Midterm

Charles Hwang

10/20/2022

Charles Hwang

Dr. Xi

STAT 408-001

2022 October 20

Problem 1

Problem 1a

X_i : $n \times 1$ matrix of values $i = 1, 2, \ldots, n$ for the predictor variable

Y_i : $n \times 1$ matrix of values $i = 1, 2, \ldots, n$ for the response variable

β_0 : intercept term for the linear regression model

β_1 : coefficient term for the predictor variable

ϵ_i : error term

Known: X_i (Lecture 6, Slide 14)

Unknown: Y_i, β_0, β_1 (Lecture 6, Slide 14)

Fixed: β_0, β_1, X_i (Lecture 6, Slide 40)

Random: $\hat{\beta}_0, \hat{\beta}_1, \epsilon_i$ (Lecture 6, Slide 40)

Problem 1b

- Linearity ($Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$)
- Homoscedasticity ($\epsilon_i \sim \mathrm{N}(0, \sigma^2)$) (Lecture 6, Slide 36)
- Independence of errors ($\epsilon_i \perp \epsilon_j \forall\ i \neq j$)

Problem 2

Problem 2a

We can see from the `Pr(>|t|)` column in the given output that none of the predictors are significant at the $\alpha = 0.05$ level. However, when conducting an F-test, we can see from the given output ($F = 5.59$, $p = 0.01902$) that we reject H_0 at the $\alpha = 0.05$ level and conclude that these four predictors do have a relationship with the response variable. There is no conflict between these two answers because the t-tests are on each individual variable while the F-test is on the four variables together.

Problem 2b

This code conducts an F-test on comparing the model from problem 2a with four variables to a reduced model in which right leg strength in pounds (`RStr`) and left leg strength in pounds (`LStr`) are considered to be a single variable (Lecture 8, Slides 12-15). This test wants to see which of these two models is better. The null hypothesis is that H_0 : the reduced model (ω) is better and the alternative hypothesis is that H_1 : the full model (Ω) is better.

Problem 2c

Since the residual standard error and the F-test had $df_2 = n - p - 1 = 8$ degrees of freedom, I guess the sample size of the dataset is $n = df + p + 1 = 8 + 4 + 1 = 13$. However, we would not be able to use a statistical test to compare the model from problem 2a to a model with `Hang` as the response and the same four predictors, as the response variables of the two models are different.

Problem 3

Problem 3a

From the given output, we predict that happiness would increase approximately `1.919279` points on a 10-point scale from 1 to 10 if the reported rating for `love` went from "lonely" to "secure" *or* from "secure" to "deep feeling of belonging and caring", holding all other variables constant.

Problem 3b

This code changes the `love` variable so that responses of "lonely" or "secure" are considered `0` and responses of "deep feeling of belonging and caring" are considered `1`. From the given output, we predict that happiness would increase approximately `2.296435` points on a 10-point scale from 1 to 10 if the reported rating for `clove` went from "lonely" *or* "secure" to "deep feeling of belonging and caring", holding all other variables constant. This interpretation appears to differ slightly from the model in problem 3a as there are two levels for the categorical variable as opposed to three. There is also a baseline level (`0`) that produces no change in the outcome variable.

Problem 4

(Lecture 9; Slides 16, 18, 22-23)

1. I believe point 1 is **both** an outlier and influential point. We can see this point clearly differs significantly from the rest of the points. We can also see a hypothetical linear regression fit would be unduly affected by this point.
2. I believe point 2 is an **outlier** only. We can see this point clearly differs significantly from the rest of the points. However, it would not significantly affect a hypothetical linear regression line as the point would fall roughly along it.
3. I believe point 3 is **neither** an outlier nor an influential point. We can see this point is situated among the rest of the points and would not significantly affect a hypothetical linear regression line.

Problem 5

Derivative Form

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ (Lecture 6, Slide 22)

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1(0)$ Since this is the null model, $\bar{x} = 0$

$\hat{\beta}_0 = \bar{y}$ $\square$

Matrix Form

$$\hat{\beta} = \begin{bmatrix}\beta_0\end{bmatrix}, \mathbf{X} = \begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix}_{(n\times 1)}, \mathbf{X}^{\mathrm{T}} = \begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}_{(1\times n)}, y = \begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}_{(n\times 1)} \quad \text{(Lecture 7, Slides 7-9)}$$

$$\hat{\beta} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}y \quad \text{(Lecture 7, Slide 15)}$$

$$\begin{bmatrix}\beta_0\end{bmatrix} = (\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix})^{-1}\begin{bmatrix}1 & 1 & \dots & 1\end{bmatrix}\begin{bmatrix}y_1\\y_2\\\vdots\\y_n\end{bmatrix}$$

$$\begin{bmatrix}\beta_0\end{bmatrix} = (\begin{bmatrix}n\end{bmatrix})^{-1}\begin{bmatrix}\sum_{i=1}^{n} y_i\end{bmatrix}$$

$$\begin{bmatrix}\beta_0\end{bmatrix} = \begin{bmatrix}\frac{1}{n}\end{bmatrix}\begin{bmatrix}\sum_{i=1}^{n} y_i\end{bmatrix}$$

$$\hat{\beta}_0 = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 = \bar{y} \ \square$$