

## Practice: Building a Predictive Model Using the Predictive Regression Models Task

Use the **ameshousing3** data set to build a model that predicts the sale prices of homes in Ames, Iowa, that are 1500 square feet or below, based on various home characteristics.

1. Use the Predictive Regression Models task to partition **ameshousing3** into a training data set of approximately 2/3 and a validation data set of approximately 1/3. Use 8675309 as the random seed, and **Reference Coding** for Parameterization of Effects. Specify **SalePrice** as the dependent variable, and assign the classification and continuous variables as listed below. Add all variables as model effects.

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	

Use stepwise regression as the selection method, Akaike's information criterion (AIC) to add and or remove effects, and average squared error for validation data to select the best model. Produce criteria and coefficient plots. Edit the generated code to add the REF=FIRST option after PARAM=REF in the CLASS statement.

- a. In the Navigation pane, select **Tasks and Utilities**.
- b. Expand **Tasks**.
- c. Expand **Statistics** and select the **Predictive Regression Models** task.
- d. Select the **stat1.ameshousing3** table.
- e. Expand **Partition Data** and select **Validation Data**.
- f. In the Identify validation or test data cases drop-down list, select **Specify a sample proportion**, and enter **0.3333** as the **Proportion of validation cases** value.
- g. Select the **Specify the random seed** check box, and enter **8675309** as the **Random seed** value.
- h. Assign **SalePrice** to the Dependent variable role.
- i. Assign **Heating\_QC**, **Central\_Air**, **Fireplaces**, **Season\_Sold**, **Garage\_Type\_2**, **Foundation\_2**, **Masonry\_Veneer**, **Lot\_Shape\_2**, **House\_Style2**, **Overall\_Qual2**, and **Overall\_Cond2** to the Classification variables role.
- j. Assign **Lot\_Area**, **Gr\_Liv\_Area**, **Bedroom\_AbvGr**, **Garage\_Area**, **Basement\_Area**, **Total\_Bathroom**, **Deck\_Porch\_Area**, **Age\_Sold** to the Continuous variables role.
- k. Expand **Parameterization of Effects**. In the Coding drop-down list, select **Reference Coding**.

- l. On the MODEL tab, select **Custom Model** and then click **Edit** to open the Model Effects Builder.
- m. Select all variables, and click **Add** to add them to the model. Click **OK**.
- n. On the SELECTION tab under MODEL SELECTION, use the Selection method drop-down list to select **Stepwise regression**
- o. Use the Add/remove effects with drop-down list to select **Akaike's information criterion (AIC)**.
- p. Use the Select best model by drop-down list to select **Average square error for validation data**.
- q. Expand **SELECTION PLOTS** and select **Criteria plots** and **Coefficient plots**.
- r. Click the **EDIT** button in the CODE window and add the option **ref=first** after the **param=ref** option in the CLASS statement.
- s. Run the code.

Here are the [results](#).

2. Which model did PROC GLMSELECT choose?

PROC GLMSELECT chose the model at Step 10, which has the following effects: **Intercept, Basement\_Area, Gr\_Liv\_Area, Age\_Sold, Garage\_Area, Overall\_Cond2, Fireplaces, Overall\_Qual2, House\_Style2, Deck\_Porch\_Area, and Heating\_QC**.

3. Resubmit the modified code. Do not make any changes to it. Does it produce the same results as before?

The results are the same. Every time you run a specific PROC GLMSELECT step using the same seed value, the pseudo-random selection process is replicated and you get the same results.

4. Return to the CODE window that contains the edited code. In the PROC GLMSELECT statement, change the value of SEED= and submit the modified code. Does it produce the same results as before?

Because you used a different seed, the results are almost certainly different from the previous results.

Hide Solution