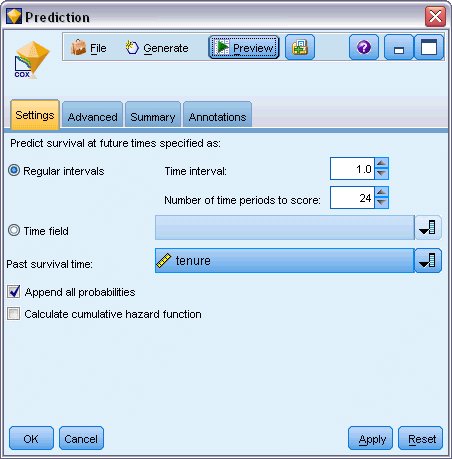
* [Tracking the Expected Number of Customers Retained](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/common/examples/coxreg_telco_prediction.htm)

**Tracking the Expected Number of Customers Retained**

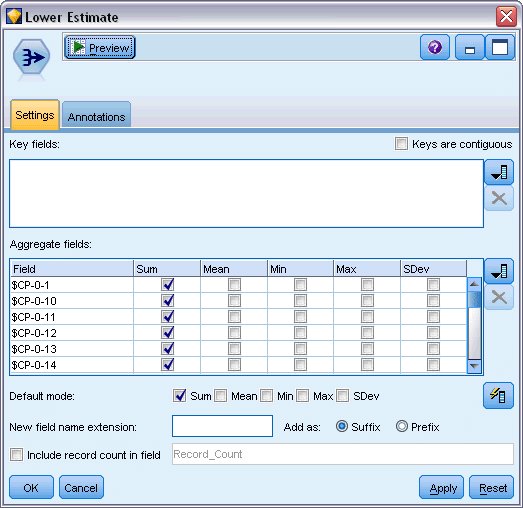
Once satisfied with a model, you want to track the expected number of customers in the dataset that are retained over the next two years. The null values, which are customers whose total tenure (future time + *tenure*) falls beyond the range of survival times in the data used to train the model, present an interesting challenge. One way to deal with them is to create two sets of predictions, one in which null values are assumed to have churned, and another in which they are assumed to have been retained. In this way you can establish upper and lower bounds on the expected number of customers retained.

*Figure 1. Cox nugget: Settings tab*



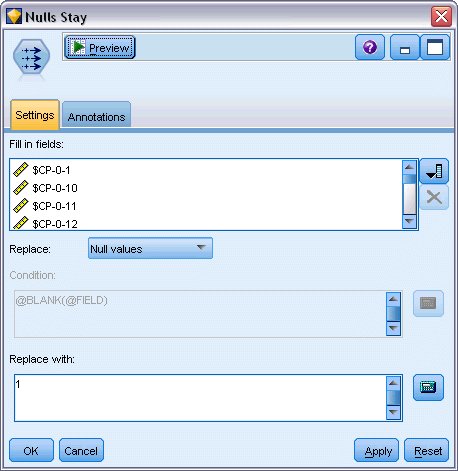
1. Double-click the model nugget in the Models palette (or copy and paste the nugget on the stream canvas) and attach the new nugget to the Source node.
2. Open the nugget to the Settings tab.
3. Make sure **Regular Intervals** is selected, and specify 1.0 as the time interval and 24 as the number of periods to score. This specifies that each record will be scored for each of the following 24 months.
4. Select *tenure* as the field to specify the past survival time. The scoring algorithm will take into account the length of each customer's time as a customer of the company.
5. Select **Append all probabilities**.

*Figure 2. Aggregate node: Settings tab*



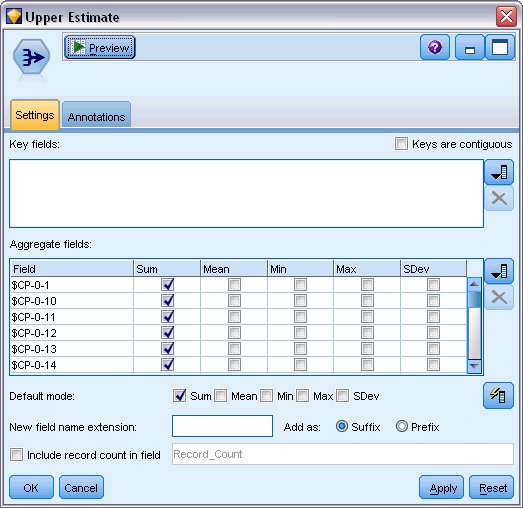
1. Attach an Aggregate node to the model nugget; on the Settings tab, deselect **Mean** as a default mode.
2. Select *$CP-0-1* through *$CP-0-24*, the fields of form *$CP-0-n*, as the fields to aggregate. This is easiest if, on the Select Fields dialog, you sort the fields by Name (that is, alphabetical order).
3. Deselect **Include record count in field**.
4. Click **OK**. This node creates the "lower bound" predictions.

