MATH1324 Assignment 3

Code

Can money buy happiness?

Abdul Raheman Nasir Gazi(s3705448)

RPubs link information

- Rpubs link: http://rpubs.com/literalmoniker/assign3moneyhapp (http://rpubs.com/literalmoniker/assign3moneyhapp)

Introduction

- Today I am going to try and answer the age-old question. “Can money buy happniess?”
- Researchers have been working on this question for a century or so
- There have been research about how spending money affects your dopamine level which makes you feel rewarded and in-turn makes you happy
- I am interested in finding how money correlates to the happiness in the world. I will do that by using GDP.per.Capita and Happiness.Score from the dataset

Problem Statement

- I am interested in finding correlation between **money** and **happiness**
- From the dataset I got from:- https://www.kaggle.com/unsdsn/world-happiness/downloads/world-happiness-report.zip/2#2017.csv (https://www.kaggle.com/unsdsn/world-happiness/downloads/world-happiness-report.zip/2#2017.csv) I will work on two speicific variables:- GDP.per.Capita and Happiness.Score

Data

- Data has been published by **The World Happiness Report**
- Data has been accesed from:-https://www.kaggle.com/unsdsn/world-happiness/downloads/world-happiness-report.zip/2#2017.csv (https://www.kaggle.com/unsdsn/world-happiness/downloads/world-happiness-report.zip/2#2017.csv)
- Only **2016.csv** has been used for the statistics and analysis
- Varibales in the dataset:
    - **Country** - A total of 155 countries were surveyd, this variable consists of all the 155 countries
    - **Happiness.Rank** - Countries have been ranked as per their Happiness.Score
    - **Happiness.Score** - People were asked about happy they are from 0 to 10, 10 being the happiest. An average was calculated for every individual country and then ranked
    - **Whisker.high** - Higher Confidence Interval of the Happiness Score
    - **Whisker.low** - Lower Confidence Interval of the Happiness Score
    - **Economy..GDP.per.Capita.** - The extent to which GDP contributes to the calculation of the Happiness Score.

Data Cont.

- Remaining variables

  - **Family** - The extent to which Family contributes to the calculation of the Happiness Score
  - **Health..Life.Expectancy.** - The extent to which Perception of Corruption contributes to Happiness Score
  - **Freedom** - The extent to which Freedom contributed to the calculation of the Happiness Score
  - **Generosity** - The extent to which Generosity contributed to the calculation of the Happiness Score
  - **Trust..Government.Corruption.** - The extent to which Perception of Corruption contributes to Happiness Score
  - **Dystopia.Residual** - Least happy people and the personal factors affecting their happiness scored
- The dataset contains 155 observations and 12 variables
- Data is taken in the year 2017
- Happiness.Rank and Economy..GDP.per.Capita. are the two variables we are going to work upon

Below code shows the importing of the dataset

Hide

```
library(readr)
happiness <- read_csv("C:/Users/Suraj/Downloads/2017.csv")
View(happiness)
```

Descriptive Statistics and Visualisation

- Using descriptive statistics we found that mean value of **Happiness.Score** is 5.35 which is moderate if we scale it between 0 to 10
- We also found that 7.53 is the maximum happiness score of the country Norway which sits at number one in ranking
- Rest of the descriptive statistics is given below

Hide

```
knitr::opts_chunk$set(error = TRUE)
happinesssummary <- happiness %>%
            summarise(
             Min = min(Happiness.Score,na.rm = TRUE),
             Q1 = quantile(Happiness.Score,probs = .25,na.rm = TRUE),
             Median = median(Happiness.Score, na.rm = TRUE),
             Q3 = quantile(Happiness.Score,probs = .75,na.rm = TRUE),
             Max = max(Happiness.Score,na.rm = TRUE),
             Mean = mean(Happiness.Score, na.rm = TRUE),
             SD = sd(Happiness.Score, na.rm = TRUE),
             n = n())
happinesssummary
```

Min	Q1	Median	Q3	Max	Mean	SD	n
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<int>
2.693	4.5055	5.279	6.1015	7.537	5.354019	1.13123	155

