STAT 700, Fall 2011
Homework 4 Problems
due Wednesday September 28

2 Problems. Please follow the Lab report directions off the homework web page.

1. Return to housing demand data. An economist is interested in the relationship between the demand for housing (as measured by housing starts), price, and national disposable income.

It is available off the class web page:
`http://www.rohan.sdsu.edu/~babailey/stat700/housing.dat`

Let $Y$ be the housing demand in appropriate units, $AP$ be a variable representing average price, and $DI$ be a variable representing disposable income. There are $n = 6$ observations. We will consider the model,

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

where $\boldsymbol{\beta}' = (\beta_0 \ \beta_1 \ \beta_2)'$.

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

(a) In R, compute $\boldsymbol{P}$ and $(\boldsymbol{I} - \boldsymbol{P})$ for the housing demand data. Are each of these matrices symmetric and idempotent? Demonstrate the answers with matrix operations.

(b) Now, for some fun. Fill in just the **Sum of Squares** column below by using R matrix and vector operations. Show all your work.

| Source | Sum of Squares | Degrees of Freedom | Mean Squares | $F$ |
|---|---|---|---|---|
| Regression | | | | |
| Residuals | | | | |
| Total (corrected) | | | | |

2. 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.

Turn page over.

(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. 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. Compare the method of Holms with the method of Bonferroni. What do you conclude?

(f) There is a $p$-value adjustment method that controls the false discovery rate. Compare this method with the method of Bonferroni. What does the help file indicate about this method?