*Figure 3. Filler node: Settings tab*



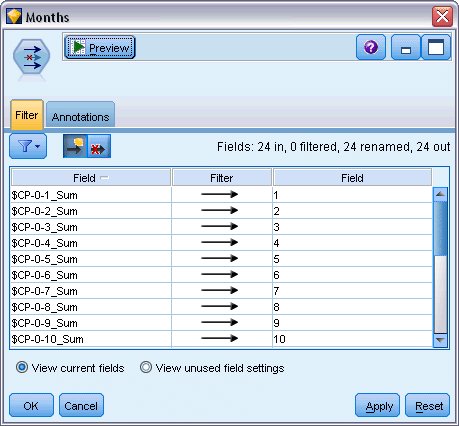
1. Attach a Filler node to the Coxreg nugget to which we just attached the Aggregate node; on the Settings tab, select *$CP-0-1* through *$CP-0-24*, the fields of form *$CP-0-n*, as the fields to fill in. This is easiest if, on the Select Fields dialog, you sort the fields by Name (that is, alphabetical order).
2. Choose to replace **Null values** with the value 1.
3. Click **OK**.

*Figure 4. Aggregate node: Settings tab*



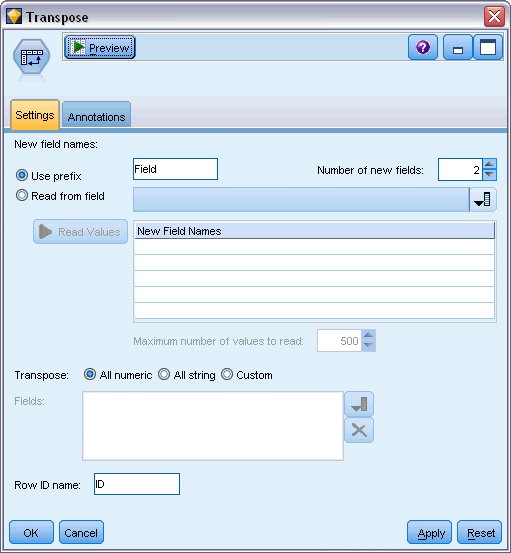
1. Attach an Aggregate node to the Filler node; on the Settings tab, deselect **Mean** as a default mode.
2. Select *$CP-0-1* through *$CP-0-24*, the fields of form *$CP-0-n*, as the fields to aggregate. This is easiest if, on the Select Fields dialog, you sort the fields by Name (that is, alphabetical order).
3. Deselect **Include record count in field**.
4. Click **OK**. This node creates the "upper bound" predictions.

*Figure 5. Filter node: Settings tab*



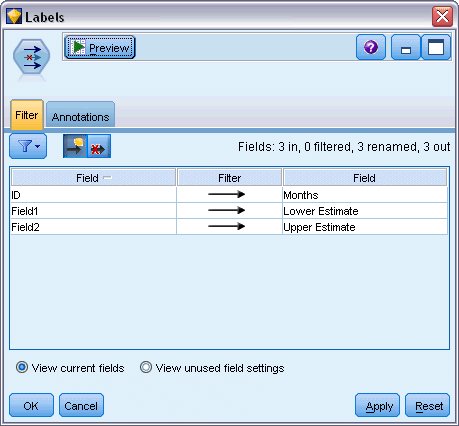
1. Attach an Append node to the two Aggregate nodes, then attach a Filter node to the Append node.
2. On the Settings tab of the Filter node, rename the fields to *1* through *24*. Through the use of a Transpose node, these field names will become values for the *x*-axis in charts downstream.

