

STAT 632, Homework 4

Directions: Please submit your completed assignment to Canvas as a **single file**. For the concept questions, your solutions may be typed (using LaTeX or equation editor in Word and converted to PDF), or handwritten and then scanned. For the data analysis questions, which require R, you must type your solutions. I suggest using R Markdown and knitting to PDF/html. Include all R code in your answers to each data analysis question.

Reading: Sections 5.1 and 5.2 from *A Modern Approach to Regression*.

Exercise 1.¹ The multiple linear regression model can be written as $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$, where $\text{Var}(\mathbf{e}) = \sigma^2 \mathbf{I}$ and \mathbf{I} is the $n \times n$ identity matrix. The fitted values are given by

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$

- (a) Show that $\mathbf{H}\mathbf{H}' = \mathbf{H}\mathbf{H} = \mathbf{H}$. Note that a matrix that has this property is called **idempotent**.
- (b) Show that $E(\hat{\mathbf{Y}}) = \mathbf{X}\boldsymbol{\beta}$
- (c) Show that $\text{Var}(\hat{\mathbf{Y}}) = \sigma^2 \mathbf{H}$

Exercise 2. In lecture 10 we showed the variance-covariance matrix for the $(p+1) \times 1$ vector, $\hat{\boldsymbol{\beta}}$, of least squares estimates is given by $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$. Derive the 2×2 variance-covariance matrix for least squares estimates, $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0 \ \hat{\beta}_1)'$, for simple linear regression:

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) \end{pmatrix}$$

Additionally, use your result to verify that $\text{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{SXX} \right)$ and $\text{Var}(\hat{\beta}_1) = \sigma^2 / SXX$, where $SXX = \sum_{i=1}^n (x_i - \bar{x})^2$. [Hint: it might be useful to use the identity $\sum_{i=1}^n (x_i - \bar{x})^2 = \sum x_i^2 - n\bar{x}^2$]

Exercise 3. For this exercise use the **Boston** data set from the **MASS** package (you can read about this data set in the help menu). Consider the multiple linear regression model with **medv** as the response, and **dis**, **rm**, **tax** and **chas** as predictor variables.

- (a) In R, compute the vector of least squares estimates $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Then verify that the results are the same as the parameter estimates provided by the **lm()** function.
- (b) In R, compute the variance-covariance matrix $\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ (plug in $\hat{\sigma}^2 = \text{RSS}/(n-p-1)$ as the estimate for σ^2). Then verify that the square root of the diagonal entries of this matrix are the same as the standard errors provided by the **lm()** function.

¹From *A Modern Approach to Regression with R*, Chapter 6, Exercise 1, with slight modifications