Question 2

2023-10-17

{r setup, include=FALSE} knitr::opts_chunk$set(echo = TRUE)

Reading the data:

```
#Reading the data:

#install.packages("openxlsx")
library(openxlsx)
inflation_years <- read.xlsx("Inflation by Year.xlsx")
inflation_quarters <- read.xlsx("Inflation by Quarter.xlsx")
head(inflation_years)
```

```
##   Year Inflation
## 1 1989       5.7
## 2 1990       8.0
## 3 1991       7.5
## 4 1992       4.6
## 5 1993       2.6
## 6 1994       2.2
```

```
head(inflation_quarters)
```

```
##   Year Quarter Inflation
## 1 1989      Q1       5.8
## 2 1989      Q2       5.8
## 3 1989      Q3       5.6
## 4 1989      Q4       5.7
## 5 1990      Q1       6.1
## 6 1990      Q2       8.0
```

1.

```
n=nrow(inflation_years)/2
first_half_inflation_years=inflation_years[1:n,2]
second_half_inflation_years=inflation_years[n:34,2]
t_test_result <- t.test(first_half_inflation_years, second_half_inflation_years)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  first_half_inflation_years and second_half_inflation_years
## t = 0.8256, df = 29.49, p-value = 0.4157
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.7792023  1.8354114
## sample estimates:
## mean of x mean of y
##  3.005882  2.477778
```

The p.value is larger than 0.05, therefore is not possible to reject the null hypothesis. Therefore there isn't a significant difference between the means, and the confidence interval for the difference of the means is (-0.7,1.8).

2.

```
n=nrow(inflation_quarters)/2
first_half_inflation_quarters=inflation_quarters[1:n,3]
second_half_inflation_quarters=inflation_quarters[n:nrow(inflation_quarters),3]
t_test_result <- t.test(first_half_inflation_quarters, second_half_inflation_quarters)
t_test_result
```

```
##
##  Welch Two Sample t-test
##
## data:  first_half_inflation_quarters and second_half_inflation_quarters
## t = 0.95158, df = 135.34, p-value = 0.343
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3556055  1.0151914
## sample estimates:
## mean of x mean of y
##  2.985507  2.655714
```

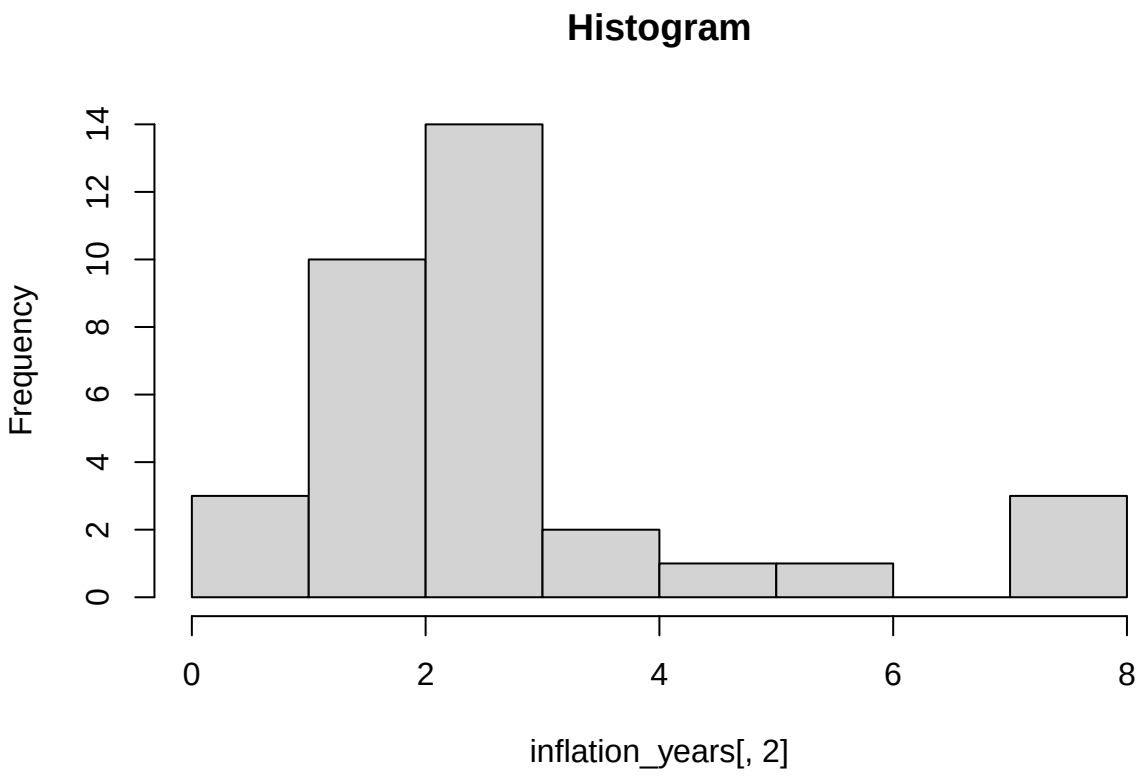
The p-value is 0.34, therefore is not possible to reject the null hypothesis. And the means of the groups are not significantly different from each other.

