Practice: Using the SCORE Statement in PROC GLMSELECT

Re-create the model that was built in the previous practice with a few changes. Create an item store, and then use the item store to score the new cases in **ameshousing4**. You'll use code to score the data in two different ways (using PROC GLMSELECT and PROC PLM) and compare the results.

1. Build the model.
    a. In the Navigation pane, select **Tasks and Utilities**.
    b. Expand **Tasks**.
    c. Expand **Statistics** and select the **Predictive Regression Models** task.
    d. On the DATA tab, 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 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	

    j. Expand **Parameterization of Effects**. In the Coding drop-down list, select **Reference Coding**.
    k. On the MODEL tab, select **Custom Model** and then click **Edit** to open the Model Effects Builder.
    l. Select all the variables, and click **Add** to add them to the model. Click **OK**.
    m. On the SELECTION tab under MODEL SELECTION, use the Selection method drop-down list to select **Stepwise regression**.
    n. Use the Add/remove effects with drop-down list to select **Akaike's information criterion** (AIC).
    o. In the Select best model by drop-down list, select **Average square error for validation data**.
    p. Expand **SELECTION PLOTS** and select **Criteria plots** and **Coefficient plots**.
    q. On the SCORING tab, select **Save scored data** and enter the name **score1** for the output data set. It will be created in the **Work** library.
    Notice the **Save scoring code** check box. You won't use the scoring code in this practice, but you could select this check box and browse to a location where you can store files. It will create a file with scoring code that you can include in a DATA step to score the item store.
    r. Click the **EDIT** button in the CODE window and make the following changes:
        1. Add the option **ref=first** after the **param=ref** option in the CLASS statement.
        2. Add **data=stat1.ameshousing4** in the SCORE statement.
        3. Add a STORE statement to create an item store named **store1** in the **Work** library.

4. You don't need to examine the results, so you can remove the ODS statement and add the NOPRINT option in the PROC GLMSELECT statement.

s. Submit the code. Check the log to verify that the item store, **work.store1**, was created, and that the data set, **work.score1**, was created with 300 observations and 34 variables.

Generated Code

```
ods noproctitle;

proc glmselect data=STAT1.AMESHOUSING3 plots=(criterionpanel coefficientpanel)
                seed=8675309 noprint;
        partition fraction(validate=0.3333);
        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=ref ref=first;
        model SalePrice=Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
                Total_Bathroom Overall_Qual 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=stepwise
        (select=aic) hierarchy=single;
        score out=WORK.score1 predicted residual data=stat1.ameshousing4;
        code;
        store out=work.store1;
run;
```

2. Write a PROC PLM step to process the **store1** item store. Score the data in **ameshousing4**. Create an output data set named **score2**. Submit the code, and check the log to verify that **work.score2** was created with 300 observations and 33 variables.

```
proc plm restore=store1;
    score data=STAT1.ameshousing4 out=score2;
run;
```

Here are the results.

3. Write a PROC COMPARE step to compare the scoring results from PROC GLMSELECT and PROC PLM. Use a criterion of 0.0001 for the comparison.

```
proc compare base=score1 compare=score2 criterion=0.0001;
    var P_SalePrice;
    with Predicted;
run;
```

4. Submit the PROC COMPARE step and examine the resullts. Does the PROC COMPARE output indicate any differences between the predictions produced by the two scoring methods?

Here are the results. As shown in this output, the two scoring methods produce the same predictions.

Note: Depending on the version of SAS and SAS/STAT that you are using, your results might look somewhat different from the output shown here. However, the results should indicate that these data sets do not differ.

Hide Solution