*Figure 6. Transpose node: Settings tab*



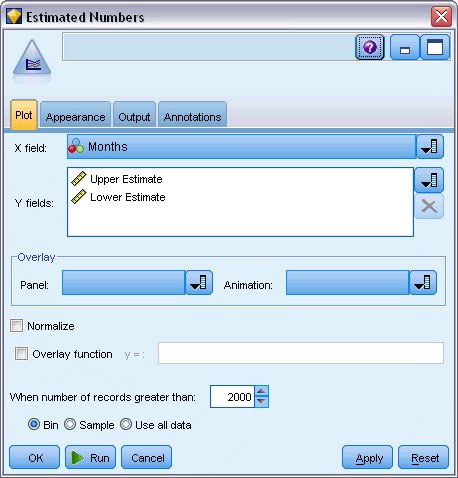
1. Attach a Transpose node to the Filter node.
2. Type 2 as the number of new fields.

*Figure 7. Filter node: Filter tab*



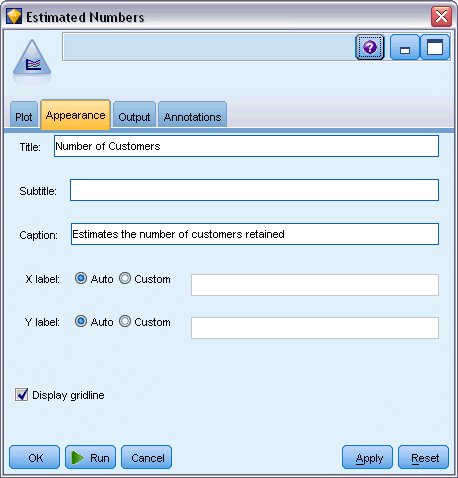
1. Attach a Filter node to the Transpose node.
2. On the Settings tab of the Filter node, rename *ID*to *Months*, *Field1* to *Lower Estimate*, and *Field2* to *Upper Estimate*.

*Figure 8. Multiplot node: Plot tab*



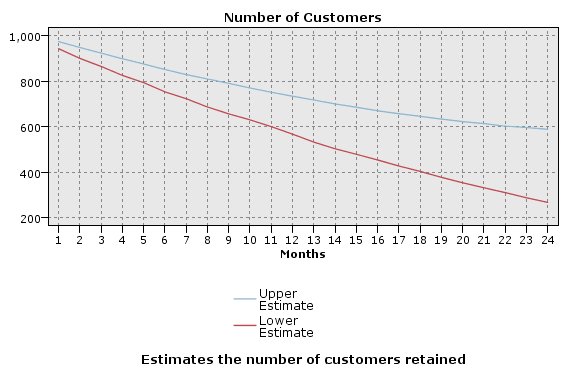
1. Attach a Multiplot node to the Filter node.
2. On the Plot tab, *Months* as the X field, *Lower Estimate* and *Upper Estimate* as the Y fields.

*Figure 9. Multiplot node: Appearance tab*



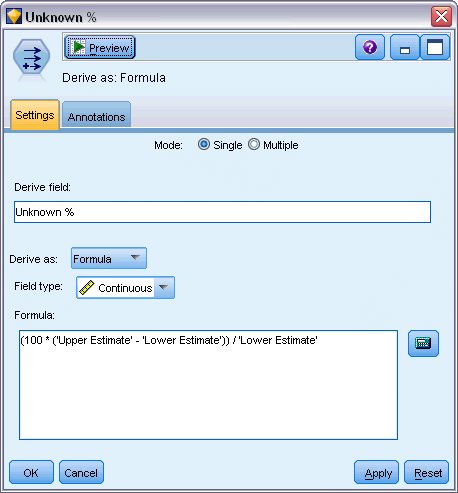
1. Click the Appearance tab.
2. Type Number of Customers as the title.
3. Type Estimates the number of customers retained as the caption.
4. Click **Run**.

