Demo: Performing Post-Processing Analysis Using PROC PLM

Filename: **st103d02.sas**

In the previous two-way ANOVA, we observed a larger seasonal effect for houses with fair and good heating systems, but the p-values weren't adjusted for multiple tests. We'll use PROC PLM to access the item store and make the adjustments without refitting the ANOVA model.

PROC GLM DATA=*SAS-data-set* <*options*>;
  CLASS *variables*;
  MODEL *dependent-variable = independent-effects*;
  LSMEANS *effects* < / *options*>;
  STORE <OUT= *item-store-name***>** < / **LABEL=***'label'*>;
RUN;

PROC PLM RESTORE=*SAS-data-set* <*statistic-keyword(s)*>;
  CLASS *variable(s)* < / *option(s)*>;
  VAR *variable(s)*;
RUN;

1. Open program st103d02.sas.

```
/*st103d02.sas*/  /*Part A*/
ods graphics on;

proc glm data=STAT1.ameshousing3
         order=internal
         plots(only)=intplot;
    class Season_Sold Heating_QC;
    model SalePrice = Heating_QC Season_Sold Heating_QC*Season_Sold;
    lsmeans Heating_QC*Season_Sold / diff slice=Heating_QC;
    format Season_Sold Season.;
    store out=interact;
    title "Model with Heating Quality and Season as Interacting Predictors";
run;
quit;

/*st103d02.sas*/  /*Part B*/
proc plm restore=interact plots=all;
    slice Heating_QC*Season_Sold / sliceby=Heating_QC adjust=tukey;
    effectplot interaction(sliceby=Heating_QC) / clm;
run;

title;
```

We ran the PROC GLM step in Part A earlier and saved the results in a temporary item store. We're in the same SAS session, so the item store is still available. In the PROC PLM statement in Part B, the RESTORE= option specifies the item store, interact. The PLOTS= option produces all the available ODS plots for the statements that we include in the step. This includes an effect plot by default, but we've added an EFFECTPLOT statement to request interaction plots sliced by heating quality, and the CLM option to request confidence limits for the means. The SLICE statement requests tables for the interaction term, Heating_QC crossed with Season_Sold, sliced by the different levels of heating quality. The ADJUST=TUKEY option will

adjust the p-values for multiple comparison tests. Remember that the SLICEBY= syntax in the SLICE statement is different from SLICE= in the LSMEANS statement.

2. Submit the PROC PLM step in Part B.

3. Review the output.

   The Store Information table describes the item store, including its name, location, the data set from which it was created, the procedure that was used to create it, the response and class variables, and the model effects. This is followed by a Class Level Information table and a series of F tests, least squares means, and diffograms. There's one set for each heating quality slice. The diffograms were produced because we specified PLOTS=ALL and a SLICE statement.

   Let's look at the first slice, where the heating quality is excellent. In the overall F test for Season_Sold by Heating_QC, the p-value is the same as in the PROC GLM results. The least squares means table shows all the pairwise comparisons of Season_Sold within the Heating_QC level, excellent. Here we get the unadjusted p-values as well as the Tukey adjusted p-values.

   The diffogram shows there are no significant differences among SalePrice means when we hold Heating_QC constant at excellent. This slice shows no significant effect of Season_Sold, so let's jump to the analysis for fair heating systems. Looking at the pairwise comparisons of Season_Sold within the Heating_QC level, fair, we see that the only statistically significant pairwise comparison is between summer and winter. The blue line in the corresponding diffogram indicates a significant difference in the mean SalePrice of fair-heating-quality properties that were sold in the summer versus the winter. Houses with fair heating quality sell for more than $70,000 less in the winter than in the summer, on average.

   What about good heating systems? There was a significant mean sale price difference between spring and summer months. The diffogram also shows that the spring versus summer is statistically different. The diffogram for the typical/average heating quality shows no significant differences among the pairs of seasons. Finally, the fit plot, produced by the EFFECTPLOT statement with the CLM option, includes the confidence limits for the means. We can see larger confidence intervals for the sales price means that were based on small sample sizes.

   Remember that it was through our initial plot that we recognized an interaction effect could potentially improve our model. So, you should always begin your analyses by visually investigating the data. By adding an interaction effect to our two-way ANOVA, we could explain a greater proportion of the variation in sale price. But finding a significant interaction is not the end of your work. You can explore the nature of the interaction through additional graphics, differences of least squares means, and tests of simple effects. In doing so, we saw that seasonal differences in sales prices do exist, but only for houses with fair or good heating quality.

Close