STAT 700
Homework 3 Problems
due Wed. Sept. 26

3 Problems. Show all work.
Please follow the Lab report directions off the homework web page for R Problems.
Please work in HW Groups!

1. Consider the linear model from class,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}.$$

Assume that the ε_i are independent $N(0, \sigma^2)$ random variables or equivalently

$$\boldsymbol{Y} \sim N_n(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I}_n).$$

Also, assume that $\boldsymbol{X}'\boldsymbol{X}$ is invertible.

The prediction of a future observation, $Y_0 = \boldsymbol{x}_0'\boldsymbol{\beta} + \varepsilon_0$ at a given vector of independent variables $\boldsymbol{x}_0'$, is given by $\widehat{Y}_0 = \boldsymbol{x}_0'\widehat{\boldsymbol{\beta}}$. Find the expected value, variance, and the distribution of $\widehat{Y}_0$.

2. Consider the linear model from class,

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}.$$

Assume that the ε_i are independent $N(0, \sigma^2)$ random variables.

We will use a dataset on one dependent variable and five independent variables The data is available off the class web page:

`https://edoras.sdsu.edu/~babailey/stat700/test.txt`

Use the R `read.table` command with the `header=T` option.

(a) In R, use matrix and vector operations to find the least estimate $\hat{\boldsymbol{\beta}}$. You will have to make your own $\boldsymbol{X}$ matrix that has a column of 1's. Note: You can check your answer with the results from the R function `lm`. (For help on matrix algebra in R see: `http://www.statmethods.net/advstats/matrix.html`)

Note: You should make sure that the $\boldsymbol{X}$ matrix that you create in R is really a matrix by using the R `is.matrix` command. It should return `TRUE`.

(b) The following quadratic forms were computed (and rounded).

(1) $\boldsymbol{Y'PY} = 404.532$
(2) $\boldsymbol{Y'(I-P)Y} = 0.480$
(3) $\boldsymbol{Y'(P-J/}n\boldsymbol{)Y} = 3.598$
(4) $\boldsymbol{Y'Y} = 405.012$
(5) $\boldsymbol{Y'(I-J/}n\boldsymbol{)Y} = 4.078$
(6) $\boldsymbol{Y'(J/}n\boldsymbol{)Y} = 400.934$

Use matrices and R functions to reproduce the above quadratic forms.

(c) Match the above quadratic form with its corresponding sum of squares: $\text{SS}_{Total(uncorrected)}$, $\text{SS}_{Total(corrected)}$, SS_{Model}, $\text{SS}_{Regression}$, $\text{SS}_{Residuals}$, and SS_{μ}.

Note: SS_{μ} is commonly called the correction factor.

(d) To test if the regression is significant at the $\alpha = 0.1$ level, use R operations and the above computed sum of squares to construct the F test statistic. You can use R to compute a p-value. You can also check your answer by using the summary from the R `lm` function. Be sure to state the null and alternative hypotheses. State your conclusion.

3. We will use the R dataset `chickwts`. We are interested in the effects of different types of feed on the growth rate of chicken. The help file for the dataset gives more details of the experiment.

Please read Lab2 One-Way ANOVA Section

(a) Make boxplots of the data. What do the boxplots suggest about the different types of feed?

(b) Determine if there are differences in the weights of chicken according to their feed. Write down the model that you are using and state the null and alternative hypotheses. State your conclusion. Give diagnostic plots of the residuals.

(c) To test all possible two-group comparisons, use the R function `pairwise.t.test` with the Bonferroni adjustment. What do you conclude? How many pairwise comparisons were made for this problem?

(d) In part (c) we used the Bonferroni adjustment. What is the Bonferroni adjustment and what does it do? (see the R function `p.adjust`).

(e) What are the other methods available to adjust the p-value?