Print

Practice 2.4 (Level 2): Identifying Outlying Values

Task

In this practice, you modify your PROC SGPLOT step to identify outlying values.

Note: Before you do this practice, you must run the code for practice 2.3 in the same SAS session.

Reminder: Make sure you've defined the **mydata** library.

1. Modify the PROC SGPLOT code from step 2 of practice 2.3 by specifying the DATALABEL= option to identify outlying values. Label the points by **Model**. For information about the DATALABEL= option, see the SAS documentation for PROC SGPLOT.

```
proc sgplot data=out;
   scatter y=Difference x=Estimate / datalabel=Model;
   xaxis min=0 max=55;
   yaxis min=-20 max=20;
   refline 0;
   title 'DATALABEL=Model';
run;
```

2. Submit the code. Do you see a pattern?

   In the plot, the points with the most extreme values for **Difference** appear to be in the category of luxury cars, such as the Mercedes-Benz 190E, the Infinity Q45, and the Dodge Stealth.

   On your own, you might want to try specifying each of the following variables in the DATALABEL= option: **Price**, **EngineSize**, and **Citympg**. As you'll see, these plots seem to indicate that the model fits poorly for cars in the higher price ranges, with large engine sizes, and in the lower range of fuel efficiency.

Hide Solution

Close