SMT HOMEWORK -6-

berkay cayan

```r
data <- read.csv("US Stock Market Dataset.csv")
```

```r
na_count <- sum(is.na(data))
print(paste("Total NA:", na_count))
```

```
[1] "Total NA: 490"
```

```r
cleardata <- na.omit(data)

any(is.na(cleardata))
```

```
[1] FALSE
```

```r
selected_data <- data[, c("Natural_Gas_Price", "Crude_oil_Price")]


head(selected_data)
```

```
  Natural_Gas_Price Crude_oil_Price
1             2.079           72.28
2             2.050           73.82
3             2.100           75.85
4             2.077           77.82
5             2.490           76.78
6             2.712           78.01
```

```r
str(selected_data)
```

```
'data.frame':   1013 obs. of  2 variables:
 $ Natural_Gas_Price: num  2.08 2.05 2.1 2.08 2.49 ...
 $ Crude_oil_Price  : num  72.3 73.8 75.8 77.8 76.8 ...
```

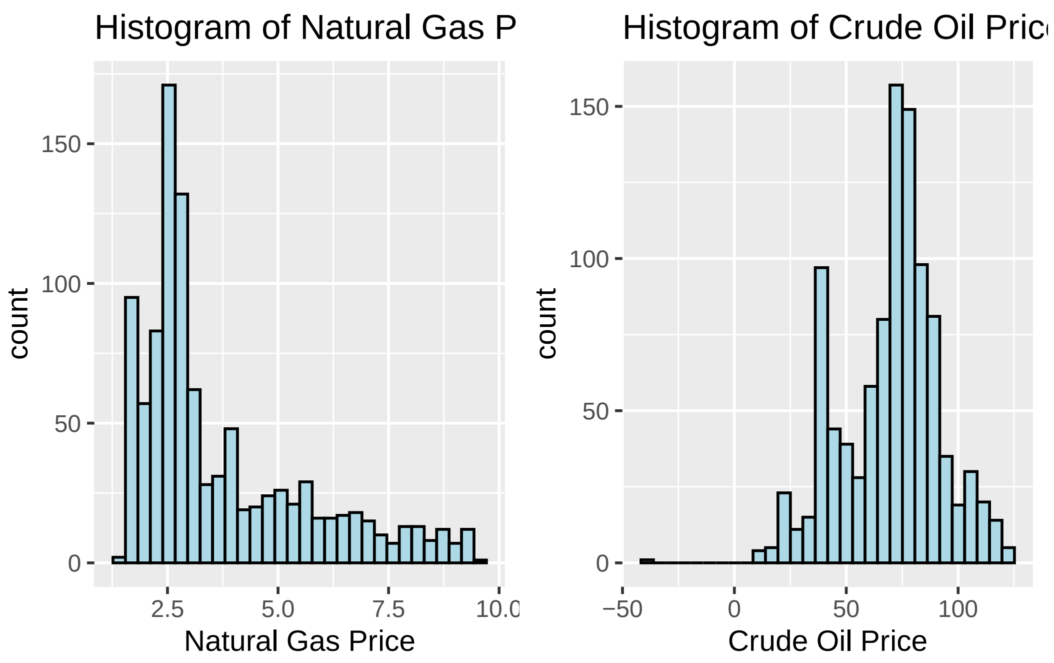
```r
summary(selected_data)
```

```
 Natural_Gas_Price Crude_oil_Price
 Min.   :1.482     Min.   :-37.63
 1st Qu.:2.447     1st Qu.: 54.76
 Median :2.880     Median : 72.91
 Mean   :3.727     Mean   : 69.85
 3rd Qu.:4.805     3rd Qu.: 82.81
 Max.   :9.647     Max.   :123.70
```

```r
histogram1 <- ggplot(data, aes(x = Natural_Gas_Price)) +
              geom_histogram(fill = "lightblue", color = "black") +
              labs(title = "Histogram of Natural Gas Price", x = "Natural Gas Price")

histogram2 <- ggplot(data, aes(x = Crude_oil_Price)) +
              geom_histogram(fill = "lightblue", color = "black") +
              labs(title = "Histogram of Crude Oil Price", x = "Crude Oil Price")

grid.arrange(histogram1, histogram2, ncol = 2)
```



```r
set.seed(123)
index = sample(1 : nrow(selected_data), round(nrow(selected_data) * 0.80))
train_data = selected_data[index,]
test_data = selected_data[-index,]
```

```r
model <- lm(Natural_Gas_Price ~ Crude_oil_Price, data = selected_data)

summary(model)
```

```
Call:
lm(formula = Natural_Gas_Price ~ Crude_oil_Price, data = selected_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.5313 -1.0785 -0.1243  0.8132  4.8477 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.595375   0.145322  -4.097 4.52e-05 ***
Crude_oil_Price  0.061873   0.001984  31.178  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.388 on 1011 degrees of freedom
Multiple R-squared:  0.4902,    Adjusted R-squared:  0.4897 
F-statistic: 972.1 on 1 and 1011 DF,  p-value: < 2.2e-16
```

The coefficient of Crude_oil_Price indicates the sensitivity of the Natural_Gas_Price variable to changes in the Crude_oil_Price variable. In this case, when the Crude_oil_Price variable increases by one unit, the Natural_Gas_Price variable increases by 0.061873 units.

