Practice Session 3

Part 1 : The relationship between two Quantitative Variables/ Linear Regression

Remember that the fitted regression line is defined by the equation:

- $\widehat{y} = a + bx$, or
- $\widehat{Response} = a + b(Explanatory)$
- Residuals= observed- predicted

Where:

- Response: is the response variable or the dependent variable
- Explanatory: is the independent variable
- a: is the y- intercept
- b: is the slope of the regression line

You may use the following R functions: `plot()`, `lm()`, `cor()`, `abline()`. And you might need to download Lock5Data using Library(Lock5Data).

Practice 1:

State if the following sentences are true or false.

a. We choose the linear model that passes through the most data points on scatter plot.

b. The residuals are observed y-values minus the y-value predicted by linear model.

c. Least square means that the square of the largest residuals is as small as it could be possibly be.

d. Some of the residuals from least linear model will be positive and some will be negative.

e. Least squares means that some of the squares of the residuals are minimized.

f. We write $\hat{y}$ to denote the predicted values and the y to denote the observed value.

Practice 2:

Use data `FloridaLakes` and the two variables `Alkalinity` and `calcium`.

```
# download the data and load it into R
library(Lock5Data)
data(FloridaLakes)
```

1. Find the correlation for the `Alkalinity` and `calcium`.

```
# your code here
```

2. Using the appropriate R function, create a linear model to predict `Alkalinity` from `calcium`.

```
# your code here
```

3. Find the coefficients of the regression and write the correct equation of this linear model.

```
# your code here
```

4. Interpret the intercept and slope of the model within the context of the data.

5. Predict the `Alkalinity` level when Calcium = 2.5.

```
# your code here
```

6. The actual `Alkalinity` level was 8.5 when Calcium = 2.5. Find the residual for this data point.

```
# your code here
```

7. Find the five number summary for the variable `Calcium`.

```
# your code here
```

8. Predict the `Alkalinity` level when Calcium = 0.5.

```
# your code here
```

9. Do both of these predictions make sense given the data?

```
# your code here
```

Part 2: Review

Practice 3:

From the data `ICUAdmissions` create descriptive statistics (such as: `boxplot`, `histogram`, `five number summary`) for the variables `Infection`, `Age`, `Race`, and `HeartRate`.

Note: You can use the function `str` ti check the structure of the data.

1. Create descriptive statistics using `one quantitative variable`.
2. Create descriptive statistics using `one categorical variable`.
3. Descriptive statistics for `two quantitative variable`.
4. Descriptive statistics for `one quantitative and one categorical variables`.

```
# download the data and load it into R
library(Lock5Data)
data(ICUAdmissions)
```