STA302 Lec5101, Methods of Data Analysis 1
Module 4: Problem set

September 25, 2024

1 Basic learning objective practice

Problem.

Which regression modeling assumptions are important to the following statements about the sampling distribution of $\hat{\beta}_0$ and $\hat{\beta}_1$? (linearity assumption, homoscedasticity, the errors are normally distributed)

1. Ensuring the estimators are unbiased.
2. Ensuring the variance of the estimators is not over-estimated or under-estimated.
3. Ensuring the sampling distribution of the estimators is normal.

Problem.

Suppose we have a simple linear model where the predictor (x) has been transformed with a square power, while the response has been transformed with a natural logarithm. The following interpretation is made: The change in the log of the mean response that results from an increase of one-unit in x is $\hat{\beta}_1$. What is wrong with this interpretation.

Problem.

An analyst has identified a violation of Normality in a linear regression model. They use Box-Cox in R to estimate the power that best approximates Normality. The value estimated is $\lambda = 0.389$. Rather than use the modified power transformation of Box-Cox on their response explicitly, the analyst chooses to apply a square root transformation to their response. The resulting diagnostic plots do not show any evidence against the regression modeling assumption. Is this a reasonable choice? Why or why not?

Problem.

Sheather: Chapter 3, Exercise 5 (b)-(e)

Answer this question in a brief R Markdown report (A few paragraphs). Include a relevant scatterplot matrix, residual plots, qq plot, and histogram. Compare your report with other members in your group for the Final Project.

Problem.

Weisberg: Chapter 7, Problem

- 7.1.1
- 7.1.2. Instead of generating new points in 7.1.2, use / modify the following simpler approach in the code below.
- Replace Sulfur by its logarithm, and consider transforming the response Tension. Use diagnostic plots to decide if transforming the response is helpful.

R code:

```
library(alr4)
data(baeskel)

x = baeskel$Sulfur
y = baeskel$Tension

```

```
fit = lm(y ~ I(x^(-1)) )
summary(fit)

plot (x, y)
points(x, fitted(fit), type="l", col="red")
```

2 Advanced learning objective practice from textbooks

Problem.

Sheather: Chapter 6, Exercise 5 (a), (b), (d)

Problem.

Weisberg: Chapter 7, Problems 7.2
Data is in the alr4 package:

```
library(alr4)
data(stopping)
```

Problem.

Weisberg: Chapter 7,

- 7.3.2. Use the Box-Cox method or other techniques such as plots.
- 7.3.3

Data is in the alr4 package:

```
library(alr4)
data(water)
```

Problem.

Pardoe: Chapter 4, Problem 4.3. Use the "internet.csv" dataset in Quercus.