1 row

Hide

```
gdpsummary <- happiness %>%
             summarise(
              Min = min(Economy..GDP.per.Capita.,na.rm = TRUE),
              Q1 = quantile(Economy..GDP.per.Capita.,probs = .25,na.rm = TRUE),
              Median = median(Economy..GDP.per.Capita., na.rm = TRUE),
              Q3 = quantile(Economy..GDP.per.Capita.,probs = .75,na.rm = TRUE),
              Max = max(Economy..GDP.per.Capita.,na.rm = TRUE),
              Mean = mean(Economy..GDP.per.Capita., na.rm = TRUE),
              SD = sd(Economy..GDP.per.Capita., na.rm = TRUE),
              n = n())
gdpsummary
```

Min <dbl>	Q1 <dbl>	Median <dbl>	Q3 <dbl>	Max <dbl>	Mean <dbl>	SD <dbl>	n <int>
0	0.6633708	1.064578	1.318027	1.870766	0.9847182	0.4207927	155

1 row

Decsriptive Statistics Cont.

- No outliers were found in both of the variables(Checked using boxplot below.)

Hide

```
boxplot(
  happiness$Economy..GDP.per.Capita.,
  happiness$Happiness.Score,
  ylab = "Scores",
  xlab = "Wealth and Happiness"
  )
axis(1, at = 1:2, labels = c("Wealth", "Happiness"))
```

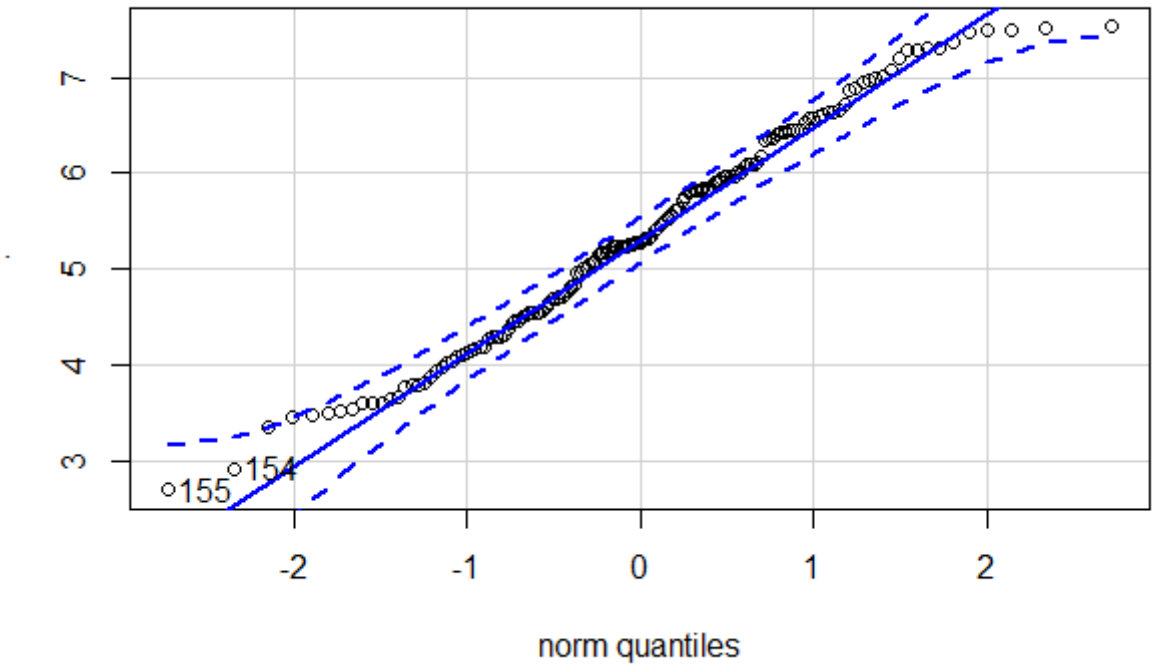
Decsriptive Statistics Cont.(2)