3.

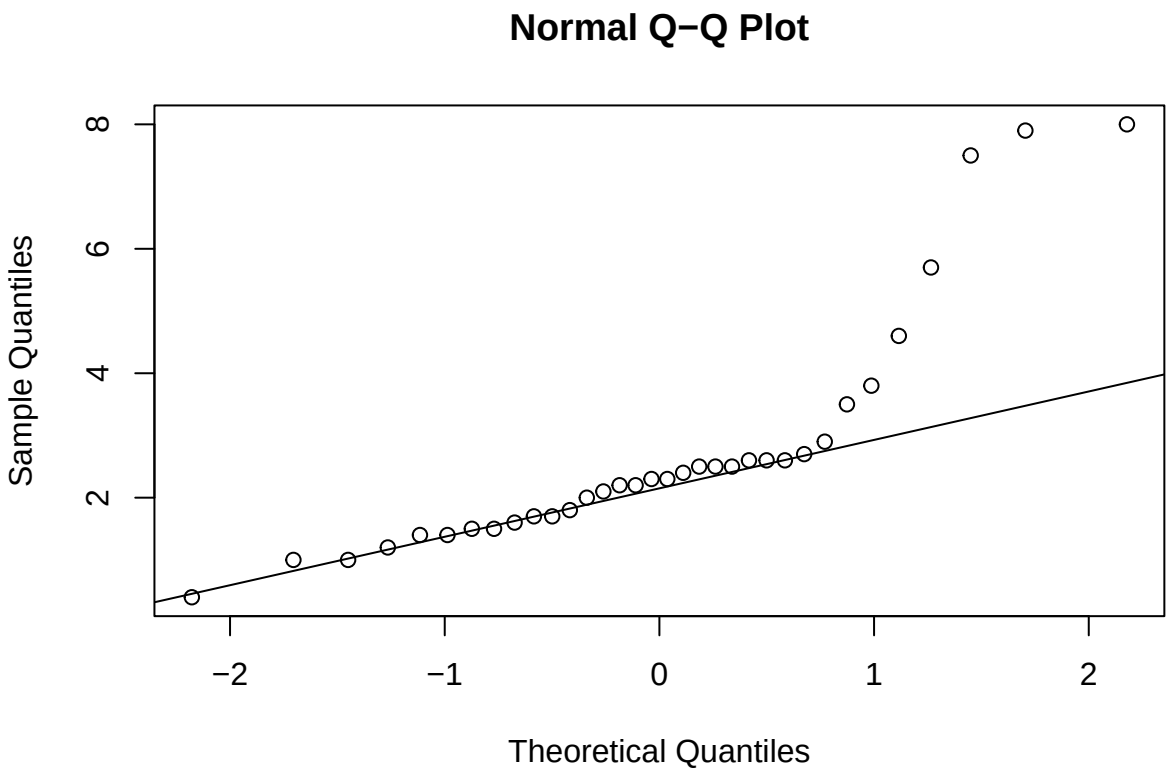
Is the Inflation by Years Data Normal?

```
# Inflation by Years

# Histogram
hist(inflation_years[,2], main = "Histogram")
```




```
# Q-Q Plot
qqnorm(inflation_years[,2])
qqline(inflation_years[,2])
```



Graphically it can be seen that data doesn't seem normal, there is an increase in inflation in recent years that deviates from normal.

```
## Shapiro-Wilk test
shapiro.test(inflation_years[,2])
```

