Demo: Calculating Collinearity Diagnostics Using the Linear Regression Task

Use the Linear Regression task to investigate the correlations between the variable **score** and the other interval variables. First combine the **score** data from the other research group with the data that we already have. Then further assess the collinearity problem and identify the predictors that are involved in the problem.

1. Run the code below. This combines the data from the other research group with the data that we've been analyzing.

```
%let interval=Gr_Liv_Area Basement_Area Garage_Area Deck_Porch_Area
         Lot_Area Age_Sold Bedroom_AbvGr Total_Bathroom ;

/*st105d03.sas*/  /* Part A*/
proc sort data=STAT1.ameshousing3;
   by PID;
run;
proc sort data=STAT1.amesaltuse;
   by PID;
run;

data amescombined;
   merge STAT1.ameshousing3 STAT1.amesaltuse;
   by PID;
run;
```

2. In the Navigation pane, select **Tasks and Utilities**.
3. Expand **Tasks**.
4. To investigate the correlations, expand **Statistics** and open the **Correlation Analysis** task.
5. Select the **work.amescombined** table.
6. Assign the interval variables (**Lot_Area**, **Gr_Liv_Area**, **Bedroom_AbvGr**, **Garage_Area**, **Basement_Area**, **Total_Bathroom**, **Deck_Porch_Area**, and **Age_Sold**) to the Analysis variables role.
7. Assign **score** to the Correlate with role.
8. Click **Run**.

Generated Code

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

proc corr data=WORK.AMESCOMBINED pearson nosimple noprob plots=none;
   var Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
           Total_Bathroom Deck_Porch_Area Age_Sold;
   with score;
run;
```

1. Expand **Statistics** and open the **Linear Regression** task.
2. Select the **stat1.ameshousing3** table.
3. Assign **SalePrice** to the Dependent variable role.
4. Assign the interval variables (**Lot_Area**, **Gr_Liv_Area**, **Bedroom_AbvGr**, **Garage_Area**, **Basement_Area**, **Total_Bathroom**, **Deck_Porch_Area**, and **Age_Sold**) and the variable **score** to the Continuous variables role.

5. On the MODEL tab, click the **Edit this model** icon, select all variables, and click **Add**. Then click **OK**.
6. On the OPTIONS tab, under STATISTICS, use the drop-down list for Display statistics and select **Default and selected statistics**.
7. Expand **Collinearity** and select the option to display **Variance inflation factors**.
8. Suppress all plots by clearing the check boxes under **Diagnostics and Residual Plots** and **Scatter Plots**.
9. Click **Run**.

Generated Code

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

proc reg data=STAT1.AMESHOUSING3 alpha=0.05 plots=none;
   model SalePrice=Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
             Total_Bathroom Deck_Porch_Area Age_Sold score / vif;
run;
quit;
```

Remove **score** from the model and rerun the task.

1. On the DATA tab, select score from the list of Continuous variables, and click the **Remove column** icon.
2. Click **Run**.

Generated Code

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

proc reg data=STAT1.AMESHOUSING3 alpha=0.05 plots=none;
   model SalePrice=Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
             Total_Bathroom Deck_Porch_Area Age_Sold  / vif;
run;
quit;
```