STAT3011 Fall2019 Week8 Lab

During today's lab, we will practice functions we learn in Chapter 12.

Generate Random Sample

We first generate explanatory variable X:

```
X = rnorm(n = 500,mean = 5,sd = 2)
```

Then we generate random errors:

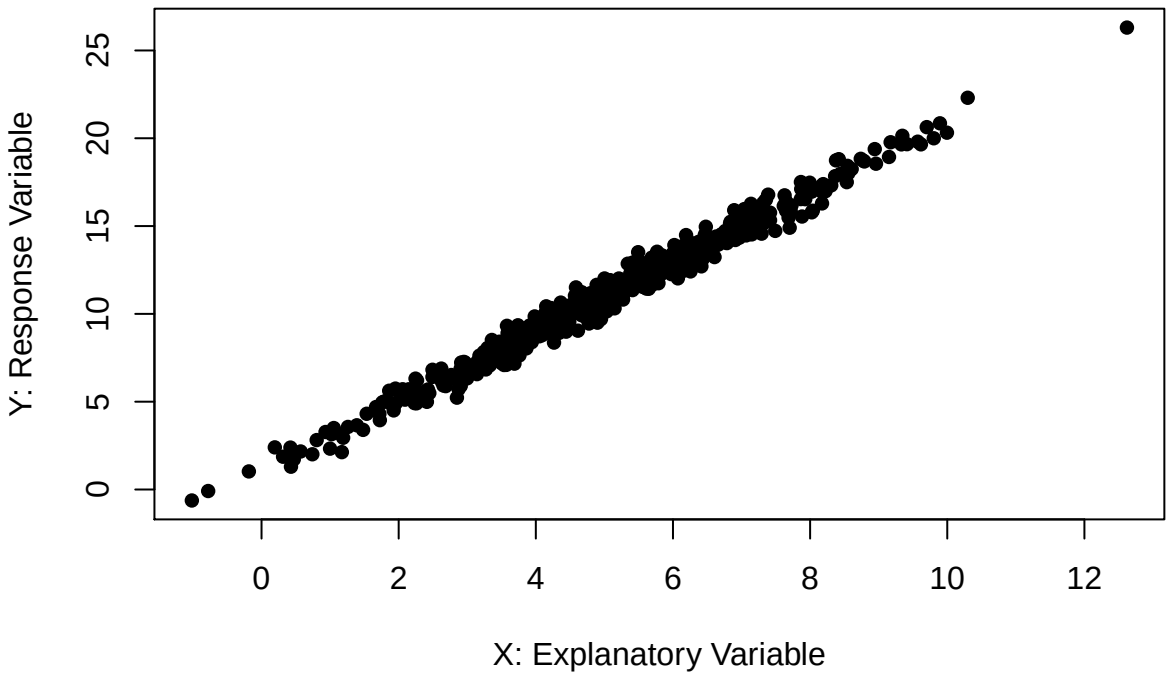
```
Error = rnorm(n = 500, mean = 0,sd = 0.5)
```

Now we set $\alpha = 1, \beta = 2$, and get our response variable Y as:

```
Y = 1 + 2*X + Error
```

Let's draw a scatter plot to show the relationship of X and Y:

```
plot(X,Y,xlab = "X: Explanatory Variable", ylab = "Y: Response Variable", pch = 16)
```



Correlation and Linear Regression

Calculate correlation of X and Y:

```
cor(X,Y)
```

```
## [1] 0.9915027
```

```
cor(Y,X)
```

```
## [1] 0.9915027
```

We can see, the correlation between X and Y does not depend on the order.

Now let's fit linear regression:

```
LSL = lm(Y ~ X)
```

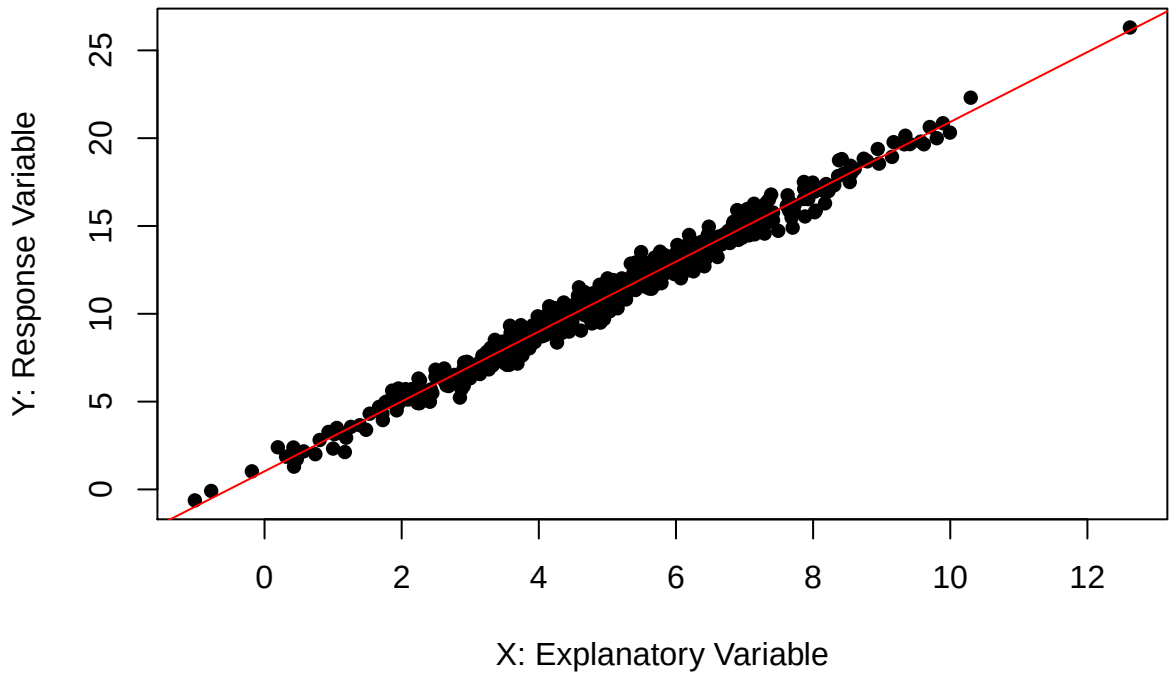
The fitted model is saved as object **LSL**, we can use function *summary* to explore the fitted model:

```
summary(LSL)
```

```
##
## Call:
## lm(formula = Y ~ X)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.47056 -0.37254  0.00832  0.36441  1.55566
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.03137    0.06357   16.23   <2e-16 ***
## X            1.98923    0.01170  170.09   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.5287 on 498 degrees of freedom
## Multiple R-squared:  0.9831, Adjusted R-squared:  0.983
## F-statistic: 2.893e+04 on 1 and 498 DF,  p-value: < 2.2e-16
```

We can see, the fitted intercept $a = 1.03$, is close to the true intercept $\alpha = 1$; fitted slope $b = 1.99$, is close to the true slope $\beta = 2$. Let's draw the fitted line in red:

```
plot(Y ~ X, xlab = "X: Explanatory Variable", ylab = "Y: Response Variable", pch = 16)
abline(LSL, col = "red")
```



Explore the Predicted Value and Residuals

We can extract the predicted values of Y from **LSL**, and get mean of the predicted values:

```
mean(LSL$fitted.values)
```

```
## [1] 11.06761
```

```
mean(Y)
```

```
## [1] 11.06761
```

We can see, mean predicted value is the same as mean of Y.

We can also get the sum of the residuals:

```
sum(LSL$residuals)
```

```
## [1] 6.766289e-15
```

We can see, sum of the residuals is 0.