Let's look through some data generated by the National Institute of Standards and Technology (NIST) that investigates the strength of ceramics. Use the code 5_homeworkCeramic.sas for the data set in SAS form. For our purposes, we are interested in the three production variables: tableSpeed, feedRate, and gritSize. Additionally, the ceramics are produced in two different batches in eight different laboratories. The batches are coded 1 and 2 to indicate the first batch produced in a given lab and the second batch, respectively.

Let's work through analyzing this data set.

a) A good first step is to go looking for anomalous observations. In this case, that will mainly be in terms of the response. One way of doing this is to look at summary statistics or histogram.

b) Next, let's visually investigate the possibility of a batch or lab effect. A useful tool for this is the boxplot. What do your findings suggest?

c) Now we need to decide what type of study this was (ideally, we would have designed the study ourselves..).

d) Write down a model for this study with pairwise interactions for the main explanatory variables of interest. Include batch and lab, but not as interactions.

e) A helpful diagnostic plot in many instances is called a "run plot". A run plot plots a variable of interest versus measurements that might exhibit correlation between observations. Here, make run plots of the residuals versus id and versus num.

f) Check the assumptions for the model you wrote down in d)

g) …. Continue the analysis reporting your conclusions (not necessarily with the model specified in d). I'm just not explicitly outlining the remaining steps.)

For review (not to be turned in). List of topics for the exam:

General Statistical Knowledge That is Fair Game:

Power / Alpha / Beta / Sample Size / Effect Size
Type 1 Error / Type II Error
Interpretation of Confidence Intervals
Correspondence between Hypothesis Tests and Confidence Intervals.
Multiple comparisons

Chapter 9:

Interpreting coefficients of categorical explanatory variables
Interpreting interactions between categorical and continuous variables.
Modeling with continuous and categorical variables with respect to scientific questions of interest.

Chapter 10:

Doing inference in multiple regression via hypothesis tests or confidence intervals
(this includes extra sums of squares F tests)
Finding confidence/prediction intervals for estimates of the mean of Y as a function of X
(and knowing the difference between them)
Confidence intervals and inference for linear combinations of parameters

Chapter 11:

Be familiar enough with the formulas for Cook's D, dfbetas, leverage, Studentized residual, and R^2 to know how they change with various other quantities
(examples: how does leverage change with n? How does the leverage affect Studentized residuals)
What are strategies for dealing with extreme observations?
Very basic/general questions about weighted regression/nonlinear regression

Chapter 12:

Multicollinearity and VIF

Chapter 13:

Crossed vs. nested design
Additive vs. interaction models
Identifying the Experimental Unit
Strategies involving blocking
(why and how would you use blocking?)
Using the output for lsmeans to draw inferences about differences between particular factor level combinations
What is the relationship between multiple regression and multiway anova?
Profile plots for two-way anova