Week 4 Assignment

Kurt Stoneburner

June 22nd 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.

Covariance

```
##             TimeReading TimeTV Happiness Gender
## TimeReading        3.05 -20.36    -10.35  -0.08
## TimeTV           -20.36 174.09    114.38   0.05
## Happiness        -10.35 114.38    185.45   1.12
## Gender            -0.08   0.05      1.12   0.27
```

The covariance is a measure of how one variable varies compared to another variable. A primary purpose of this calculation is to measure if a variable varies positively or negatively in relation to another variable. For example, TimeReading has a negative covariance with both TimeTV and Happiness. This can be interpreted as the more time is spent reading, the less time is used watching Television and Happiness is decreased.

The covariance is suitable for measuring how one variable does or does not influence another variables. This is one method used to observe how one variable may affect another. However, it is difficult to ascertain the strength of the relationship unless all the values are in the same unit of measurement.

The results indicate: As TimeReading Increases - TimeTV Decreases & Happiness Decreases As TimeTV Increases - TimeReading Decreases & Happiness Increases As Happiness Increases - TimeReading Decreases & TimeTV Increases

Gender does not appear to significantly affect: TimeReading, TimeTV, or Happiness.

B. Examine the Survey data variables. What measurement is being used for the variables? TimeReading and TimeTV appear to be discrete measurements of time. The units of measured time are unknown. Nor can it be assumed that both variables are using the same units of time. Likely, TimeReading is in Hours per day, and TimeTV is in minutes per day. But this is purely a guess.

Explain what effect changing the measurement being used for the variables would have on the covariance calculation. Would this be a problem? Explain and provide a better alternative if needed. Changing values will affect the covariance values. For example, if TimeReading was in hours, it could be converted to minutes by multiplying each value by 60. The actual covariance values increase greatly as well.

The student survey would look like this:

```
##    TimeReading(mins) TimeTV(mins) Happiness Gender
## 1                 60           90     86.20      1
## 2                120           95     88.70      0
```

```
## 3                120           85     70.17      0
## 4                120           80     61.31      1
## 5                180           75     89.52      1
## 6                240           70     60.50      1
## 7                240           75     81.46      0
## 8                300           60     75.92      1
## 9                300           65     69.37      0
## 10               360           50     45.67      0
## 11               360           70     77.56      1
```

The resulting covariance table:

```
##                    TimeReading(mins) TimeTV(mins) Happiness Gender
## TimeReading(mins)          10996.36     -1221.82   -621.01  -4.91
## TimeTV(mins)               -1221.82       174.09    114.38   0.05
## Happiness                   -621.01       114.38    185.45   1.12
## Gender                        -4.91         0.05      1.12   0.27
```

I do not believe changing the units of measurement would affect the interpretation of covariance. Since it is difficult to measure the intensity of the covariance the relative values themselves appear to be less important. The prior observations remain true (whether a variable positively or negatively affects another), just at a different unit of measurement.

C. Choose the type of correlation test to perform, explain why you chose this test, and make a prediction if the test yields a positive or negative correlation?

I would correlate happiness and TimeReading. In relative units, correlating TimeReading and happiness generated a greater difference in value from correlating Happiness with itself. Happiness had a self correlated value of 185, and happiness correlated with TimeReading was -10. I predict that TimeReading has a negative correlation on Happiness. I would use Kendall's tau to generate a correlation coefficient since it works best with small data sets of non-parametric data.

D1. Perform a correlation analysis of all variables:

```
##             TimeReading      TimeTV   Happiness      Gender
## TimeReading  1.00000000 -0.80454045 -0.28894280 -0.07824608
## TimeTV      -0.80454045  1.00000000  0.46304237 -0.02507849
## Happiness   -0.28894280  0.46304237  1.00000000  0.09847319
## Gender      -0.07824608 -0.02507849  0.09847319  1.00000000
```

D2. A single correlation between a pair of the variables

```
##
##  Pearson's product-moment correlation
##
## data:  survey_df$TimeReading and survey_df$Happiness
## t = -1.4488, df = 9, p-value = 0.1813
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.8206596  0.2232458
## sample estimates:
##        cor
## -0.4348663
```

D3. Repeat your correlation test in step 2 but set the confidence interval at 99%

```
##
##  Pearson's product-moment correlation
##
## data:  survey_df$TimeReading and survey_df$Happiness
## t = -1.4488, df = 9, p-value = 0.1813
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
##  -0.8801821  0.4176242
## sample estimates:
##        cor
## -0.4348663
```

D4. Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

The correlation matrix suggests the following relationships:

- TimeTV and Happiness: Moderate positive relationship
- TimeTV and TimeReading: Strong negative relationship.
- TimeReading and Happiness: Weak negative relationship.

D5. Calculate the correlation coefficient and the coefficient of determination, describe what you conclude about the results.

Using Kendall's Tau the correlation coefficient of TimeTV and TimeReading is:

```
## [1] -0.8045404
```

The coefficient of determination for TimeTV and TimeReading is:

```
## [1] "64.73%"
```

D6. Based on your analysis can you say that watching more TV caused students to read less? Explain.

There is a strong negative correlation between Watching Tv and Time spent reading. Time spent on one activity correlates to less time spent on the other. This makes intuitive sense since time spent on one activity cannot be spent on another (assuming only one task is performed at a time). However, the time allotments for TV and Reading represent a small portion of the day. Time spent engaged in watching TV or Reading does not necessarily represent the totality of a student's available time (at least one would certainly hope not), performing one activity (Watching TV) may limit the opportunity to access another (Reading), it cannot be the sole determinant of how a student's time is spent. The coefficient of determination of 64.73% represents the shared variance between TimeReading and TimeTV. This suggests a moderately strong negative relationship between reading and watching TV. This isn't a perfect relationship, there are other factors that influence how a student spends their time. These unaccounted factors (tertium quid) represent an influence 35.27% on student time.

I believe, the more time spent watching TV decreases the likelihood that a student will spend time reading.

D7. Pick three variables and perform a partial correlation, documenting which variable you are "controlling". Explain how this changes your interpretation and explanation of the results. The survey does not reflect how a student's happiness may affect Time spent reading or watching TV. Happiness can be used as a control in a partial correlation between TimeTV and TimeReading which results in an R value of:

```
## [1] -0.872945
```

Which indicates a small increase in correlation.

The resulting coefficient of determination

```
## [1] "76.2%"
```

indicates a stronger link between TimeTV and TimeReading.

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

The resulting p value is less than .05 indicating this measure is significant.

When controlling for happiness the negative correlation between TimeTV and TimeReading slightly increases from -0.804 to -0.872 [1]. However, there is a significant increase in the coefficient of determination from 64.73% to 76.2%.

By these measures, I'd feel comfortable stating that there is a strong indication that Time spent reading decreases the time spent watching TV[2].

[1] I realize -0.872 is actually smaller than -0.804. The only way the value increases is if I'm using a double negative: Negative correlation and a negative number.

[2] The reverse is true as well. Time spent watching TV reduces reading time.