

STAT 700, Fall 2011  
Homework 3 Problems  
due Wednesday September 21

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

1. 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,

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

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, use matrix and vector operations to find the least estimate  $\hat{\boldsymbol{\beta}}$ . Note: You can check your answer with the results from the R function `lm`.

The following is the ANOVA table for the model.

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

(b) Fill the above table. It would be fun, but not required to fill it in using R matrix and vector operations.

(c) Test if the regression is significant at the  $\alpha = 0.05$  level. Be sure to state the null and alternative hypotheses. State your conclusion.

2. Two methods,  $A$  and  $B$ , were used in determination of the latent heat of fusion of ice (Natrella, 1963). The investigators wished to find out how much the methods differed. The dataset gives the change in total heat from ice at  $-72^\circ\text{C}$  to water  $0^\circ\text{C}$  in calories per gram of mass.

Turn page over.

It is available off the class web page:

`http://www.rohan.sdsu.edu/~babailey/stat700/ice.dat`

(a) Make boxplots of the data. What do the boxplots suggest about the different methods? Hint: there is a R function `boxplot`.

(b) There is an R function `t.test` and `var.test` to compare the differences in two means and variances. To practice, use `t.test` to compare the means of the two methods. What do you conclude about the difference in the two means?

The `t.test` default is to assume unequal variances. You can test this assumption by using the `var.test` function. What do you conclude about the difference in the two variances? What do you now conclude about the difference in the two means?