* [Classifying Telecommunications Customers (Multinomial Logistic Regression)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_telco_custcat.htm)

# Classifying Telecommunications Customers (Multinomial Logistic Regression)

Logistic regression is a statistical technique for classifying records based on values of input fields. It is analogous to linear regression but takes a categorical target field instead of a numeric one.

For example, suppose a telecommunications provider has segmented its customer base by service usage patterns, categorizing the customers into four groups. If demographic data can be used to predict group membership, you can customize offers for individual prospective customers.

This example uses the stream named telco\_custcat.str, which references the data file named telco.sav. These files are available from the Demos directory of any IBM® SPSS® Modeler installation. This can be accessed from the IBM SPSS Modeler program group on the Windows Start menu. The telco\_custcat.str file is in the streams directory.

The example focuses on using demographic data to predict usage patterns. The target field custcat has four possible values that correspond to the four customer groups, as follows:

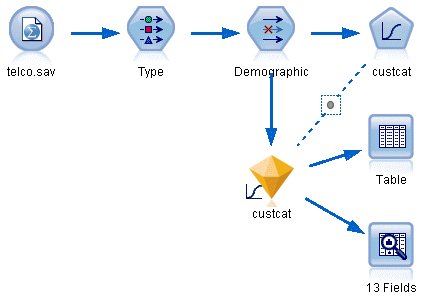
|  |  |
| --- | --- |
| Value | Label |
| 1 | Basic Service |
| 2 | E-Service |
| 3 | Plus Service |
| 4 | Total Service |

Because the target has multiple categories, a multinomial model is used. In the case of a target with two distinct categories, such as yes/no, true/false, or churn/don't churn, a binomial model could be created instead. See the topic [Telecommunications Churn (Binomial Logistic Regression)](http://127.0.0.1:54857/help/topic/com.ibm.spss.modeler.tutorial/clementine/example_telco_churn.htm#example_telco_churn) for more information.

**Building the Stream**

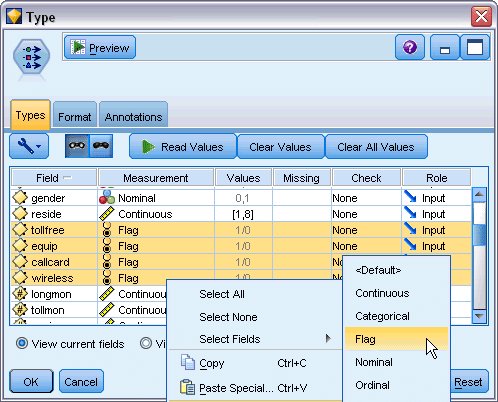
1. Add a Statistics File source node pointing to *telco.sav* in the *Demos* folder.

*Figure 1. Sample stream to classify customers using multinomial logistic regression*



* 1. Add a Type node and click **Read Values**, making sure that all measurement levels are set correctly. For example, most fields with values 0 and 1 can be regarded as flags.

*Figure 2. Setting the measurement level for multiple fields*

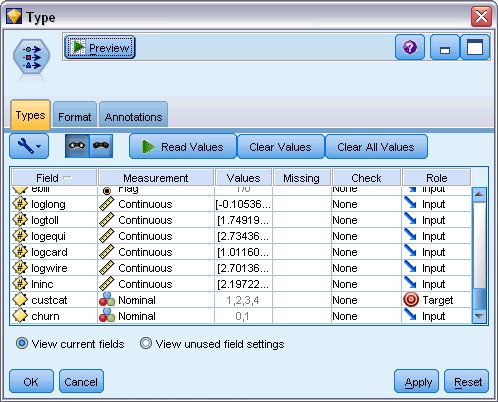


*Tip*: To change properties for multiple fields with similar values (such as 0/1), click the *Values* column header to sort fields by value, and then hold down the shift key while using the mouse or arrow keys to select all the fields you want to change. You can then right-click on the selection to change the measurement level or other attributes of the selected fields.

Notice that *gender* is more correctly considered as a field with a set of two values, instead of a flag, so leave its Measurement value as **Nominal**.

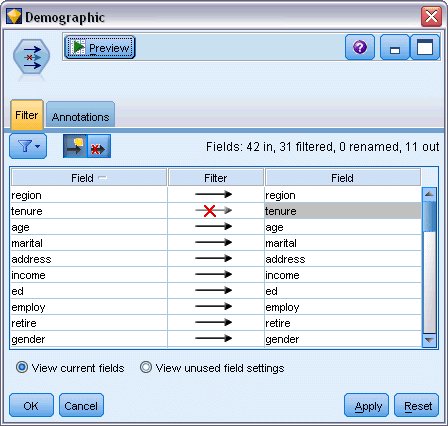
* 1. Set the role for the *custcat* field to **Target**. All other fields should have their role set to **Input**.

*Figure 3. Setting field role*



Since this example focuses on demographics, use a Filter node to include only the relevant fields (*region*, *age*, *marital*, *address*, *income*, *ed*, *employ*, *retire*, *gender*, *reside*, and *custcat*). Other fields can be excluded for the purpose of this analysis.

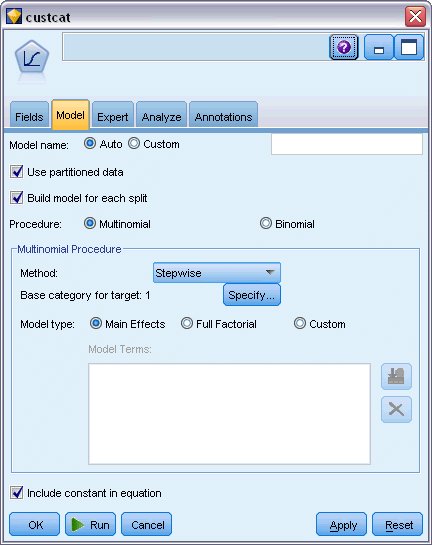
*Figure 4. Filtering on demographic fields*



(Alternatively, you could change the role to **None** for these fields rather than exclude them, or select the fields you want to use in the modeling node.)

1. In the Logistic node, click the **Model** tab and select the **Stepwise** method. Select **Multinomial**, **Main Effects**, and **Include constant in equation**as well.

*Figure 5. Choosing model options*



Leave the Base category for target as 1. The model will compare other customers to those who subscribe to the Basic Service.

1. On the Expert tab, select the **Expert** mode, select **Output**, and, in the Advanced Output dialog box, select **Classification table**.

*Figure 6. Choosing output options*

