

Assignment 4.1: Student Survey

Anna Harvey

6/28/2020

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:

```
##           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
```

You would use the covariance calculation to help determine if there is a significant relationship between two variables. If the variables have a relationship, they should vary from the mean in the same direction (if positive) or directly opposite each other (if negative). TimeTv and Happiness, Gender and Happiness, and Gender and TimeTv all have positive covariance with each other. TimeReading and TimeTv, TimeReading and Happiness, and TimeReading and Gender all have negative covariance with each other.

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

Since there is no information available about the measurements of the variable, we have to make some assumptions. It is clear that TimeTv and TimeReading are being measured differently. Most likely, TimeReading is being measured in hours while TimeTv is being measured in minutes. There's no indication of how Happiness or Gender are being measured. Covariance changes depending on the measurement of the variable and won't really work well if the measurements are not the same. If we changed TimeTv to hours, it would cause the covariance to increase in positivity. If we changed TimeReading to minutes, it would cause the covariance to increase in negativity. Since the measurement isn't standardized, this isn't a good measurement for this comparison. A better alternative would be using the correlation coefficient. **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 will use a standard Pearson correlation because we want to look at the correlation between two variables (TimeRead and TimeTv) and the variables are both continuous and quantitative. I think the test will yield a negative correlation.

d. Perform a correlation analysis of:

1. 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
```

2. A single correlation between two a pair of the variables

Correlation between TimeReading and TimeTv is:

```
## [1] -0.8830677
```

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

```
##
## Pearson's product-moment correlation
##
## data: survey$TimeReading and survey$TimeTV
## 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
```

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

There is a strong positive relationship between TimeTv and Happiness. There is a strong negative relationship between TimeTv and TimeReading. There is a medium negative relationship between Happiness and TimeReading. There is a small positive relationship between Gender and TimeTv as well as Gender and Happiness. There is a small negative relationship between Gender and TimeReading.

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

```
##          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
```

```
##          TimeReading      TimeTV  Happiness      Gender
## TimeReading  1.000000000  0.7798085292  0.18910873  0.0080357143
## TimeTV      0.779808529  1.00000000000  0.40520352  0.0000435161
## Happiness   0.189108726  0.4052035234  1.000000000  0.0246527174
## Gender      0.008035714  0.0000435161  0.02465272  1.00000000000
```

The coefficient of determination is the percentage that the amount of variability in one variable that is shared by the other. Since the coefficient of determination for TimeTv and TimeReading is approximately 78%, that affirms that there is a strong relationship between the two.

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

There seems to be a strong negative relationship between TimeTv and TimeReading. However, there are several outside variables that could have influenced this perceived relationship, including Gender and Happiness. It is also problematic that TimeTv and TimeReading were measured differently. There could be a reason why that happened that influenced the outcome.

g. 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.

```
pcor(c("TimeReading", "TimeTV", "Gender"), var(survey)) #Gender is the control variable
```

```
## [1] -0.8860628
```

Using Gender as the control variable, it increases the strong negative relationship between TimeReading and TimeTv, but only marginally. This could provide more evidence that the relationship between the two actually exists apart from the other variables, however, the increase in relationship is very small. If this negative relationship actually exists, this would suggest that more time spent watching TV would decrease the time spent reading, and vice versa.