STAT302 Lec5101, Methods of Data Analysis 1 Module 2: Problem set

September 11, 2024

1 Basic learning objective practice

Problem 1.

For each situation, write out the dimensions (i.e. number of rows and columns) of each matrix.

- ullet The $oldsymbol{X}$ matrix for a multiple linear model that includes 5 predictors from a sample of 50 observations.
- The β parameter vector for a multiple linear model that uses 3 numerical predictors and one qualitative predictor with 4 levels included only as indicator variables (not an interaction).

Problem 2.

Use the data below to answer the following questions.

y	x_1	x_2
-0.2	0.5	0.5
-1.1	0.8	0.7
-0.8	0.6	-0.5
7	-0.2	-0.5

- \bullet Compute $\boldsymbol{X}^T\boldsymbol{X}$ without using R
- Compute $X^T y$ without using R
- Use

$$(\boldsymbol{X}^T \boldsymbol{X})^{-1} = \begin{pmatrix} 0.7 & -1.2 & 0.5 \\ -1.2 & 2.9 & -1.2 \\ 0.5 & -1.2 & 1.3 \end{pmatrix}$$

to compute the least squares estimates without using R.

Problem 3 (Advanced).

Remark: This problem is challenging. We will cover problems like this in module 3. Sheather: Chapter 5, Exercise 3 (Use dataset Latour.txt) Chapter 5, Exercise 3

- Load the data set and fit the model in R.
- Interpret the the coefficients in the context of this problem.
- (b) (i, ii)

Problem 4.

Rencher: Chapter 7, Problem 7.3

Problem 5.

(Similar to Rencher: Chapter 7, Problem 7.54 (a)) Use the "GasVapor.csv" dataset. Use this data to compute $\hat{\beta}$ and $\hat{\sigma^2}$.

- Use R for computations.
- (use R to get $X^T y$ and $(X^T X)^{-1}$ and compute $\hat{\beta}$ by hand and then check you get the same thing as with R)

Problem 6.

Rencher: Problem 7.55 (a). This is "landrent" dataset in the "alr4" package. Use the R command: data(landrent).

- Part (a)
- Interpret the coefficients in the context of the problem.

Problem 7.

Pardoe: Chapter 3, Problem 3.1 (a)-(c). Use the "movies.csv" dataset.

2 Advanced learning objective practice

Problem 8.

Pardoe: Chapter 4, Problem 4.8 (a)-(e)