STAT 540, Fall 2015

Due on **12:00 A.M., Friday, September 25th**

Homework 4

Homework format for all STAT 540 homework this term: Please label all problems clearly and turn in an organized homework assignment. You don't need to spend hours producing beautifully typeset homework, but you won't get credit if we can't find or read your answer. Unless noted otherwise, turn in the following (as appropriate for the problem).

- Theoretical derivation (when asked for).
- Numerical results **with an explanation of your solution**, written in complete sentences. If computer code is absolutely necessary to provide context here, then include it–nicely formatted–within the solution (otherwise, see below).
- Appropriate graphics. Use informative labels, including titles and axis labels. Try to put multiple plots on the page by using, for example, the R command `par(mfrow=c(2,2))`.
- **Only as necessary**: Final clean computer code used to answer the problem **attached to the end of your homework**. Only include the rare code excerpts without which we wouldn't be able to figure out what you did. Annotate your code. Number and order the code in order of the problems. When in doubt, leave it out; consider that we will probably never read it.
- Some problems will be relatively open-ended, such as "Here are some data. Analyze them and write a report." I will provide further instructions about reports later. They should be self-contained, with suitable EDA, graphs, numerical results, and **scientific interpretation**. No computer code should be included. The report should be concise: "no longer than necessary".

(1) Still consider the problem (5) in HW 3 and use the file `snow1.dat` on the class web site, which contains data from a snow gauge calibration study. Answer the following questions. If you have done parts (a) and (b) below in HW3, you can skip them.

(a) These data include multiple observations at each density. Hence it is possible to compute the mean and standard deviation for each density, then use the Box-Cox method to choose a transformation that equalizes the variances. Calculate the mean and standard deviation for each of the 9 densities, then regress log sd (Y) on the log mean (Y). What transformation (if any) does this suggest?

(b) Rightly or wrongly, the investigators decide to use a log transformation. Fit the regression of Y = log gain on X= density. (You don't have to report anything about this regression). Plot the residuals vs. predicted values. Are there any concerns?

(c) Lack of fit is an important concern for these investigators. Since there are repeated measurements for each density, it is possible to use the ANOVA lack of fit test. Is a straight line sufficient to describe the relationship between log gain and density. Report the p-value and a one-sentence conclusion.

(d) Following part (i) of problem (5) in HW3, the investigators consider a third way to do the study. They could use one block from each of the 9 densities and measure each block 10 times. This is a lot cheaper than the designs in either parts 1a or 1b. They tried this design and got the following partial ANOVA tables (Y is log gain and X is density):

Model: $\mathbb{E}(Y_{ij}) = \beta_0 + \beta_1 X_i$		
Source	df.	SS
Model	1	96.507
Error	88	0.416
Total	89	96.923

Model: $\mathbb{E}(Y_{ij}) = \mu_i$		
Source	df	SS
Model	8	96.828
Error	81	0.095
Total	89	96.923

Test for lack of fit of the linear regression of log gain on density using these data. Report your F statistic, p-value and a short conclusion.

(e) The investigators notice that the MSE from the regression in part (d) is very much smaller than 0.06, the MSE from the regression on the original data using 90 blocks. Would you recommend the design in part (d) above? Explain why or why not. Hint: it may help to think about / explain why the two MSE's are so different.

(2) Textbook problems:

(a) Problem 3.15 (Solution concentration): parts (a) and (b) only.

(b) Problem 3.16 (Solution concentration continued). Note: You can ignore the λ values given in the problem and directly use `R` function `boxcox` in the `MASS` library.

(c) Problem 3.27.