highlighted in the EMWS.Trans_TRAIN data set window and in the other histograms. Close the Explore window to return to the Variables — Reg window.

b. (Optional) You can explore the histograms of other input variables.

c. Close the Variables — Reg window.

5. In the Properties Panel, scroll down to view the Train properties. Click on the **Selection Model** property in the **Model Selection** subgroup, and select **Stepwise** from the drop-down menu that appears. This specification causes SAS Enterprise Miner to use stepwise variable selection to build the logistic regression model.

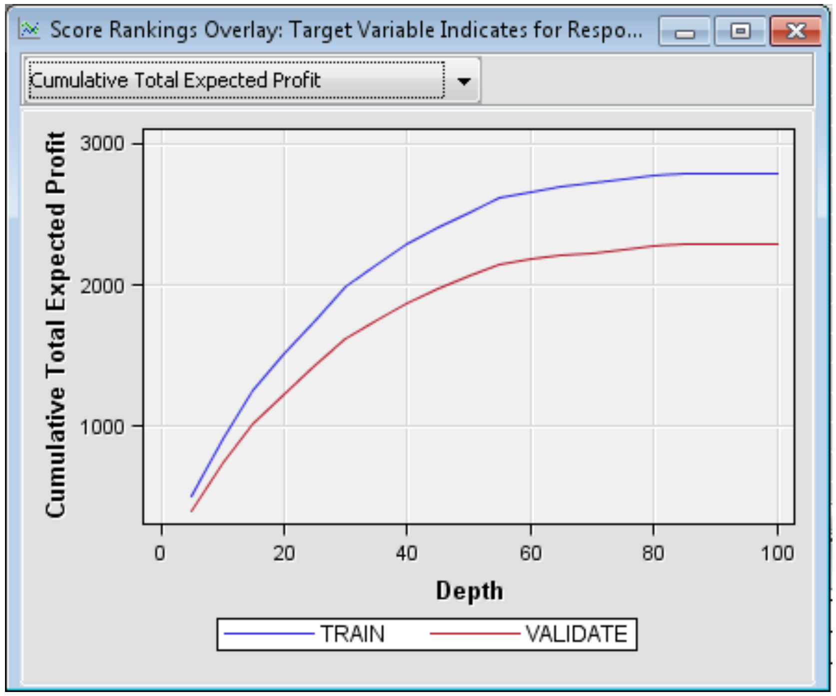
   Note: The Regression node automatically performs logistic regression if the target variable is a class variable that takes one of two values. If the target variable is a continuous variable, then the Regression node performs linear regression.

6. In the Diagram Workspace, right-click the Regression node, and select **Run** from the resulting menu. Click **Yes** in the Confirmation window that opens.

7. In the window that appears when processing completes, click **Results**. The Results window appears.

8. Maximize the Output window. This window details the variable selection process. Lines 401 – 424 list a summary of the steps that were taken.

9. Minimize the Output window and maximize the Score Rankings Overlay window. From the drop-down menu, select **Cumulative Total Expected Profit**.



The data that is used to construct this plot is ordered by expected profit. For this example, you have defined a profit matrix. Therefore, expected profit is a function of both the probability of donation for an individual and the profit associated with the corresponding outcome. A value is computed for each decision from the sum of the decision matrix values multiplied by the classification probabilities and minus any defined cost. The decision with the greatest value is selected, and the value of that selected decision for each observation is used to compute overall profit measures.

The plot represents the cumulative total expected profit that results from soliciting the best *n*% of the individuals (as determined by expected profit) on your mailing list. For example, if you were to solicit the best 40% of the individuals, the total expected profit from the validation data would be approximately $1850. If you were to solicit everyone on the list, then based on the validation data, you could expect approximately $2250 profit on the campaign.

10. Close the Results window.

Analyze with a Neural Network Model

Neural networks are a class of parametric models that can accommodate a wider variety of nonlinear relationships between a set of predictors and a target variable than can logistic regression. Building a neural network model involves two main phases. First, you must define the network configuration. You can think of this step as defining the structure of the model that you want to use. Then, you iteratively train the model.

A neural network model will be more complicated to explain to the management of your organization than a regression or a decision tree. However, you know that the management would prefer a stronger predictive model, even if it is more complicated. So, you decide to run a neural network model, which you will compare to the other models later in the example.