*Figure 10. Multiplot estimating the number of customers retained*



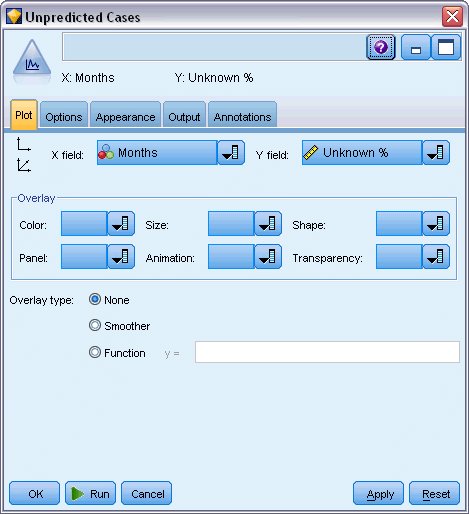
The upper and lower bounds on the estimated number of customers retained are plotted. The difference between the two lines is the number of customers scored as null, and therefore whose status is highly uncertain. Over time, the number of these customers increases. After 12 months, you can expect to retain between 601 and 735 of the original customers in the dataset; after 24 months, between 288 and 597.

*Figure 11. Derive node: Settings tab*



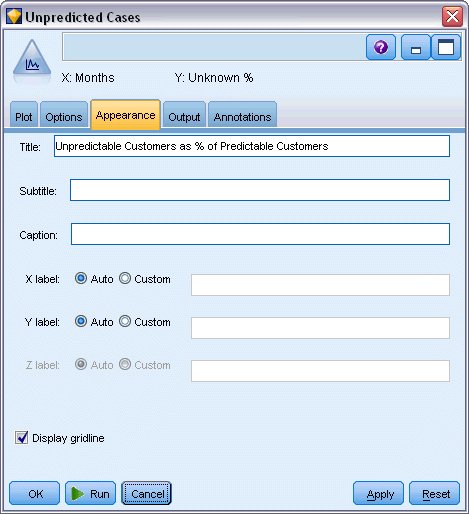
1. To get another look at how uncertain the estimates of the number of customers retained are, attach a Derive node to the Filter node.
2. On the Settings tab of the Derive node, type *Unknown %* as the derive field.
3. Select **Continuous** as the field type.
4. Type (100 \* ('Upper Estimate' - 'Lower Estimate')) / 'Lower Estimate' as the formula. *Unknown %* is the number of customers "in doubt" as a percentage of the lower estimate.
5. Click **OK**.

*Figure 12. Plot node: Plot tab*



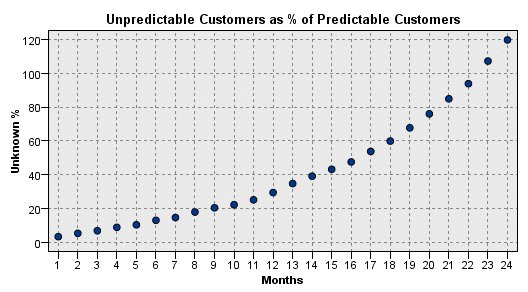
1. Attach a Plot node to the Derive node.
2. On the Plot tab of the Plot node, select *Months* as the X field and *Unknown %* as the Y field.
3. Click the **Appearance** tab.

*Figure 13. Plot node: Appearance tab*



1. Type Unpredictable Customers as % of Predictable Customers as the title.
2. Execute the node.

*Figure 14. Plot of unpredictable customers*

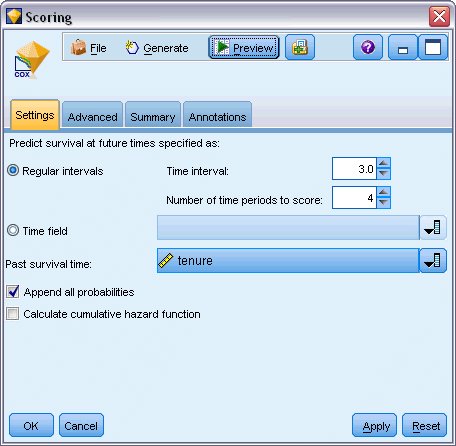


Through the first year, the percentage of unpredictable customers increases at a fairly linear rate, but the rate of increase explodes during the second year until, by month 23, the number of customers with null values outnumber the expected number of customers retained.

