STAT 3008: Applied Regression Analysis
2019-20 Term 2
Assignment #1

Due: March 6th, 2020 (Friday) at 5:30pm

This assignment covers material from Chapter 1 and Section 2.1-2.3 of the lecture notes.
You need to show your calculation in details order to obtain full scores.
Please include the R-codes and R-outputs of Problems 3(a) and3(b).

Problem 1 [35 points]: Suppose the following regression model is fitted to a data set with observations $\{(x_i, y_i), i = 1, 2, \ldots, n\}$:

$$y_i = \beta x_i^2 + e_i, \quad e_i \overset{iid}{\sim} N(0, \sigma^2)$$

(a) [11 points] Based on the least squares method and the fact that $\boxed{\text{RSS}/\sigma^2 \sim \chi^2_{n-1}}$ (*df = n-1 since df = n from the data and df = 1 from estimating* β), compute the least squares estimates for β and σ^2.

(b) [5 points] Is $\hat{\beta}$ an unbiased estimator for β? Verify.

(c) [4 points] Show that the regression line passes through the point

$\left(\left[\overline{x^4}\right]^{1/2}, \overline{x^2 y}\right) = \left(\left[\frac{1}{n}\sum_{i=1}^n x_i^4\right]^{1/2}, \frac{1}{n}\sum_{i=1}^n x_i^2 y_i\right)$, but does NOT pass through the point $(\bar{x}, \bar{y})$.

(d) [7 points] Derive the maximum likelihood estimates (MLE) $\tilde{\beta}$ and $\tilde{\sigma}^2$.

(e) [8 points] Suppose $(x_1, x_2, x_3, x_4, x_5) = (-2, -1, 0, 1, 2)$ and $(y_1, y_2, y_3, y_4, y_5) = (8, 1, 1, 3, 9)$. What the values of the least squares estimates $\hat{\beta}$ and $\hat{\sigma}^2$? Does the sum of residuals equal to zero?

Problem 2 [10 points]: Consider the residuals $\{\hat{e}_i\}$ from a simple linear regression

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \qquad i = 1, 2, \ldots, n$$

where $\hat{\beta}_1 = \text{SXY}/\text{SXX}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ are the OLS estimates for β_1 and β_0.

Show that $\{\hat{e}_i, i=1,2,\ldots n\}$ are uncorrelated with the explanatory variables $\{x_i, i= 1,2,\ldots n\}$.

That is, $$\hat{\rho}(x, \hat{e}) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \bar{x})(\hat{e}_i - \bar{\hat{e}}) = 0.$$

Problem 3 (R problem) [20 points]: The *R* library 'alr3' contains the "brains" data, which provided the average body weight (in kg) and average brain weight (in gram) for 62 species of mammals. *(https://rdrr.io/rforge/alr3/man/brains.html)*
Suppose that we are interested in how the log of average brain weight (*y=log(brains$BrainWt)*) is affected by the log of average body weight (*x=log(brains$BodyWt)*).

(a) [10 points] Based on the R codes similar to those from Ch2 page 23, obtain the OLS estimates $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}^2$.

(b) [6 points] Based on the *plot* and the *abline* functions as in Ch1 page 27, generate the scatterplot of the data, and add the regression line obtained in part (a) to the plot.

(c) [4 points] Suppose an outlier is defined as observation (x_i, y_i) with $|\hat{e}_i| > 2\hat{\sigma}$. Do you think there is outlier in the data set? Verify.
(***Note:*** *A more precise definition of outlier will be introduced in Chapter 7, which removes the impact of the outlier (x_i, y_i) itself when estimating* $\hat{\sigma}$).

Problem 4 [35 points]: Suppose we want to fit the our data $\{(x_i, y_i), i=1, 2, \ldots n\}$ based on the following simple linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + e_i, \text{ with } \mathrm{E}(e_i) = 0, \ \ \mathrm{Var}(e_i) = \sigma^2 \text{ and } \{e_i, \ i = 1, \ldots, n\} \text{ are uncorrelated}$$

Given that $n = 11$, $\bar{x} = 73.14545$, $\bar{y} = 3.95455$, $\sum_{i=1}^{n} x_i^2 = 60961.94$, $\sum_{i=1}^{n} y_i^2 = 202.25$, $\sum_{i=1}^{n} x_i y_i = 3373.75$.

(a) [8 points] Compute SXY, SXX and SYY.

(b) [6 points] Show that the OLS estimates $\hat{\beta}_1$ = 0.09100. What are the OLS estimates for β_0 and σ^2?

(c) [4 points] Compute $\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_0 \mid \mathbf{X})$ and $\hat{\mathrm{V}}\mathrm{ar}(\hat{\beta}_1 \mid \mathbf{X})$.

(d) [5 points] Suppose (x, y) = (74.5, 2.0) is an observation in the data set. Based on the definition of outlier as in Problem 3(c), do you think the observation is an outlier? Explain.

[part(e) and (f)]] Suppose the point (50.3, 3.0) is added to the data set, and the new OLS estimates $\hat{\beta}_0^*$, $\hat{\beta}_1^*$ and $(\hat{\sigma}^*)^2$ are obtained from the 12 observations.

(e) [8 points] Show that $\hat{\beta}_1^*$ = 0.08190.

(f) [4 points] What are the values of OLS estimate $\hat{\beta}_0^*$ and $(\hat{\sigma}^*)^2$?

\- **End of the Assignment** -