- **Happiness.Score** was found to be normally distributed as per the Q-Q Plot
- **Economy..GDP.per.Capita.** had some slight skew between 1 to 2 quantiles

Hide

```
happiness$Happiness.Score  %>% qqPlot(dist="norm")
```

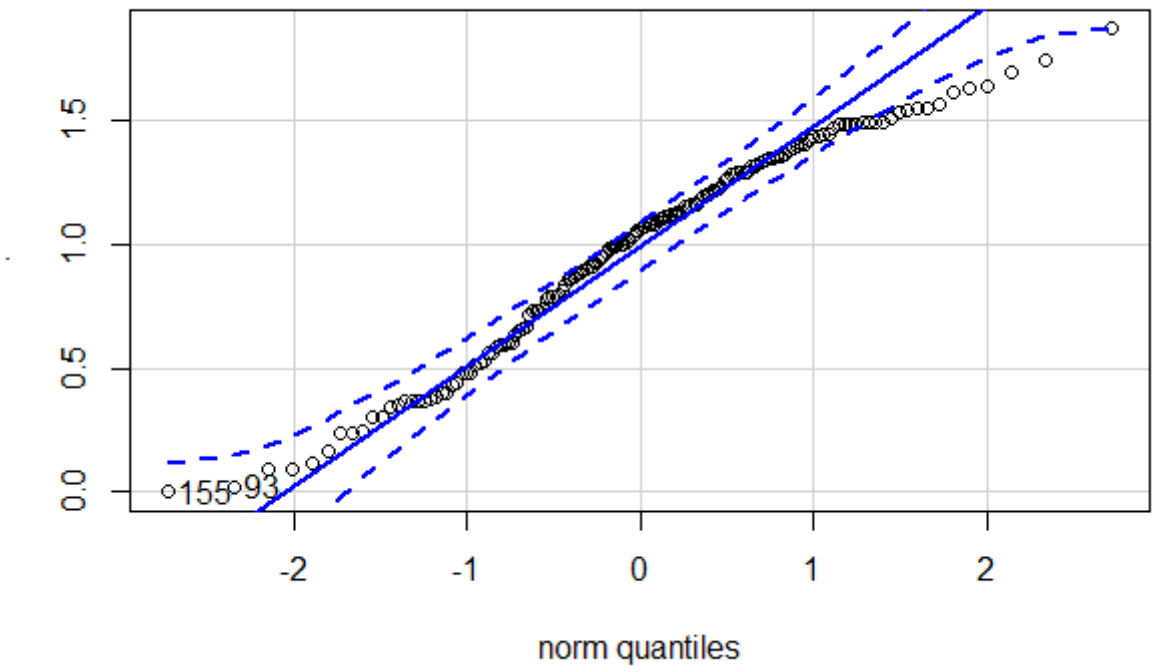
```
[1] 155 154
```



Hide

```
happiness$Economy..GDP.per.Capita.  %>% qqPlot(dist="norm")
```

```
[1] 155  93
```



T-Test

Paired sample T-Test is used to find mean difference between the two variables. The mean difference came out to be -4.369.

Hide

```
t.test(happiness$Economy..GDP.per.Capita., happiness$Happiness.Score,
       paired = TRUE,
       alternative = "two.sided")
```

```

    Paired t-test

data:  happiness$Economy..GDP.per.Capita. and happiness$Happiness.Score
t = -65.809, df = 154, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.500461 -4.238141
sample estimates:
mean of the differences
              -4.369301
```