The residual standard error represents the standard deviation of the error terms in the model. In this case, it is 1.388.

The coefficient of determination, or R-squared, is the percentage of variance in the dependent variable explained by the independent variables. In this model, the R-squared value is 49.02%. Therefore, the model has a relatively high explanatory power.

The F-statistic is a test statistic that evaluates the statistical significance of the model. In this model, the F-statistic value is 972.1 and the p-value is < 2.2e-16. These results indicate that the model is statistically significant.

```r
anova_result <- aov(Natural_Gas_Price ~ Crude_oil_Price, data = selected_data)


print(summary(anova_result))
```

```
                  Df Sum Sq Mean Sq F value Pr(>F)    
Crude_oil_Price    1   1873  1872.8   972.1 <2e-16 ***
Residuals       1011   1948     1.9                   
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Degrees of freedom are given for Crude_oil_Price and error terms. In this case, there is 1 degree of freedom for Crude_oil_Price and 1011 degrees of freedom for error terms.

Sum of squares expresses the total sum of squares of the relevant independent variable or error terms. In this case, there are 1873 sum of squares for Crude_oil_Price and 1948 sum of squares for error terms.

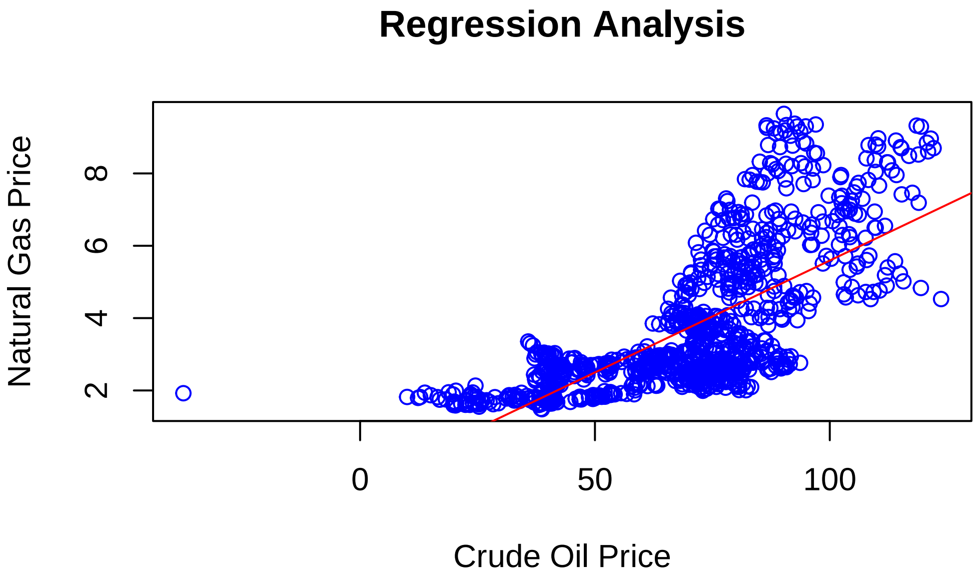
Mean square is obtained by dividing the total sum of squares by the relevant degrees of freedom, representing the variance of the relevant factor or error terms. It is 1872.8 for Crude_oil_Price and 1.9 for error terms.

The F-statistic represents the ratio of the variance of the relevant factor to the variance of the error terms. In this case, the ratio of the variance of Crude_oil_Price factor to the variance of error terms is 972.1.

The p-value evaluates the significance of the F-statistic. In this case, the p-value is very small (<2e-16), indicating that the model is statistically significant.

```r
plot(selected_data$Crude_oil_Price, selected_data$Natural_Gas_Price,
     xlab = "Crude Oil Price", ylab = "Natural Gas Price",
     main = "Regression Analysis", col = "blue")

abline(anova_result, col = "red")
```



```r
ancova_result <- lm(Natural_Gas_Price ~ Crude_oil_Price, data = selected_data)

print(summary(ancova_result))
```

```
Call:
lm(formula = Natural_Gas_Price ~ Crude_oil_Price, data = selected_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.5313 -1.0785 -0.1243  0.8132  4.8477 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -0.595375   0.145322  -4.097 4.52e-05 ***
Crude_oil_Price  0.061873   0.001984  31.178  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.388 on 1011 degrees of freedom
Multiple R-squared:  0.4902,	Adjusted R-squared:  0.4897 
F-statistic: 972.1 on 1 and 1011 DF,  p-value: < 2.2e-16
```

The estimated coefficient for Crude_oil_Price is 0.061873. This indicates that for every one-unit increase in Crude_oil_Price, Natural_Gas_Price is expected to increase by approximately 0.061873 units.

This column indicates how statistically significant each coefficient is. High t-values indicate that the respective coefficient is significant. For example, a high t-value (31.178) is obtained for the Crude_oil_Price coefficient, indicating that the Natural_Gas_Price variable is associated with Crude_oil_Price.

It evaluates the significance of each coefficient. Small p-values indicate that the respective coefficient is statistically significant. For instance, the p-value for the Crude_oil_Price coefficient is very small (< 2e-16), indicating that the coefficient is statistically significant.

It expresses the percentage of variance in the dependent variable explained by the independent variables. It indicates how well the model fits the data.