Refer to the Real Estate Sales data set (Appendix C, Data Set C.7).

Use Sales Price (variable 2) as the response variable.

Use the following variables as factors:
  Quality (variable 10)
  Style1 (dichotomized version of variable 11: 1=Style 1; 0=all other styles)

Use the following variables as covariates:
  Finished Square Feet (variable 3)
  Lot Size (variable 12)

1. Using SAS PROC GLM, fit the ANCOVA model that includes the main effects Quality and Style1, the interaction effect between Quality and Style1, and the covariates Finished Square Feet and Lot Size. Present the ANOVA table.

2. Check the data for potential influential observations. (For the purpose of consistency, assume we can find good reasons to eliminate any observations with |RStudent|>4, Cook’s D>1.0, CovRatio>2.0 or Leverage>0.5.)

3. Check the normality of the residuals. If you determine that the residuals do not follow a normal distribution, discuss the consequences.
4. Create a new variable CELL, which takes on values 1 – 6, corresponding to the 6 possible combinations of levels of Quality and Style1. Plot the residuals vs. CELL. Interpret.

5. Using the variable CELL created in part 4, test whether the variances of the adjusted residuals are equal across all 6 cells.

6. With $\alpha=0.05$,
   a) Determine if the covariates Finished Square Feet and Lot Size are needed in the model. Remove any of these variables that are not significant (one at a time, in a sequential fashion), and refit the model with only the significant covariate(s), if any.

   b) Determine whether there is a significant interaction between Quality and Style1.

   If there is (meaning a p-value < 0.05), create a plot of the adjusted cell means which graphically illustrates the nature of this interaction. Interpret this plot.

   If there is not, remove this interaction effect from the model, and refit with only main effects and any significant covariates.

   c) Determine whether there are any significant main effects. If there are, use post hoc tests to determine where the mean differences are. Control the overall error rate using Tukey's method. Interpret any significant results.