CMSE 381: HW3

1 (10 pts) Exercise 3.7.3

2 (8 pts) Exercise 3.7.4

3 (10 pts) Exercise 3.7.9 (Don't do (d)). Please read through chapter 3.6 before doing those applied questions.

4 (12 pts) Exercise 3.7.10 (a, b, c, d, e, f)

5 (10 pts) Exercise 3.7.13

6 (10 pts) Exercise 3.7.14 (a, b,c, d, e, f)

7 (20 pts) We assume $Y = X^T\beta + \epsilon$ with $\epsilon \sim N(0, \sigma^2)$. We collect a set of training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$ and let $\mathbf{Y} = (y_1, \ldots, y_n)^T$ and $\mathbf{X} = (x_1, \ldots, x_n)^T$, where $x_i = (1, x_{i1}, \ldots, x_{ip})^T$. We want to fit a multiple linear regression.

   a. Derive the $\hat{\boldsymbol{\beta}}$ minimizing the RSS, where $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \ldots, \hat{\beta}_p)^T$.
   b. Prove that this $\hat{\boldsymbol{\beta}}$ is an unbias estimator for the true $\boldsymbol{\beta}$.
   c. Derive Var($\hat{\boldsymbol{\beta}}$).
   d. Let $p = 1$, and using (c.) to prove the equation (3.8). Hint: you may need the formula for the inverse of a 2x2 matrix.
   e. Prove that $\boldsymbol{\epsilon} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$ is orthogonal to the column space of $\mathbf{X}$. Namely,

$$\boldsymbol{\epsilon}^T\mathbf{X} = \mathbf{0}.$$

8 (20 pts) We will now use simulation to test the confidence interval and prediction interval. We assume the data are generated according to

$$Y = \beta_0 + X\beta_1 + \epsilon, \quad \epsilon \sim N(0, 0.2^2).$$

You can using the following code to simulate a training set of 100 data with fixed $X = (x_1, \ldots, x_{100})^T$, a new response variable Y and the expection of Y for $X = 1$. set.seed(36)
$X = rnorm(100)$ # we will fix this X
$Y = 1 + 0.2 * X + 0.2 * rnorm(100)$
$Y.new = 1 + 0.2 * 1 + 0.2 * rnorm(1)$
$F.X = 1 + 0.2 * 1$ # this is expectation of Y when $X = 1$

a [2 pts] Using this training data to get the 90% confidence interval and prediction interval for $X = 1$, and check whether the confidence interval contains $F.X$ and whether the prediction interval contains $Y.new$.

b [10 pts] Repeat the procedure 2000 times for generating training data, $Y.new$, $F.X$, 90% confidence interval and prediction interval. Report how many time the confidence interval contains $F.X$ and how many time the prediction interval contains $Y.new$.

c [8 pts] Using the same setting to generate 50 different training datasets, and generate two plot similar to Fig 3.3 in the textbook.

9 (Challenging problem, not required) Another important property for the least squares estimates (the one in Q7, and we denote $\hat{\boldsymbol{\beta}}$ as $\hat{\boldsymbol{\beta}}^{\mathrm{ols}}$) is stated in the Gauss-Markov Theorem. We first introduce some concepts:

- An estimator that is a linear function of $\mathbf{Y} = (y_1, \ldots, y_n)^T \in \mathbb{R}^n$ is said to be a **linear estimator**.
- A linear estimator $c^T\mathbf{Y}$ is an unbiased estimator of $a^T \boldsymbol{beta}$ if and only if $E(c^T\mathbf{Y}) = a^T\boldsymbol{\beta}$ for $\forall \boldsymbol{\beta} \in \mathbb{R}^p$.

Theorem 1 (Gauss-Markov) *Based on the linear model assumption $y = x^T\boldsymbol{beta}^* + \epsilon$, among all the linear unbiased estimators for $a^T\boldsymbol{\beta}^*$ (for example $\tilde{\boldsymbol{\theta}} = c^T\mathbf{Y}$ linear in terms of $\mathbf{Y}$), $a^T\hat{\boldsymbol{\beta}}^{ols}$ has the smallest variance. Namely, for $\forall c \in \mathbb{R}^n$ such that $E(c^T\mathbf{Y}) = a^T\boldsymbol{\beta}$, we have*

$$Var(a^T\hat{\boldsymbol{\beta}}^{ols}) \leq Var(c^T\mathbf{Y}).$$

Why is this important? Because it says that our predictions $x_0^T\hat{\boldsymbol{\beta}}$ at any $x_0 \in \mathbb{R}^p$ will be unbiased for the true mean $x_0^T\boldsymbol{\beta}^*$ at x_0. Thus, the Gauss-Markov Theorem ensures that among all the linear unbiased predictions, $x_0^T\hat{\boldsymbol{\beta}}^{\mathrm{ols}}$ results the smallest variance. We also call $\hat{\boldsymbol{\beta}}_{\mathrm{ols}}$ the Best Linear Unbiased Estimator (BLUE). Prove this theorem.