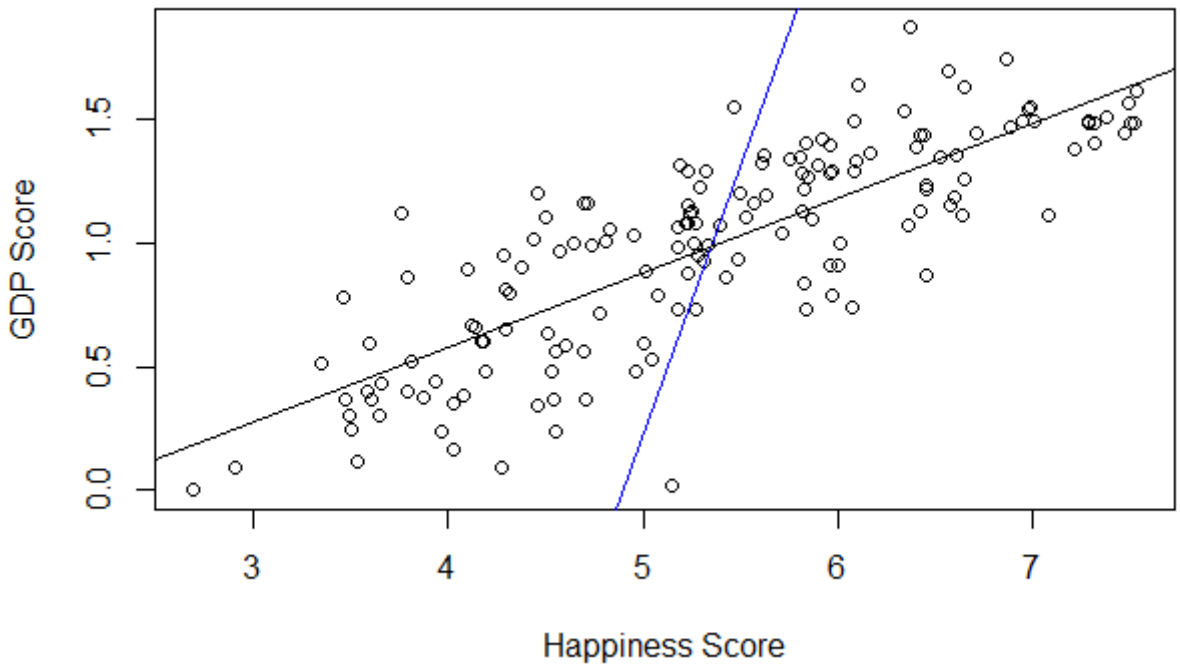
Linear Regression

Hide

```
y2 <- happiness$Happiness.Score^2
x2 <- happiness$Economy..GDP.per.Capita.^2
xy <- happiness$Happiness.Score*happiness$Economy..GDP.per.Capita.
sum_x <- sum(happiness$Economy..GDP.per.Capita.)
sum_y <- sum(happiness$Happiness.Score)
sum_x_sq <- sum(happiness$Economy..GDP.per.Capita.^2)
sum_y_sq <- sum(happiness$Happiness.Score^2)
sum_xy <- sum(happiness$Happiness.Score*happiness$Economy..GDP.per.Capita.)
n <- length(happiness$Economy..GDP.per.Capita.) #Sample size

Lxx <- sum_x_sq-((sum_x^2)/n)
Lyy <- sum_y_sq-((sum_y^2)/n)
Lxy = sum_xy - (((sum_x)*(sum_y))/n)
b = Lxy/Lxx
a = mean(happiness$Economy..GDP.per.Capita. - b*mean(happiness$Happiness.Score))

plot(Economy..GDP.per.Capita. ~ Happiness.Score,
     data = happiness, xlab = "Happiness Score", ylab = "GDP Score")

abline(a = a, b = b, col= "blue")
```

Hide

```
abline(lm(happiness$Economy..GDP.per.Capita. ~ happiness$Happiness.Score))
```



Hide

```
happgdp <- lm( Economy..GDP.per.Capita. ~ Happiness.Score, data = happiness)
happgdp %>% summary()
```

```
Call:
lm(formula = Economy..GDP.per.Capita. ~ Happiness.Score, data = happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-0.90072 -0.16663  0.00354  0.16685  0.61731

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     -0.63338    0.09593  -6.603 6.27e-10 ***
Happiness.Score  0.30222    0.01753  17.238  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2461 on 153 degrees of freedom
Multiple R-squared:  0.6601,    Adjusted R-squared:  0.6579
F-statistic: 297.1 on 1 and 153 DF,  p-value: < 2.2e-16
```

Hide

```
R2 <- (b*Lxy)/Lyy
R2
```

```
[1] 0.6601055
```

Hypothesis Testing

H0:The data do not fit the linear regression model

HA:The data fit the linear regression model

- The p-value turned out to be p<0.001 which is less than 0.05 level of significance.There was statistically signifact evidence that the data fit the linear regression model.