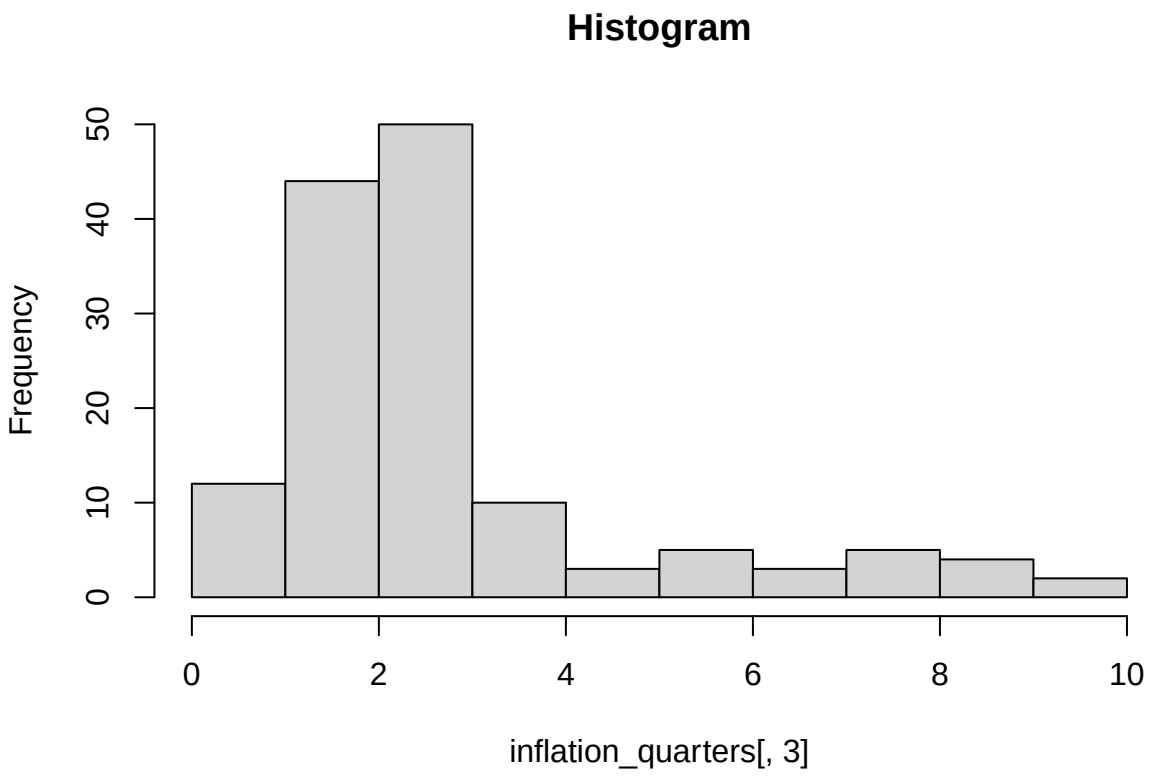
```
##
##  Shapiro-Wilk normality test
##
## data:  inflation_years[, 2]
## W = 0.77101, p-value = 7.393e-06
```

The p-value is significant, we can reject the null hypothesis, and conclude that the data is NOT normal.

