Demo: Looking for Influential Observations Using the Linear Regression Task

Use the Linear Regression task to look for influential observations in the **ameshousing3** data set.

1. In the Navigation pane, select **Tasks and Utilities**.
2. Expand **Tasks**.
3. Expand **Statistics** and open the **Linear Regression** task.
4. Select the **stat1.ameshousing3** table.
5. Assign **SalePrice** to the Dependent variable role.
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 Continuous variables role.
7. On the MODEL tab, use the Model Effects Builder to specify the appropriate model. Click the **Edit this model** icon, select all variables, and click **Add**. Click **OK**.
8. On the OPTIONS tab, expand **Diagnostic and Residual Plots** and clear the check boxes for **Diagnostic plots** and **Residuals for each explanatory variable**.
9. Expand **More Diagnostics Plots** and select all four check boxes. This will display diagnostic plots with labels for influential observations.
10. Expand **Scatter Plots** and clear the check box for **Observed values by predicted values**.
11. On the SELECTION tab, use the Selection method drop-down list and choose **Stepwise selection**.
12. For the Add/remove effects with value, choose **Significance level**.
13. On the CODE tab, click the **Edit SAS code** icon.
14. In the PROC REG step, enter **cooksd** within the parentheses where the plots are listed.
15. Add **COOKSDPLOT** to the list in the ODS SELECT statement.
16. Add the following code after the ODS SELECT statement to write the data from the influence plots into data sets:

```
ods output RStudentByPredicted=Rstud
          COOKSDPLOT=Cook
          DFFITSPLOT=Dffits
          DFBETASPANEL=Dfbs;
```

17. Click **Run**.

Generated Code

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

proc glmselect data=STAT1.AMESHOUSING3 outdesign(addinputvars)=Work.reg_design
            plots=(criterionpanel);
   model SalePrice=Lot_Area Gr_Liv_Area Bedroom_AbvGr Garage_Area Basement_Area
               Total_Bathroom Deck_Porch_Area Age_Sold / showpvalues selection=stepwise
        (slentry=0.05 slstay=0.05 select=sl);
run;

proc reg data=Work.reg_design alpha=0.05 plots(only label)=(rstudentbypredicted cooksd dffits dfbetas);
   ods select RStudentByPredicted DFFITSPlot DFBETASPanel COOKSDPLOT;
   ods output RStudentByPredicted=Rstud
            COOKSDPLOT=Cook
            DFFITSPLOT=Dffits
            DFBETASPANEL=Dfbs;
   model SalePrice=&_GLSMOD /;
   run;
quit;
```

```
proc delete data=Work.reg_design;
run;
```