R Report 3 (Word Document)

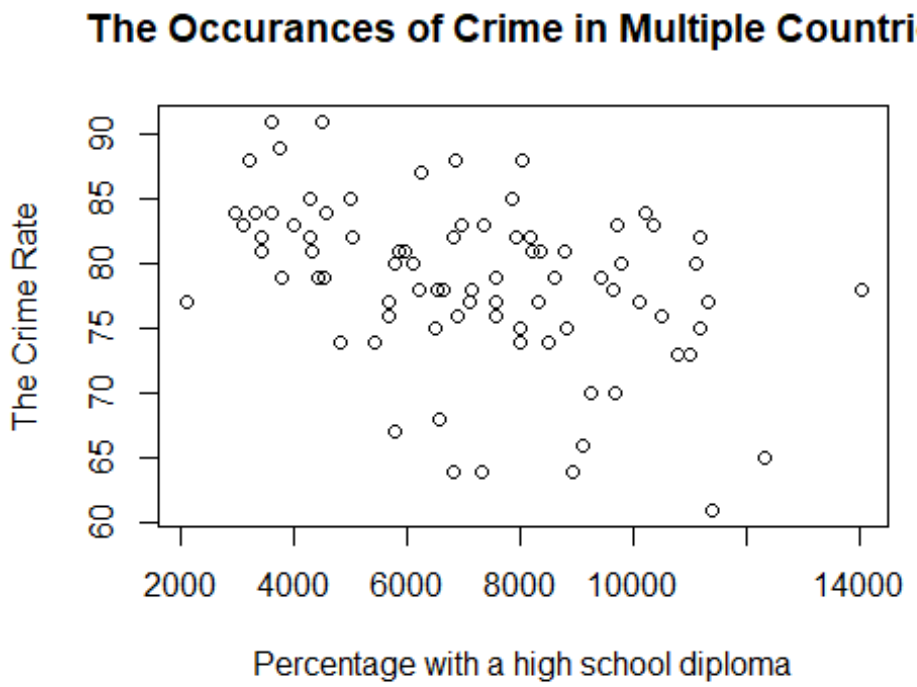
Jeffrey Ugochukwu

6/27/2019

#I

##a

The scatter plot for the the occurnaces of crime in many countries world wide is given below with the X Axis defined as the percentage of individuals in the country with a high school diploma and the y axis being the average crime rate in per country




```
## NULL
```

##b

The estimated regression line for the scatter plot of the the occurnaces of crime in many countries world wide is given below:

```
##
## Call:
## lm(formula = rate ~ dip, data = crime)
##
```

```
## Coefficients:
## (Intercept)          dip
##     20517.6       -170.6
```

##c

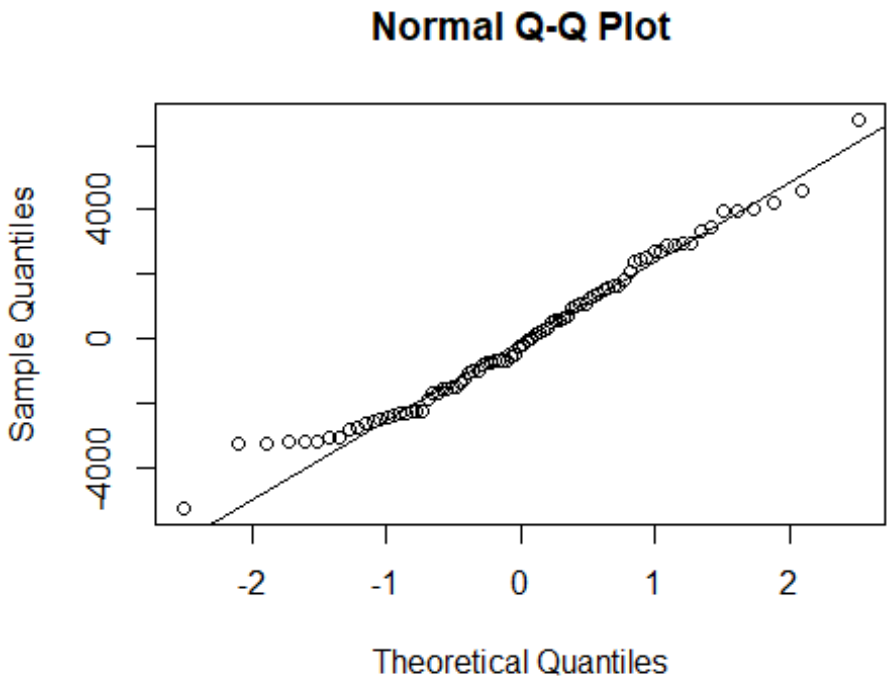
The slope of the regression line for the occurances of crime in multiple countries can be interpreted as the average decrease of the crime rate is 170.6 with a one percent increase of individuals with a high school diploma. The Y-intercept of the regression line for the occurances of crime in multiple countries can't actually be interpreted in this case since the range of the range of the percentage of individuals with a high school diploma doesn't cover zero.

