

# Stat 435 HW1

Zach Fechko (011711215)

## General guidelines

This HW covers contents related to simple linear regression and multiple linear regression.

Please show your work in order to get points. Providing correct answers without supporting details does not receive full credits. The homework assignment also trains your programming skills. So, please complete both the conceptual exercises and applied exercises. It is absolutely not OK to just submit your codes only. This will result in a considerable loss of points on your assignments or projects.

For an assignment or project, you DO NOT have to submit your answers or reports using typesetting software. However, your answers must be well organized and well legible for grading. Specifically, if you are not able to knit a .Rmd/.rmd file into an output file such as a .pdf, .doc, .docx or .html file that contains your codes, outputs from your codes, your interpretations on the outputs, and your answers in text (possibly with math expressions), you can organize your codes, their outputs and your answers in a document. Regardless of if you are using typesetting software or not, please organize your work in the format given below:

Problem or task or question ...

Codes ...

Outputs ...

Your interpretations ...

Please upload your answers in a document to the course space where assignment or project is provided.

## Conceptual exercises

1. Consider a simple linear regression.

- (a) Provide the formulae for the *least squares estimate (LSE)* of the coefficients, based on observations  $y_i, i = 1, \dots, n$  for the response  $Y$  and observations  $x_i, i = 1, \dots, n$  for the predictor  $X$ .
- (b) Are these estimated coefficients unbiased?
- (c) Provide the formulae for the standard errors of these estimated coefficients, and state the conditions under which the formulae are valid.

- (d) Explain how the sample variance of the observations  $x_i, i = 1, \dots, n$  for  $X$  affects the standard errors of the estimated intercept and slope.
  - (e) Give a test on if the slope in the model is zero, by stating the null hypothesis, the statistic with its degrees of freedom, and the decision rule. Under what conditions will this test work well?
  - (f) In the simple linear regression model  $Y = \beta_0 + \beta_1 X + \varepsilon$ ,  $\varepsilon$  is the random error term with unknown standard deviation  $\sigma$ . Provide an estimator for  $\sigma$ , and describe how the sample size  $n$  of the observations affects the accuracy of this estimator.
2. Consider a simple linear regression model. State the definition of the  $R^2$  statistic by providing also the definitions of RSS and TSS, and explain its connection with sample correlation and the LSE of the slope.
  3. Consider a multiple linear regression model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon,$$

where  $E(\varepsilon) = 0$  and  $Var(\varepsilon) = \sigma^2$ ,  $\beta_0$  is the intercept, and  $\beta_1, \dots, \beta_p$  are the regression coefficients. Suppose we want to test that a particular subset of  $q$  of the  $p$  regression coefficients are zero.

- (a) Provide a test statistic for this by describing in detail each quantity that makes up the statistic.
  - (b) Under what conditions will the test statistic work well?
4. Question 3(a) of Section 3.7 (the section for exercises) of the Textbook.

## Applied exercises

1. Exercise 9 of Section 3.7 of the Textbook.
2. Exercise 10 of Section 3.7 of the Textbook.
3. Exercise 14 of Section 3.7 of the Textbook.