RMarkdown Assignment Template

Osmond Oke

June 28th 2020

## A. Use R to calculate the covariance of the Survey variables and provide an explanation of why you would use this calculation and what the results indicate.

## [1] -20.36364

The covariance calculation is used to determine the direction of the linear relationship between two variables.The value of -20.36364 simply indicates the strength of the negative linear relationship. A negative covariance indicates that one value is reducing as the other value increases.

## B.Examine the Survey data variables. What measurement is being used for the variables?

The measurement being used is time, with TimeReading probably being measured in hours, and TimeTv probably being measured in minutes.

* Explain what effect changing the measurement being used for the variables would have on the covariance calculation.

Changing the time would not necessarily have a large impact on the covariance calculation, as the number will definitely change, but the shape/trendline will not really be affected.

* Would this be a problem?

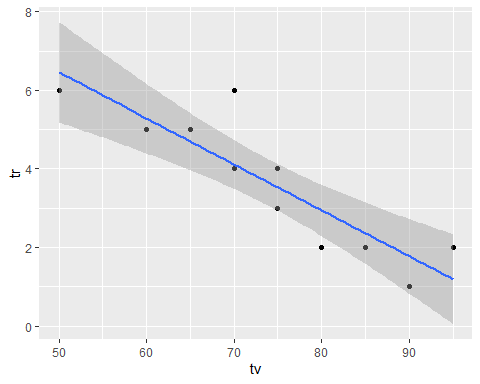
This could be a problem because it might show results that would appear opposite from the actual results.

* Explain and provide a better alternative if needed.

In my opinion, a better way around it would be to convert both values (hours and minutes) into the same units, in order to get more accurate results.

## C.Choose the type of correlation test to perform

## `geom\_smooth()` using formula 'y ~ x'



## [1] -0.8830677

##   
## Pearson's product-moment correlation  
##   
## data: tr and tv  
## t = -5.6457, df = 9, p-value = 0.0003153  
## alternative hypothesis: true correlation is not equal to 0  
## 95 percent confidence interval:  
## -0.9694145 -0.6021920  
## sample estimates:  
## cor   
## -0.8830677

## function (x, ...)   
## UseMethod("cor.test")  
## <bytecode: 0x00000000141f8570>  
## <environment: namespace:stats>

The test to use would be the Pearson Test

* Explain why you chose this test

I chose the pearson test because the data (as observed in the plot) is normally distributed ratio data, so the pearson test would be ideal.

* Make a prediction if the test yields a positive or negative correlation?

The test yields a negative linear correlation, as the TimeTV value reduces, as the TimeReading value increases.

## D. Perform a correlation analysis of:

* All variables

## TimeReading TimeTV Happiness Gender  
## TimeReading 1.00000000 -0.883067681 -0.4348663 -0.089642146  
## TimeTV -0.88306768 1.000000000 0.6365560 0.006596673  
## Happiness -0.43486633 0.636555986 1.0000000 0.157011838  
## Gender -0.08964215 0.006596673 0.1570118 1.000000000

* Single correlation between a pair of the variables

Correlation between TimeReading and TimeTV:

## [1] -0.8830677

Correlation between TimeReading and Happiness:

## [1] -0.4348663

Correlation between TimeTV and Happiness:

## [1] 0.636556

* Repeat your correlation test in step 2 but set the confidence interval at 99%

##   
## Pearson's product-moment correlation  
##   
## data: tr and tv  
## t = -5.6457, df = 9, p-value = 0.0003153  
## alternative hypothesis: true correlation is not equal to 0  
## 99 percent confidence interval:  
## -0.9801052 -0.4453124  
## sample estimates:  
## cor   
## -0.8830677

* Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

## [,1]  
## [1,] 1

The results obtained from the correlation matrix indicate that there is a perfect linear relationship between the variables, as they possess a correlation coefficient of 1

## E. Calculate the correlation coefficient and the coefficient of determination,

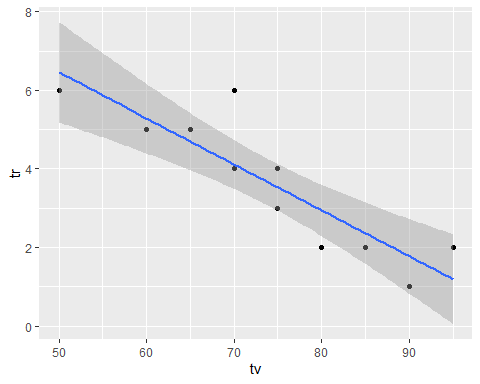
## TimeReading TimeTV Happiness Gender  
## TimeReading 1.000000000 0.7798085292 0.18910873 0.0080357143  
## TimeTV 0.779808529 1.0000000000 0.40520352 0.0000435161  
## Happiness 0.189108726 0.4052035234 1.00000000 0.0246527174  
## Gender 0.008035714 0.0000435161 0.02465272 1.0000000000

* Describe what you conclude about the results

The coefficient of determinant tells us by how much the dependent variable can be predicted from the independent variables. The results show a strong relationship between the TimeTV and the TimeReading, but shows a weak relationship between both of those variables and happiness, and an even weaker relationship between them and Gender. This means that Happiness and Gender have almost no impact on the TimeReading and TimeTV.

## F. Based on your analysis can you say that watching more TV caused students to read less? Explain.

## `geom\_smooth()` using formula 'y ~ x'



From the data observed in the plot, it can be inferred that watching more tv caused students to read less. This is because the plot depicts the strong negative linear relationship between the time spent reading and the time spent watching tv; as the time spent watching tv increased, the time spent reading reduced.

## G. Pick three variables and perform a partial correlation, documenting which variable you are “controlling”.

Performing a partial correlation of TimeTV and TimeReading, controlling Happiness:

pc

## [1] -0.872945

pc^2\*100

## [1] 76.2033

pcor.test(pc,1,11)

## $tval  
## [1] -5.061434  
##   
## $df  
## [1] 8  
##   
## $pvalue  
## [1] 0.0009753126

```

* Explain how this changes your interpretation and explanation of the results.

This has a very minimal impact on the results, meaning that Happiness has almost no effect on the relationship between TimeTV and TimeReading.