##d

The outliers that appear on the Residuals vs FittedsValues graph of the data of the Multiple Occurances of Crimes world wide would the points (7400,5000) and (7400,-5000)

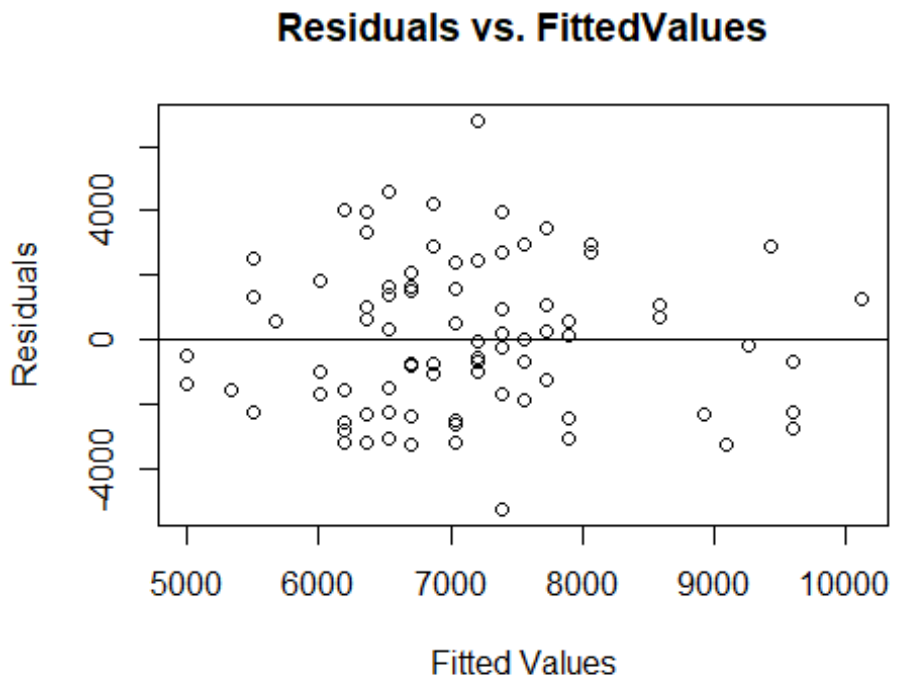
##e

The QQ plot for the occurances of crime in multiple countries is given below. The points for the plot itself does appear to be normally distributed because the majority of the points appear to be close together in straight line on the true line of regression.



##f

The errors vs.the fitted values plot for the occurances of crime in multiple countries is given below. The variance of the errors of this plot does appear to be constant since there points of the errors don't create a distinct pattern on the line of where the y axis is zero.



##g

The interpretation for the confidence interval of the slope for the regression line of the occurances of crime in multiple countries can be interpreted as: I am 95% confident that the average decrease of the crime rate with a one percent increase of individuals with a high school diploma is between -253.2798 and -87.87061. The interval does show a signifcant linear relationship between the individuals with a high school diploma and the crime rate per country because as the amount of indviduals with a high school diploma increases by one percent, the crime rate per country decreases by 170.6.

```
##         2.5 %    97.5 %
## dip -253.2798 -87.87061
```

#II

##a

The null hypothesis (Ho) for the average percent of people with a high school diploma can be determined as: the true average percentof people with a high school diploma is 80%. The alternate hypothesis (Ha) for the average percent of people with a high school diploma can be determined as: the true average percentof people with a high school diploma isn't 80%.

##b

The test statistic for the hypothesis test regarding the average percent of people with a high school diploma is given below:

```
##
##  One Sample t-test
##
## data:  crime$dip
## t = -2.0696, df = 83, p-value = 0.0416
## alternative hypothesis: true mean is not equal to 80
## 95 percent confidence interval:
##  77.24518 79.94529
## sample estimates:
## mean of x
##  78.59524
```

##c

Our p-value for the the test statistic of the hypothesis test regarding the average percent people with a high school diploma is 0.0416 (refrence for the p-value can be found in the data of part b of question 2. The interpretation of the p-value for the the test statistic of the hypothesis test regarding the average percent of people with a high school diploma can be stated as: the probability that the true average percentof people with a high school diploma is 80% is 0.0416.

##d

Since our p-value (0.0416) is less than our signifcance level (alpha = 0.10), we reject the null hypothesis (the claim that the true average percent of people with a high school diploma is 80%).

##e Since we rejected the null hypothesis (the claim that the true average percent of people with a high school diploma is 80%), there's convincing evidence that the true average percent of people with a high school diploma isn't 80%.

##f

(Calculation of the 90% confidence interval: X = 80, df = 83, n = 84, t = -2.0696, sd = 6.221066 Formula: x +/- t*(sd/sqrt(n)). Work: 80 +/- -2.0696*(6.221066/sqrt(84)) = Interval: (77.46615,79.72432)) When I calculated the 90% confidence interval of finding the true average percent of people with a high school diploma (the interval being between 77.466165 and 79.72432), this does support my decision of rejecting the null hypothesis because the interval for the true average is less than the value of the claim being that the true average is at 80%; therefore, the true average percent of people with a high school diploma isn't at 80%.

##g

Based off of the result of rejecting the null hypothesis (the claim that the true average percent of people with a high school diploma is 80%), I have made a Type I error. This would mean that we would assume that the true average percentage of people with a high school diploma isn't at 80%, but it actually is.

#Appendix

```
crime <- read.csv("C:/Users/ugoch/Downloads/crime.csv")
crime_data = crime
crime_hist = plot(crime_data, main = "The Occurances of Crime in Multiple
Countries", xlab = "Percentage with a high school diploma", ylab = "The Crime
Rate")
crime_hist
crime_regression = lm(rate~dip, data = crime)
crime_regression
qqnorm(crime_regression$residuals)
qqline(crime_regression$residuals)
plot(crime_regression$fitted.values,crime_regression$residuals, xlab =
"Fitted Values", ylab= "Residuals", main = "Residuals vs. FittedValues")
abline(h=0)
CI = confint(crime_regression, parm = "dip", level = 0.95)
CI
t.test(crime$dip, mu = 80)
```