Demo: Building a Predictive Model Using the Predictive Regression Models Task

Build a predictive regression model of **SalePrice** from both catagorical and interval predictors. Use **ameshousing3** as the training data set and **ameshousing4** as the validation data set. Use backward elimination with SBC for the training data as the model-building criterion, and choose the model with the smallest average squared error for the validation data set. Create an item store to use in subsequent processing.

---

1. In the Navigation pane, select **Tasks and Utilities**.
2. Expand **Tasks**.
3. Expand **Statistics** and select the **Predictive Regression Models** task.
4. Select the **stat1.ameshousing3** table. **Note:** You'll add the validation data set in a later step.
5. Assign **SalePrice** as the dependent variable.
6. Assign the classification and continuous variables as listed below.

Classification Variables	Continuous Variables
Heating_QC	Lot_Area
Central_Air	Gr_Liv_Area
Fireplaces	Bedroom_AbvGr
Season_Sold	Garage_Area
Garage_Type_2	Basement_Area
Foundation_2	Total_Bathroom
Masonry_Veneer	Deck_Porch_Area
Lot_Shape_2	Age_Sold
House_Style2	
Overall_Qual2	
Overall_Cond2	

7. On the DATA tab, expand **Parameterization of Effects** and verify that **GLM coding** is selected.
8. On the MODEL tab, select **Custom Model** and then click **Edit** to open the Model Effects Builder.
   - Select all of the variables and click **Add**.
   - Verify that the **Intercept** check box is selected.
   - Click **OK**.
9. On the SELECTION tab under MODEL SELECTION, select **Backward elimination** in the Selection method drop down list, and under Add/remove effects with, select **Schwarz Bayesian information criterion**.
10. Expand **SELECTION PLOTS** and select **Criteria plots** and **Coefficient plots**.
11. Expand **DETAILS** and then expand **Model Effects Hierarchy**. Under Model effects hierarchy, select **Do not maintain hierarchy of effects**. The default is Maintain hierarchy of effects.
12. Click the **Edit** button in the code window to open the editor, and make the following changes manually:
    - Add the **valdata=** option to the PROC GLMSELECT statement to specify **stat1.ameshousing4** as the validation data set. **Note:** Currently SAS Studio does not include the option to specify a separate validation data set.

- Add **choose=validate** within the parentheses containing **select=sbc** in the MODEL statement to use the average squared error of the validation data set as the model selection tool.
- Add the **ref=first** option to the CLASS statement to treat the first level of each variable in the classification variable list as the reference level.
- Add a STORE statement to save the analysis results in a SAS item store, **stat1.amesstore**.
13. Submit the code.

Generated Code

```
ods noproctitle;
ods graphics / imagemap=on;

proc glmselect data=STAT1.AMESHOUSING3 plots=(criterionpanel coefficientpanel)
               valdata=stat1.ameshousing4;
   class Heating_QC Central_Air Fireplaces Season_Sold Garage_Type_2 Foundation_2
                Masonry_Veneer Lot_Shape_2 House_Style2 Overall_Qual2 Overall_Cond2 /
                param=glm ref=first;
   model SalePrice=Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
                Total_Bathroom Deck_Porch_Area Age_Sold Heating_QC Central_Air Fireplaces
                Season_Sold Garage_Type_2 Foundation_2 Masonry_Veneer Lot_Shape_2
                House_Style2 Overall_Qual2 Overall_Cond2 / selection=backward
(select=sbc choose=validate) hierarchy=none;
store out=stat1.amesstore;
run;
```