**Scoring**

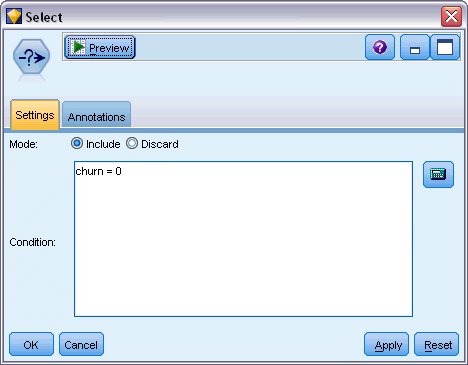
Once satisfied with a model, you want to score customers to identify the individuals most likely to churn within the next year, by quarter.

*Figure 1. Coxreg nugget: Settings tab*



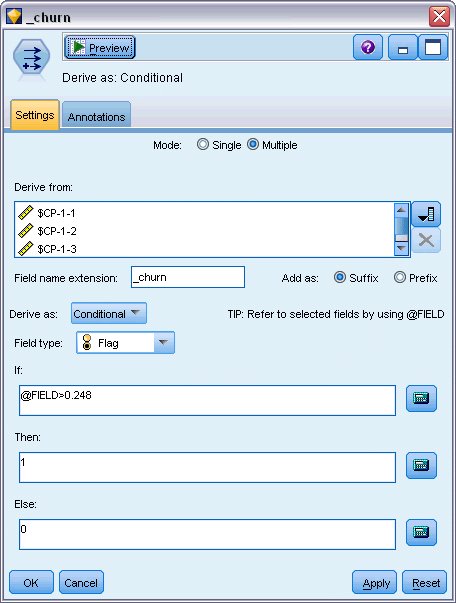
1. Attach a third model nugget to the Source node and open the model nugget.
2. Make sure **Regular Intervals** is selected, and specify 3.0 as the time interval and 4 as the number of periods to score. This specifies that each record will be scored for the following four quarters.
3. Select *tenure* as the field to specify the past survival time. The scoring algorithm will take into account the length of each customer's time as a customer of the company.
4. Select **Append all probabilities**. These extra fields will make it easier to sort the records for viewing in a table.

*Figure 2. Select node: Settings tab*



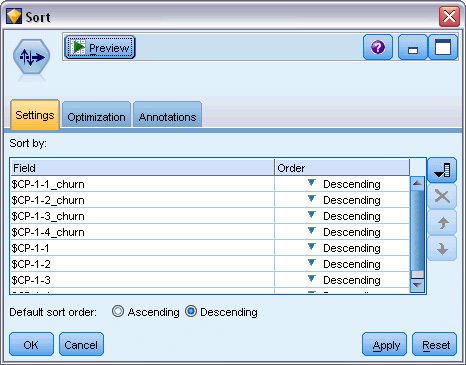
1. Attach a Select node to the model nugget; on the Settings tab, type churn=0 as the condition. This removes customers who have already churned from the results table.

*Figure 3. Derive node: Settings tab*



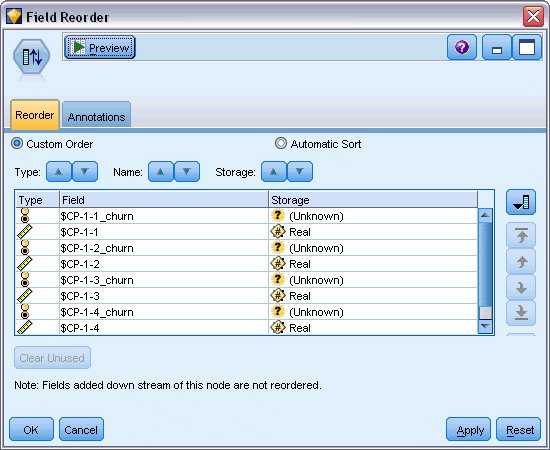
1. Attach a Derive node to the Select node; on the Settings tab, select **Multiple** as the mode.
2. Choose to derive from *$CP-1-1* through *$CP-1-4*, the fields of form *$CP-1-n*, and type \_churn as the suffix to add. This is easiest if, on the Select Fields dialog, you sort the fields by Name (that is, alphabetical order).
3. Choose to derive the field as a **Conditional**.
4. Select **Flag**as the measurement level.
5. Type @FIELD>0.248 as the **If** condition. Recall that this was the classification cutoff identified during Evaluation.
6. Type 1 as the **Then** expression.
7. Type 0 as the **Else** expression.
8. Click **OK**.

