SS3859A/9859A Fall 2019

Assignment 2

Due: Oct 23 by 11:59 p.m.

Total: 20 pt

Note: Please use either MS word or PDF format for your submission.

1. For problem 1, run the following R codes and use the $mtcars2$ dataset to answer the questions (a)-(d).

R codes:

```
set.seed(50)
idx <- sample(32,25,replace=FALSE)
mtcars2 <- mtcars[idx,]
mtcars2$cyl <- as.factor(mtcars2$cyl)
```

The $mtcars2$ data set contains 25 observations. Consider a regression model where the response is mpg and two predictors are weight and cylinder. Note that this time, we treat the cylinder predictor as a **categorical** variable with three categories: (4,6,8). Use the MLR model: $Y_i = \beta_0 + \beta_{wt}x_{i1} + \beta_{cyl6}w_{i1} + \beta_{cyl8}w_{i2} + \epsilon_i$ where x_{i1} is weight, w_{i1} is 1 if cyl=6 and 0 otherwise, and w_{i2} is 1 if cyl=8 and 0 otherwise.

(a) Obtain the fitted value of mpg at weight = 3, cylinder = 6. (1 pt)

(b) Is cyl an important predictor given that wt is used as a predictor? Answer by conducting an appropriate test at $\alpha = 0.05$. (1 pt)

Suppose we wonder if there is a significant interaction between the weight and cylinder predictors. Now consider a larger model $Y_i = \beta_0 + \beta_{wt}x_{i1} + \beta_{cyl6}w_{i1} + \beta_{cyl8}w_{i2} + \beta_{wt:cyl6}x_{i1}w_{i1} + \beta_{wt:cyl8}x_{i1}w_{i2} + \epsilon_i$. Answer (c) and (d) using this model.

(c) Obtain the fitted value of mpg at weight = 3, cylinder = 8. (1 pt)

(d) Test the null hypothesis: "There is no significant interaction effect between two predictors." Use the significance level $\alpha = 0.05$. (1 pt)

2. For problem 2, import the data in R from
https://raw.githubusercontent.com/hgweon2/ss3859/master/hw2-data-1.csv
The data set contains 100 observations with 4 variables: y (response), $x1$, $x2$ and $x3$. Consider the MLR model: $Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1} x_{i2} + \beta_5 x_{i1} x_{i3} + \beta_6 x_{i2} x_{i3} + \beta_7 x_{i1} x_{i2} x_{i3} + \epsilon_i$.

(a) Given $x2 = 50$ and $x3 = 7$, one unit increase in $x1$ increases the estimated mean of y by A units. Find A. (2 pt)

(b) Obtain the residual plot and normal QQ plot. Check the linearity, equal variance and normality assumptions. (1 pt)

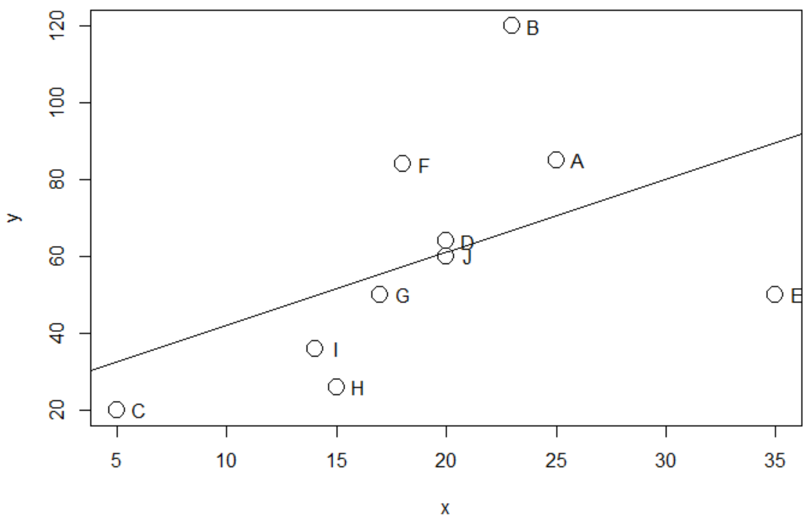
(d) Was the three-way interaction term needed? Why/why not? (1 pt)

(e) Test the null hypothesis: $\beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$ at $\alpha = 0.05$. (2 pt)

3. For problem 3, import the data in R from:
https://raw.githubusercontent.com/hgweon2/ss3859/master/hw2-data-2.csv
The data set contains 100 observations with 2 variables: y (response) and x (predictor). Consider the SLR model: $Y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$. Obtain the fitted model. Using the residual plot and the normal QQ plot, check the linearity, normality and equal variance assumptions. (Justify your answer). Interpret the test results. (3 pt)

4. Repeat problem 3. This time, import the data from:
https://raw.githubusercontent.com/hgweon2/ss3859/master/hw2-data-3.csv. (3 pt)

5. Consider the simple regression model: $Y_i = \beta_0 + \beta_1 x_{i1} + \epsilon_i$ where $\epsilon_i \sim N(0, \sigma^2)$. Suppose that we have collected 10 observations. The followings show the scatterplot with the fitted line as well as some output.



ID	A	B	C	D	E	F	G	H	I	J
x	25	23	5	20	35	18	17	15	14	20
y	85	120	20	64	50	84	50	26	36	60
diag(H)	0.16	0.13	0.47	0.10	0.55	0.10	0.11	0.13	0.15	0.10
residuals	14.49	53.29	-12.55	2.98	-39.49	26.78	-5.32	-25.53	-13.63	-1.02

Note that H is the hat matrix. The diagonal values of H and the residuals were rounded to two decimal places.

(a) Is there any observation that has a high leverage (higher than $\frac{2p}{n}$)? If so, what are they? (1 pt)

(b) Suppose Y_B (i.e. Y for observation B) changes to 50. What would the leverage for the observation be? (1 pt)

(c) Report the values of standardized residuals for observations B, C, E and H (Use $Y_B = 120$). (1 pt)

(d) Report the values of Cooks distance for observations B, C, E and H. Among these points, is there any influential point? Use $\frac{4}{n}$ as the criterion. (1 pt)