ST117 Lab 7 Workbook

26 February, 2024

1. Linear Regression Model

Generate a sample from the linear regression model:

$$Y = 2X + \varepsilon,\ \varepsilon \sim \mathcal{N}(0, 5)$$

```
#TODO: write your codes here
```

(a) Find the regression line to predict $\hat{y}$ by x, i.e. find $\hat{a}_1$, $\hat{b}_1$ such that

$$Y = \hat{b}_1 X + \hat{a}_1$$

manually using the covraiance and correlation estimators. Then verify the values using the `lm` function in R.

```
#TODO: write your codes here
```

(b) Find the regression line to predict $\hat{x}$ by y, i.e. find $\hat{a}_2$, $\hat{b}_2$ such that

$$X = \hat{b}_2 Y + \hat{a}_2$$

manually using the covraiance and correlation estimators. Then verify the values using the `lm` function in R.

```
#TODO: write your codes here
```

(c) Plot the data along with the two regression lines. Which regression line appears to be better?

```
#TODO: write your codes here
```

(d) The Least Squared Estimation can also help us derive the estimates of $\hat{a}$ and $\hat{b}$. We just need to find the value of a and b that minimizes the following expression:

$$\sum_{i=1}^{n} \Big(y_i - (a + bx_i)\Big)^2$$

Hint: Use the `optim` function.

```
#TODO: write your codes here
```

2. Mean Squared Error

Consider a random sample $X_1, X_2, \dots, X_n$ from a normal distribution with mean μ and variance σ^2. We have the following two estimators of the population variance obtained from the given sample information. From our previous lab, we derived the Maximum Likelihood Estimator(MLE) of the population variance as:

1. $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Additionally, we have the sample variance given by:

$$2.\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2,$$

where $\overline{X}$ represents the sample mean.

(a) Simulate a sample of 10 data from $\mathrm{N}(160, 10^2)$. Calculate the values of these two estimators and their Mean Squared Error(MSE).

```
#TODO: write your codes here
```

(b) Try sample sizes of 100, 1000 and 10000, and show the results in a table. Can you provide an interpretation?

```
#TODO: write your codes here
```