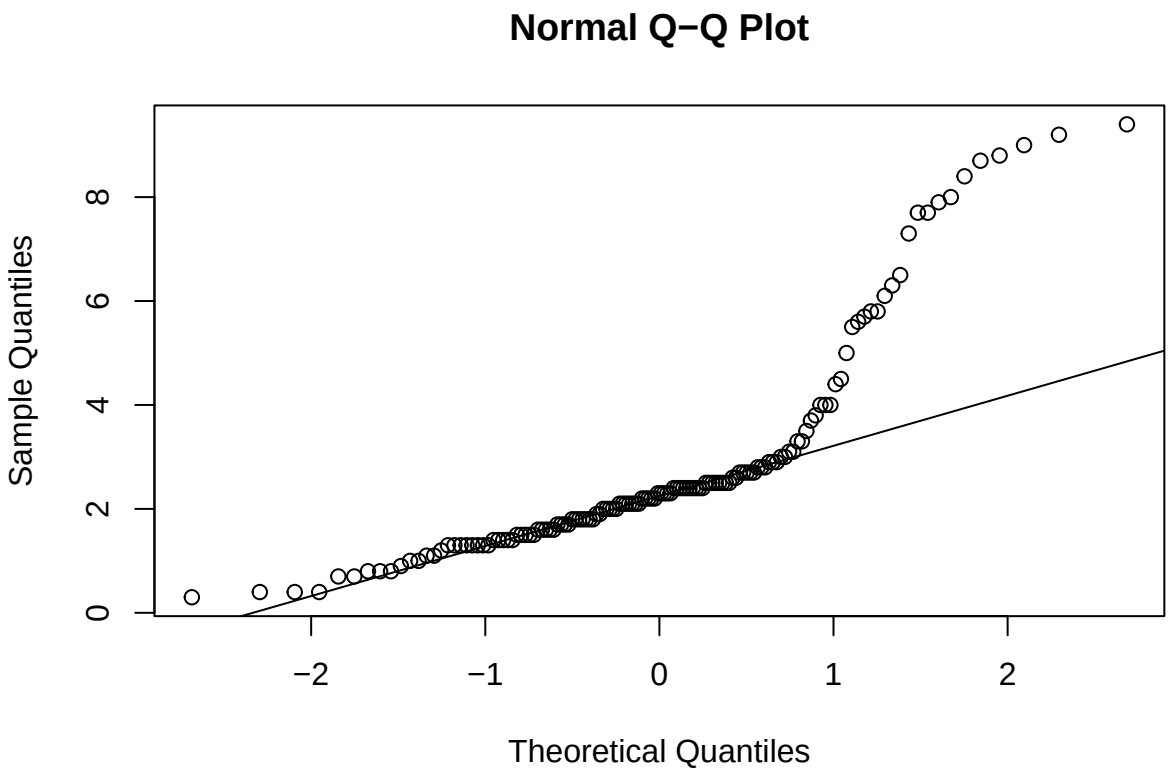
Is the Inflation by Quarters Data Normal?

```
# Inflation by Quarters

# Histogram
hist(inflation_quarters[,3], main = "Histogram")
```




```
# Q-Q Plot
qqnorm(inflation_quarters[,3])
qqline(inflation_quarters[,3])
```




```
## Shapiro-Wilk test
shapiro.test(inflation_quarters[,3])
```

```
##
##  Shapiro-Wilk normality test
##
## data:  inflation_quarters[, 3]
## W = 0.78485, p-value = 5.901e-13
```

The p-value is significant, and graphically it is observed that the inflation by quarters is NOT normal.

4.

Model for Inflation by years

```
model <- lm(Inflation ~ Year, data = inflation_years)
summary(model)
```