*Figure 4. Sort node: Settings tab*



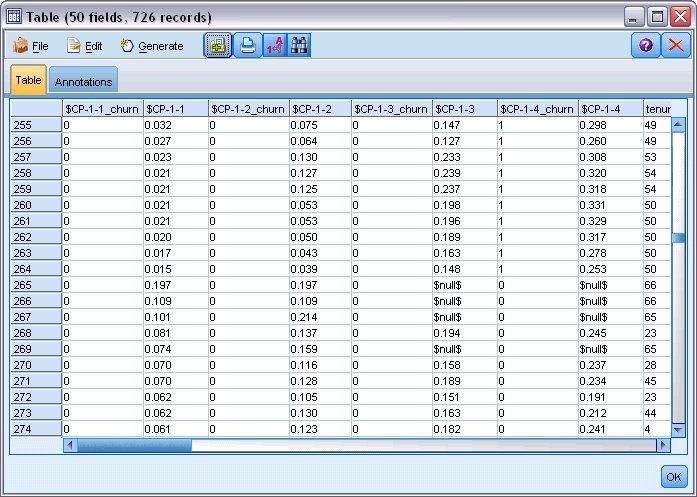
1. Attach a Sort node to the Derive node; on the Settings tab, choose to sort by *$CP-1-1\_churn* through *$CP-1-4-churn* and then *$CP-1-1* through *$CP-1-4*, all in descending order. Customers who are predicted to churn will appear at the top.

*Figure 5. Field Reorder node: Reorder tab*



1. Attach a Field Reorder node to the Sort node; on the Reorder tab, choose to place *$CP-1-1\_churn* through *$CP-1-4* in front of the other fields. This simply makes the results table easier to read, and so is optional. You will need to use the buttons to move the fields into the position shown in the figure.

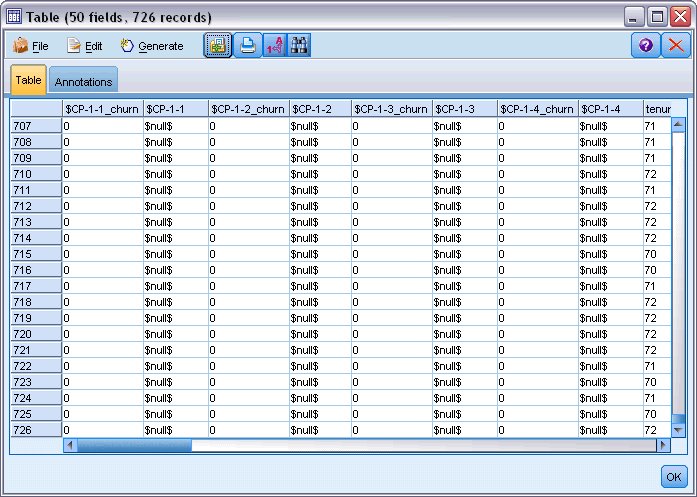
*Figure 6. Table showing customer scores*



1. Attach a Table node to the Field Reorder node and execute it.

264 customers are expected to churn by the end of the year, 184 by the end of the third quarter, 103 by the second, and 31 in the first. Note that given two customers, the one with a higher propensity to churn in the first quarter does not necessarily have a higher propensity to churn in later quarters; for example, see records 256 and 260. This is likely due to the shape of the hazard function for the months following the customer's current tenure; for example, customers who joined because of a promotion might be more likely to switch early on than customers who joined because of a personal recommendation, but if they do not then they may actually be more loyal for their remaining tenure. You may want to re-sort the customers to obtain different views of the customers most likely to churn.

*Figure 7. Table showing customers with null values*



At the bottom of the table are customers with predicted null values. These are customers whose total tenure (future time + *tenure*) falls beyond the range of survival times in the data used to train the model.

# Summary

Using Cox regression, you have found an acceptable model for the time to churn, plotted the expected number of customers retained over the next two years, and identified the individual customers most likely to churn in the next year. Note that while this is an acceptable model, it may not be the best model. Ideally you should at least compare this model, obtained using the Forward stepwise method, with one created using the Backward stepwise method.

Explanations of the mathematical foundations of the modeling methods used in IBM® SPSS® Modeler are listed in the *IBM SPSS Modeler* Algorithms Guide.