Hide

```
pf(q = 297.1,1,153,lower.tail = FALSE)
```

```
[1] 1.117922e-37
```

Hide

```
(R2/(1-R2)*(153/1))
```

```
[1] 297.1396
```

Hide

```
happgdp %>% anova()
```

```
Analysis of Variance Table

Response: Economy..GDP.per.Capita.
                 Df  Sum Sq Mean Sq F value    Pr(>F)
Happiness.Score   1 17.9999 17.9999  297.14 < 2.2e-16 ***
Residuals       153  9.2683  0.0606
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Hypthesis Testing Cont.

- Intercept/constant is reported as a = -0.633. The value represents the average Economy..GDP.per.Capita score when happiness score is 0.
- To test the statistical significance of the constant, we set the following statistical hypotheses: H0:α=0 HA:α≠0
- The hypothesis is tested using t statistic, reported as t=-6.602591, p<.001. The constant was found statistically significaant at 0.05 level. This concludes that there is statistically significant evidence that the constant is not 0.
- Using Residual vs Fitted graphs it was found that the relationship between fitted and residual values is flat, which provides us with an indication that the modelling is a linear relationship.
- There were no major deviations found from normality using the Q-Q plot
- Scale location red line was found to be flat and standardised residuals are consistent across fitted values.
- Outliers were not found to be influential

Hide