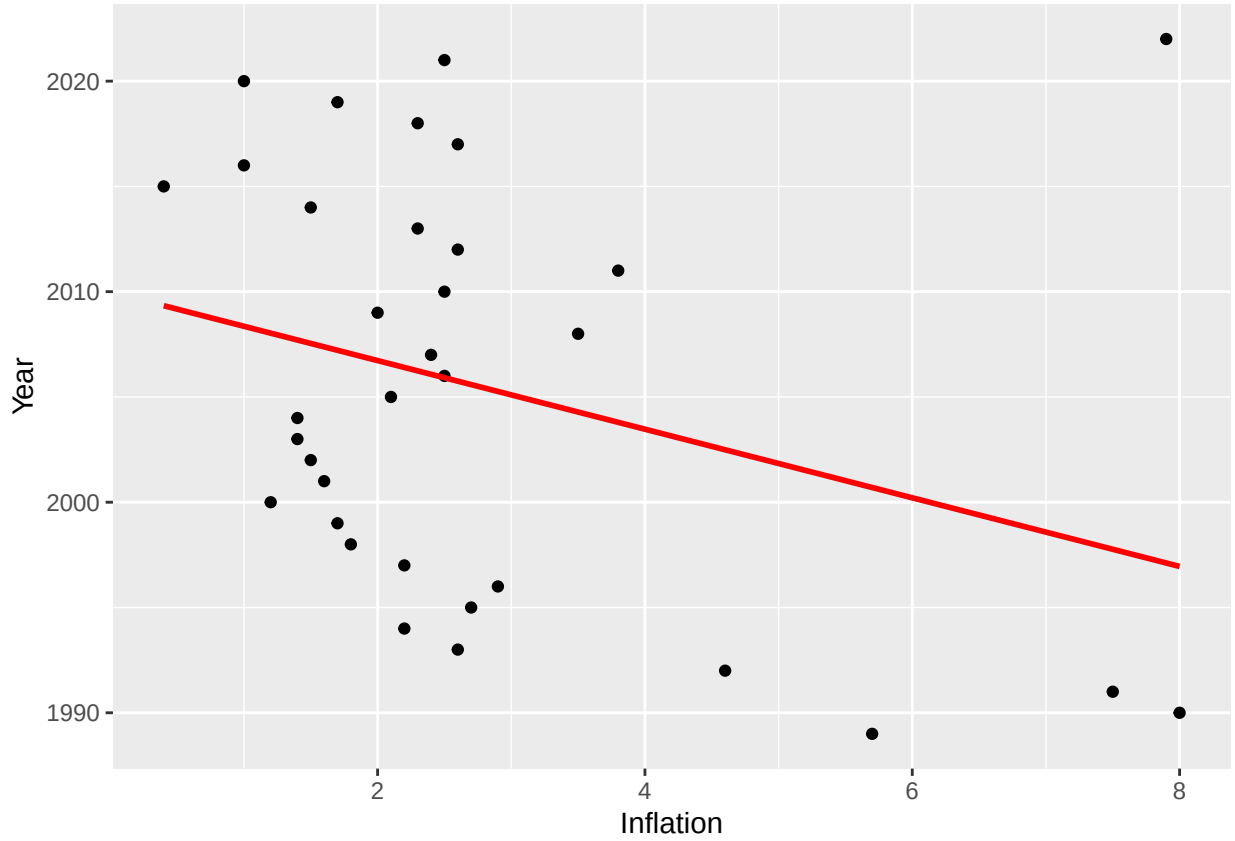
```
##
## Call:
## lm(formula = Inflation ~ Year, data = inflation_years)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8763 -1.2057 -0.4793  0.4629  6.1171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 120.66225   64.03887   1.884   0.0686 .
## Year         -0.05879    0.03193  -1.841   0.0749 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.827 on 32 degrees of freedom
## Multiple R-squared:  0.09579,    Adjusted R-squared:  0.06754
## F-statistic:  3.39 on 1 and 32 DF,  p-value: 0.07487
```

```
library(ggplot2)

# Create a scatterplot
ggplot(inflation_years, aes(x = Inflation, y = Year)) +
  geom_point() +  # Plot data points
  geom_smooth(method = "lm", se = FALSE, color = "red")  # Add the regression line
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Conclusion: It is observed that the year is not a significant variable to predict the Inflation and the regression

is not accurate, in fact the Multiple-R-Squared indicates that the variable year only explain the 9% of the variability within the inflation variable.

Model for Inflation by Quarters:

Simple linear regression: Inflation - Year

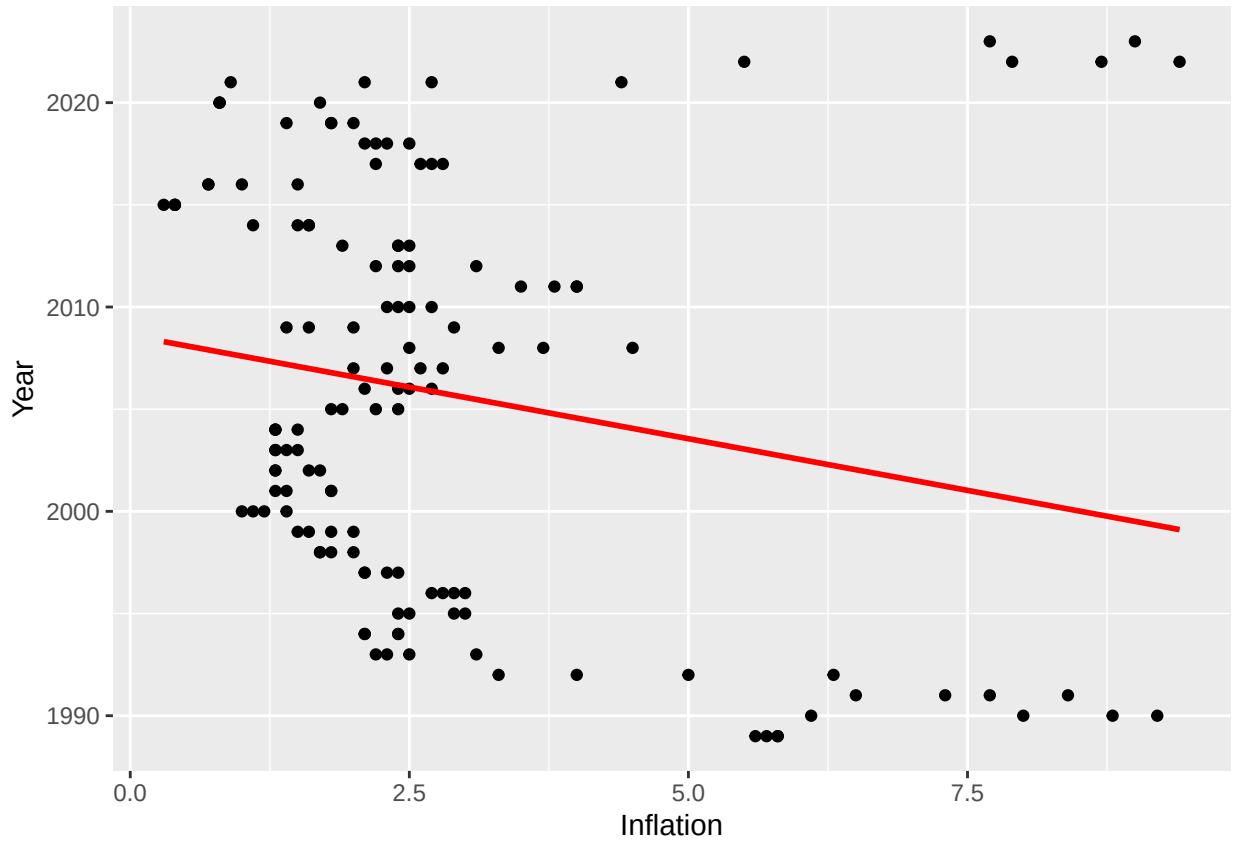
```
##
model <- lm(Inflation ~ Year, data = inflation_quarters)
summary(model)
```

