

## Spring 16 – AMS256 Homework 1

1. Consider the simple linear regression model  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ ,  $i = 1, \dots, n$ . Let  $z_i = y_i/x_i$ . Write a linear model for  $z_i$ .
2. A design matrix for a multiple regression model takes the form

$$\mathbf{X} = \begin{bmatrix} 1 & -1 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

- (a) What is the rank of  $\mathbf{X}$ ?
  - (b) Give the dimension of  $C(\mathbf{X})$  and a nonzero vector in it.
  - (c) Give the dimension of  $\mathcal{N}(\mathbf{X})$  and a nonzero vector in it.
  - (d) Compute  $\mathbf{X}^T \mathbf{X}$  and find a generalized inverse for it.
  - (e) Let  $\mathbf{P}$  denote the orthogonal projection matrix onto  $C(\mathbf{X})$ . Which of the following vectors could be  $\mathbf{P}\mathbf{y}$  for an appropriate response vector  $\mathbf{y}$ ?  $(3, 1, 1, 2, 2)^T$ ,  $(1, -1, 1, -1, 1)^T$ ,  $(1, 0, 0, 2, 2)^T$ .
  - (f) Using a right-hand side that you gave above find all the solutions to  $\mathbf{X}\boldsymbol{\beta} = \mathbf{P}\mathbf{y}$ .
3. Consider the simple linear regression model  $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ . Show that if  $x_i = s + ti$ , for some values of  $s$  and  $t$ , then  $y_i = \gamma_0 + \gamma_1 i + \epsilon_i$  is an equivalent parameterization.
  4. Consider the model  $y_i = \theta x_i^2 + \epsilon_i$ , where the  $x_i$ s are fixed constants and  $\epsilon_i \sim N(0, \sigma^2)$  for all  $i = 1, \dots, n$  (and they are i.i.d.). Find the LSE of  $\theta$ . Find the MLE of  $\theta$ .
  5. Consider the simple linear regression model,

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \dots, n.$$

Suppose that the response and predictor variables are transformed as follows:  $y_i^* = a + by_i$  and  $x_i^* = c + dx_i$ . Find the LSEs of  $\beta_0^*$  and  $\beta_1^*$  in the model

$$y_i^* = \beta_0^* + \beta_1^* x_i^* + \epsilon_i^*, i = 1, \dots, n.$$

How do these estimates relate to the LSE of  $\beta_0$  and  $\beta_1$ ?

6. Illustrate the partitioning of the sum of squares for simple linear regression by calculating the ANOVA table for the following data. Parents are often interested in predicting the eventual heights of their children. The following are measurements of heights for 8 individuals:

Height (inches) at age 2 ( $x$ )	39	30	32	34	35	36	36	30
Height (inches) as an adult ( $y$ )	71	63	63	67	68	68	70	64

Prove that  $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2 = S_{xy}^2 / S_{xx}$ . Find  $R^2$ .

7. Get the passenger car data from <http://lib.stat.cmu.edu/DASL/Datafiles/carmpgdat.html>
- (a) Fit a linear regression model to predict MPG (miles per gallon) from HP (horse power). Summarize your analysis including a plot of the data with the fitted line.
  - (b) Repeat the analysis but use  $\log(\text{MPG})$  as the response. Compare the analyses.
  - (c) Fit a multiple regression model to predict MPG from the other variables and summarize your analysis.
8. Let  $\hat{\beta}_0$  and  $\hat{\beta}_1$  be the LSEs of  $\beta_0$  and  $\beta_1$  in the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i,$$

with  $E(\epsilon_i) = 0$ , and  $\text{Var}(\epsilon_i) = \sigma^2$  for all  $i = 1, \dots, n$ , and  $\text{Cov}(\epsilon_i, \epsilon_j) = 0$  for all  $i \neq j$ . Let  $y_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0 + \epsilon_0$ , with  $E(\epsilon_0) = 0$  and  $\text{Var}(\epsilon_0) = \sigma^2$ . Find  $E(y_0)$  and  $\text{Var}(y_0)$ .