VE406
Dr Tong Zhu

JOINT INSTITUTE 交大密西根学院

Quiz 2
Due: Oct 13, 2020

Name and ID: ______________________________

Question1 (10 points)

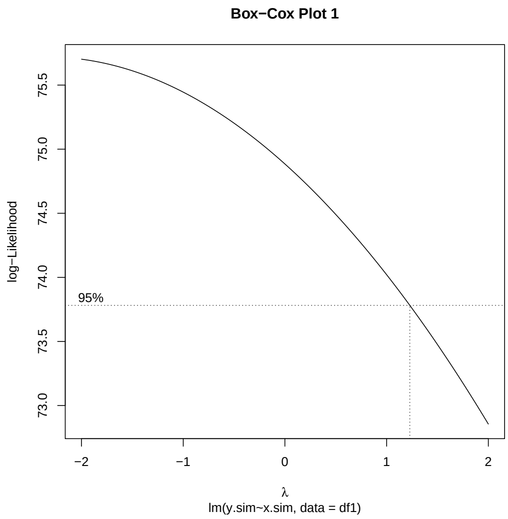
Determine each of the following statements whether it is True or False.

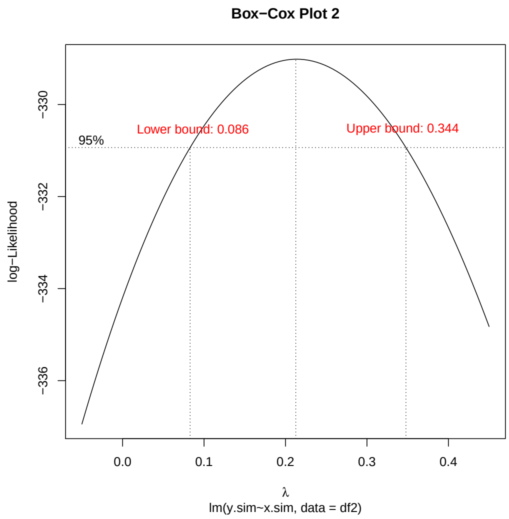
(a) (1 point) ◯ TRUE. ◯ FALSE.

The Box-Cox transformation was originally designed to transform data to be closer to normality, or data that are heteroskedastic to be closer to homoskedasticity.

(b) (1 point) ◯ TRUE. ◯ FALSE.

Plot 1 gives evidence Box-Cox transformation is no use in this case.





(c) (1 point) ◯ TRUE. ◯ FALSE.

Plot 2 gives the probability that the best λ is between 0.086 and 0.344 is 0.95.

(d) (1 point) ◯ TRUE. ◯ FALSE.

The following will not generate any error message in R, and the intervals produced are: the 95% confidence interval for the intercept, the 95% confidence interval for $\mathbb{E}[Y \mid X = 0.5]$, and the 95% prediction interval for $Y \mid X = 0.5$, respectively.

```
> x = rnorm(100, mean = 0, sd = 10)
> y = rnorm(100, 100+0.2*x, sd = 5)
>
> df1 = data.frame(x=x, y=y)
>
> xy.LM = lm(y~x, data = df1)
>
> confint(xy.LM, "(Intercept)", level = 0.95)
```

```
              2.5 %    97.5 %
(Intercept) 98.8492 100.7739
```

```
> predict(xy.LM, data.frame(x = 0.5), interval = "confidence")
```

```
       fit      lwr      upr
1 99.91127 98.95392 100.8686
```

```
> predict(xy.LM, data.frame(x = 0.5), interval = "predict")
```

```
        fit      lwr      upr
1 99.91127 90.31066 109.5119
```

(e) (1 point) ◯ TRUE. ◯ FALSE.

In multiple linear regression, the F-test gives a way to judge whether all the partial regression coefficients are different from zero, provided all the assumptions are satisfied.

(f) (1 point) ◯ TRUE. ◯ FALSE.

In multiple linear regression, the coefficient of determination is a measure of goodness of it between our model and a given dataset used to constructed the model. It is NOT a way to judge whether the underlying model assumptions are satisfied by the model.

(g) (1 point) ◯ TRUE. ◯ FALSE.

In multiple linear regression, the coefficient of determination can be used to compare between two models constructed using the same dataset, provided the underlying model assumptions are satisfied by both models.

(h) (1 point) ◯ TRUE. ◯ FALSE.

The adjusted coefficient of determination is generally larger than the unadjusted coefficient of determination.

(i) (1 point) ◯ TRUE. ◯ FALSE.

The following will not generate any error message in R,

```
> x = rnorm(100, mean = 0, sd = 10)
> y = rnorm(100, 100+0.2*x, sd = 5)
> z = factor(sample(1:3, size = 100, replace = TRUE))
>
> xyz.LM = lm(y~x*z)
```

and it will construct following the regression model in R

$$y = \beta_0 + \beta_1 x + \beta_2 z_2 + \beta_3 z_3 + \beta_4 (z_2 \cdot x) + \beta_5 (z_3 \cdot x) + \varepsilon$$

where

$$z_2 = \begin{cases} 1 & \text{if } z = 2, \\ 0 & \text{otherwise.} \end{cases} \quad \text{and} \quad z_3 = \begin{cases} 1 & \text{if } z = 3, \\ 0 & \text{otherwise.} \end{cases}$$

In this model, the data is partitioned into three categories according to z, and a straight line is fitted to each of the three portions of the data. The intercept of the straight line for the portion of data when $z = 2$ is given by

$$\beta_0 + \beta_2$$

and the slope is given

$$\beta_4$$

(j) (1 point) ◯ TRUE. ◯ FALSE.

The essential multicollinearity refers to a situation in which two or more of the independent variables in a regression model are highly correlated linearly.