```
##
## Call:
## lm(formula = Inflation ~ Year, data = inflation_quarters)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.1327 -1.2862 -0.4963  0.2083  7.2640
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 87.84702   34.45861   2.549   0.0119 *
## Year        -0.04239    0.01718  -2.467   0.0149 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.01 on 136 degrees of freedom
## Multiple R-squared:  0.04285,    Adjusted R-squared:  0.03581
## F-statistic: 6.088 on 1 and 136 DF,  p-value: 0.01485
```

```
library(ggplot2)

# Create a scatterplot
ggplot(inflation_quarters, aes(x = Inflation, y = Year)) +
  geom_point() +  # Plot data points
  geom_smooth(method = "lm", se = FALSE, color = "red")  # Add the regression line
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



The model for Inflation predicted by year, is less accurate in the data "Inflation by quarters" than in "Inflation by yaers", we observe again that the year variable is not significant. The multiple R-Squared is 0.04, thus the variable year in this model only explains 4% of the Inflation's variability.

Simple Linear Regression: Inflation - Quarter

```
inflation_quarters$Quarter=as.factor(inflation_quarters$Quarter)
model <- lm(Inflation ~ Quarter, data = inflation_quarters)
summary(model)
```

```
##
## Call:
## lm(formula = Inflation ~ Quarter, data = inflation_quarters)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5686 -1.2478 -0.5279  0.1244  6.6118
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.871429   0.349827   8.208 1.65e-13 ***
## QuarterQ2   -0.002857   0.494730  -0.006    0.995
## QuarterQ3   -0.103782   0.498354  -0.208    0.835
## QuarterQ4   -0.083193   0.498354  -0.167    0.868
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Residual standard error: 2.07 on 134 degrees of freedom
## Multiple R-squared:  0.0005218,  Adjusted R-squared:  -0.02185
## F-statistic: 0.02332 on 3 and 134 DF,  p-value: 0.9952
```

```
library(ggplot2)
```

This model is the less accurate model, it does't fit the data, the quarter variable is not signifficant for the model, the multiple R-squared is 0 and the p-value is 0.9, then there is no relationship betweeen tha variable quarters and the inflation. This is different from the two previous models where the p-value was significant.