```
happgdp %>% summary() %>% coef()
```

```
                 Estimate Std. Error   t value     Pr(>|t|)
(Intercept)     -0.6333762 0.09592844 -6.602591 6.271248e-10
Happiness.Score  0.3022205 0.01753249 17.237739 1.110391e-37
```

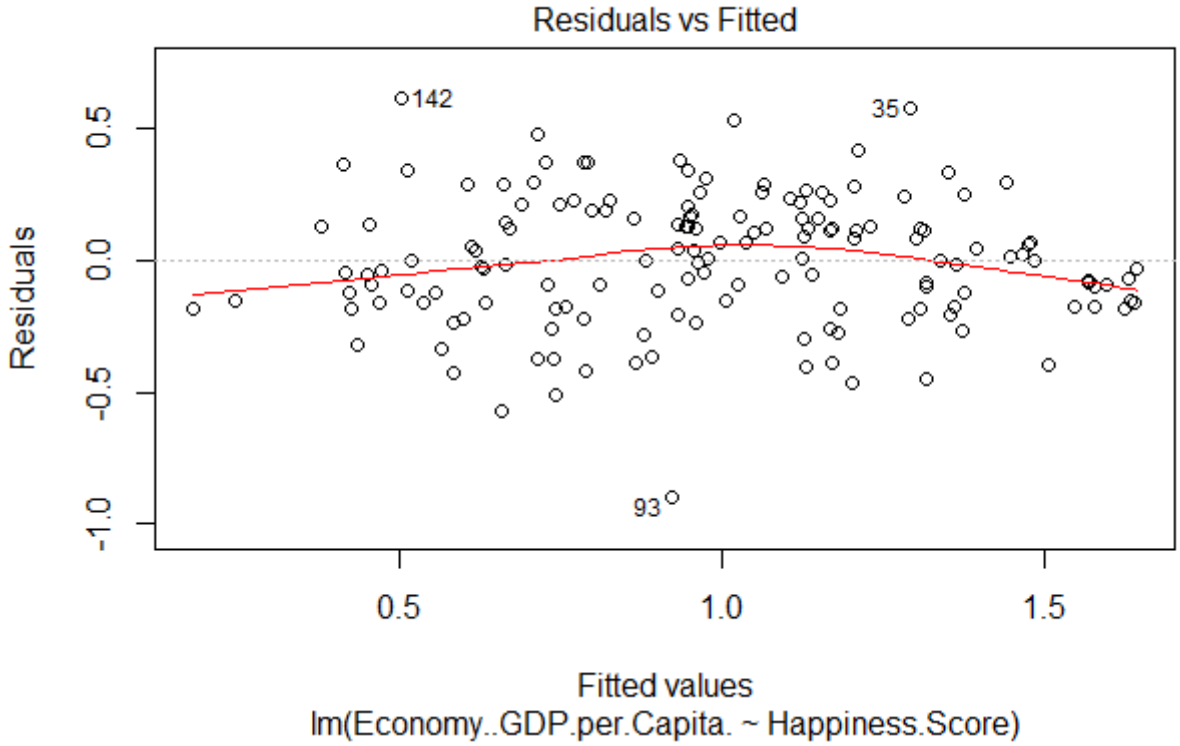
Hide

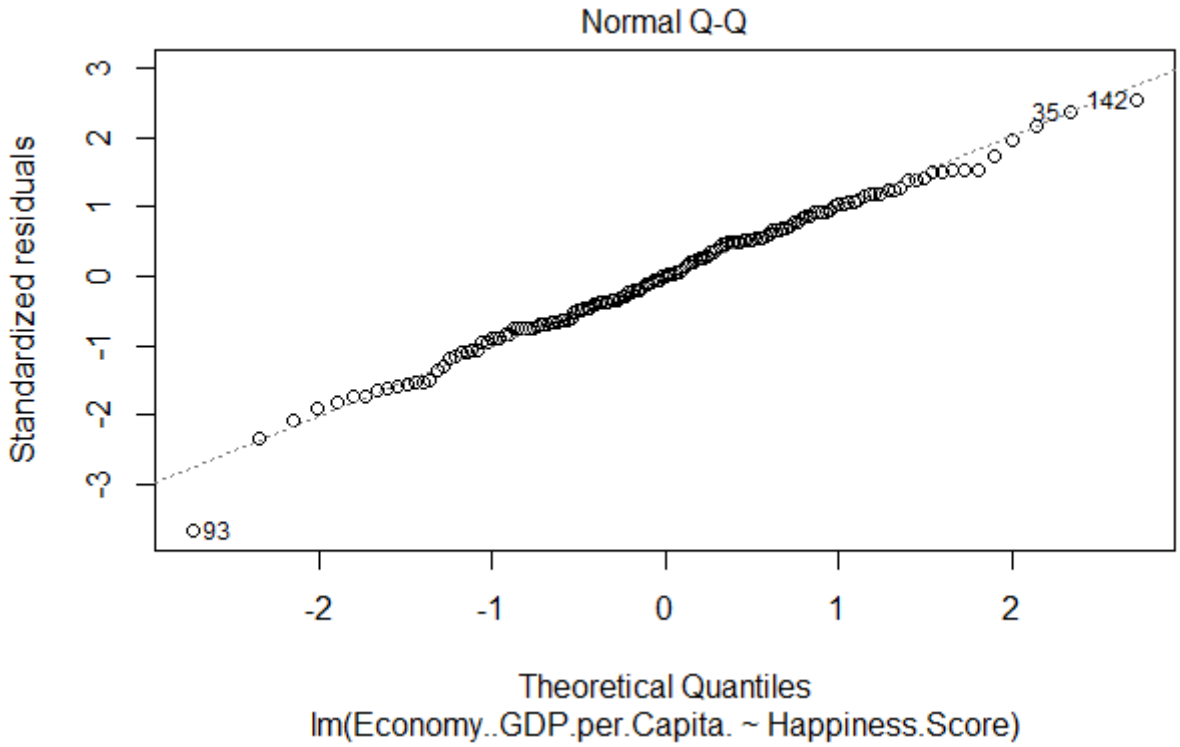
```
happgdp %>% confint()
```

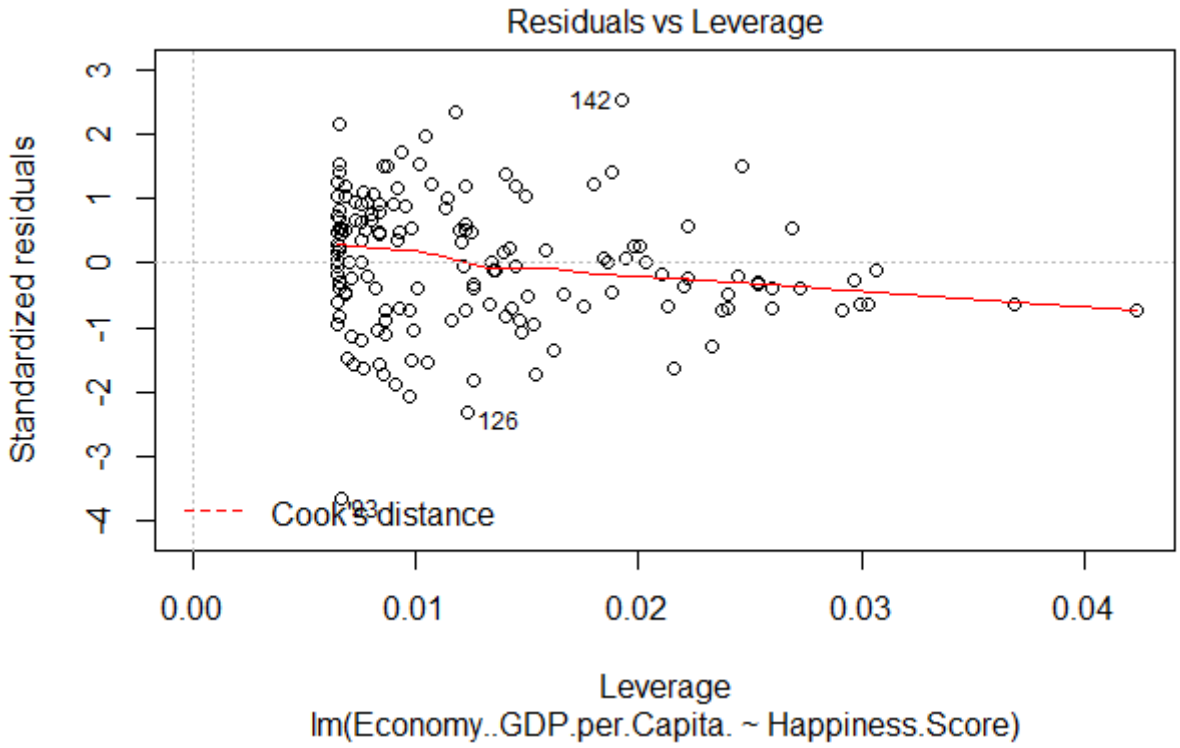
```
                    2.5 %     97.5 %
(Intercept)     -0.8228915 -0.4438609
Happiness.Score  0.2675835  0.3368575
```

Hide

```
plot(happgdp)
```


Residuals vs Fitted
Residuals
142
35
93
0.5
0.0
-0.5
-1.0
0.5
1.0
1.5
Fitted values
lm(Economy..GDP.per.Capita. ~ Happiness.Score)


Normal Q-Q
Standardized residuals
35
142
93
3
2
1
0
-1
-2
-3
-2
-1
0
1
2
Theoretical Quantiles
lm(Economy..GDP.per.Capita. ~ Happiness.Score)



Discussion

After performing certain tests and working on number of hypthesis it was found that there is a strong correlation between the two variables i.e Happiness.Score and Economy..GDP.per.Capita.

Though there is a correlation between happiness and money there are numerous other factors which also need to be taken into account.

It has been found that money alone dosen't affect happiness in pure correlation. There is a causal relationship but not pure.

Countries that rank at last are also affected by wars which is alos a major factor when you take one's happiness into account. No matter how much money you have if you're stranded in middle of a war zone you won't be happy. That's why there are certain limitations. Other factors like family, government trust also comes into account when you are working on a country's happiness score.

In conclusion, there was a strong correlation between wealth and happiness. But other factors also need ot be taken into account to find the pure score of one's happiness, wealth alone is not enough.