ASSIGNMENT 5

Joshua Greenert

7/23/2022

Set the working directory.

```
knitr::opts_knit$set(root.dir =
normalizePath("C:/Users/Josh/Documents/GitHub/R-Projects/Statistics For Data
Science"))
```

Show the head of the data.

```
##   TimeReading TimeTV Happiness Gender
## 1           1     90     86.20      1
## 2           2     95     88.70      0
## 3           2     85     70.17      0
## 4           2     80     61.31      1
## 5           3     75     89.52      1
## 6           4     70     60.50      1
```

Show the covariance

```
##               TimeReading       TimeTV  Happiness      Gender
## TimeReading   3.05454545 -20.36363636 -10.350091 -0.08181818
## TimeTV      -20.36363636 174.09090909 114.377273  0.04545455
## Happiness   -10.35009091 114.37727273 185.451422  1.11663636
## Gender       -0.08181818   0.04545455   1.116636  0.27272727
```

Using this calculation allows us to see the relationships between the data. When there is a positive covariance, the two sets of data appear to move together; whereas, if the covariance is negative, the two sets of data move away from one another. The results indicate that the variables TimeTV and Happiness appear to have a positive covariance while TimeReading and Happiness have a negative covariance.

Examine the Survey data variables

TimeReading - appears to be in units of hours per week.
TimeTV - appears to be in units of hours per week.
Happiness - appears to be a percentage of overall happiness reported.
Gender - appears to be binary between male or female.

Since both measurements of time are in the same units, the covariance would not be impacted on the result of comparing the two measurements. When comparing these against a percentage, however, there may be a significant difference in the output expected. Nevertheless, the variables are suitable as they are since they show a decent reflection of the dataset.

Choose a correlation test to perform: pearson

This test is a wise choice as it measures the strength of the linear relationship between normally distributed variables and is the default for most correlation tests.

Perform a correlation analysis

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 pair of variables

```
## [1] "Correlation between tv and happiness"

## [1] 0.636556

## [1] "Correlation between reading and happiness"

## [1] -0.4348663
```

Repeat the correlation test with confidence internal at 99%

```
## [1] "Correlation between tv and happiness with Confidence intervals"

##
##  Pearson's product-moment correlation
##
## data:  surveyResults$TimeTV and surveyResults$Happiness
## t = 2.4761, df = 9, p-value = 0.03521
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
##  -0.1570212  0.9306275
## sample estimates:
##      cor
## 0.636556

## [1] "Correlation between reading and happiness with Confidence intervals"

##
##  Pearson's product-moment correlation
```

```
##
## data:  surveyResults$TimeReading and surveyResults$Happiness
## t = -1.4488, df = 9, p-value = 0.09067
## alternative hypothesis: true correlation is less than 0
## 99 percent confidence interval:
##  -1.0000000  0.3422209
## sample estimates:
##        cor
## -0.4348663
```

Describe what the calculations suggest

The correlation tests provided data that suggested it was likely that there was both a negative and positive correlation among the variables chosen. Further tests with confidence intervals also suggest that there is a positive correlation between tv time and happiness, while there is a negative correlation between reading and happiness.

Calculate the correlation coefficient and coefficient of determination; describe the results

```
## [1] "Correlation coefficient between tv and happiness"

## [1] 40.52035

## [1] "Correlation coefficient between reading and happiness"

## [1] 18.91087
```

From the data above, we can assess that tv is positively correlated with happiness but that it can only account for 40.52% of the variation in terms of happiness. Alternatively, we can assess that reading is negatively correlated with happiness and that it can only account for 18.91% of the variation in terms of happiness. There remains 59.48% variability for tv and 81.09% variability that is unexplained by the two samples provided.

Can you say that watching more TV caused students to read less?

The data that is provided doesn't conclude whether Tv engagement effected reading. Instead, the data suggests happiness may not be as prevalent within reading as it is in TV.

Partial Correlation Test with "reading" set as the control

```
## [1] 0.5976513

## [1] "The result squared"

## [1] 0.3571871
```

From the data, we can determine that reading accounted for approximately 35.72% of the variability between happiness and TV. Therefore, it does have an effect on the results and suggests that TV usage may cause students to read less.