ASSIGNMENT 4.1

Blaine Blasdell

2020-06-28

Assignment 4.1

Question 4.1a

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.

A4.1a. Covariance tells us the combined deviation from the mean for the two variables.

This will tell us if the two variables have a relation. Reading and TV watching, reading and happiness, and reading and gender all have negative covariance meaning they have a negative relation. TV watching and happiness, TV watching and gender, and happiness and gender all have a positive covariance and positive relationship. At first look, reading and watching TV have a negative relationship, in that the more the student reads, the less likely they are to watch TV.

```
read_tv
```

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

```
read_hap
```

```
## [1] -0.4348663
```

```
read_gen
```

```
## [1] -0.08964215
```

```
tv_hap
```

```
## [1] 0.636556
```

```
tv_gen
```

```
## [1] 0.006596673
```

```
hap_gen
```

```
## [1] 0.1570118
```

Q4.1b. 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.

##A4.1b. Part 1

1. Variables - Measure
2. Time Reading - Hours
3. Time Watching TV - Minutes
4. Happiness - Percentage
5. Gender - Binary (0/1) for Male/Female

##A4.1b. Part 2
Covariance does not use a standard scale so changing the measurement would change the covariance and would give you different values.
Standardizing the values such as converting the time reading to minutes would give a better representation between time reading and watching TV.
The interesting thing is that I converted the reading time to minutes and the covariance did not change for the Reading verse time watching TV. A better option would be to use a correlation coffeicient instead of covariance, as it accounts for standardization.

```
readmin_tv
```

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

Q4.1c. 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?

##A4.1c.
I chose to use Kendall's tau as it is a very small sample size. I believe, which seems to make sense to me that this will be a negative correlation as I would guess, the more you watch TV the less you would read.

```
read_to_tv
```

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

As I predicted this is a negative correlation. Since the value is greater then .5, this is a strong correlation or relation.

Q4.1d. Perform a correlation analysis of:

A4.1d. Part 1 - All Variables

```
read_to_tv_all
```

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

A4.1d. Part 2 - A single correlation between two a pair of the variables

```
read_to_tv
```

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

A4.1d. Part 3 - Repeat your correlation test in step 2 but set the confidence interval at 99%

```
read_to_tv_con
```

```
##
##  Pearson's product-moment correlation
##
## data:  student_df$MinRead and student_df$TimeTV
## t = -5.6457, df = 9, p-value = 0.0001577
## alternative hypothesis: true correlation is less than 0
## 99 percent confidence interval:
##  -1.0000000 -0.5131843
## sample estimates:
##        cor
## -0.8830677
```

A4.1d. Part 4 - Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

The correlation Matrix suggests:

1. There is a strong negative correlation between reading and watching television, meaning the more they watch television the more less likely they are to read.

2. Interestingly enough there a moderate negative relationship between reading and the students happiness.

3. There is an extremely weak negative correlation between time reading and the gender.

4. There is a strong positive correlation between watching television and the students happiness.

5. There is an extremely weak positive correlation between gender of the student and how much TV they watched.

6. There is an extremely weak positive relationship between gender of the student and how happy they were.

A4.1e. Calculate the correlation coefficient and the coefficient of determination, describe what you conclude about the results..

```
read_to_tv_cor
```

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

```
read_to_tv_coef
```

```
## [1] 77.98085
```

```
cor_mat_all
```

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

```
cor_mat_all_coef
```

```
##             TimeReading       TimeTV  Happiness       Gender     MinRead
## TimeReading 100.0000000  77.98085292  18.910873   0.80357143 100.0000000
## TimeTV       77.9808529 100.00000000  40.520352   0.00435161  77.9808529
## Happiness    18.9108726  40.52035234 100.000000   2.46527174  18.9108726
## Gender        0.8035714   0.00435161   2.465272 100.00000000   0.8035714
## MinRead     100.0000000  77.98085292  18.910873   0.80357143 100.0000000
```

While there is a very strong negative correlation between reading and watching television; reading shares a 78% variability with watching television. So 22% of other factors could affect the relationship between reading and watching television.

A4.1f. Based on your analysis can you say that watching more TV caused students to read less? Explain. . .

Based on this analysis, the students who watched television were less likely to read. And students who read more, were less likely to watch television. There is a negative relationship between the two variables based on the correlation. Based on the coefficient of determination there is only about 22% of variability in other factors, such as gender and happiness.

A4.1f. 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.

```
pc
```

```
## [1] -0.4277985
```

```
pc^2*100
```

```
## [1] 18.30116
```

```
pcor.test(pc,1,11)
```

```
## $tval
## [1] -1.338679
##
## $df
## [1] 8
##
## $pvalue
## [1] 0.2174682
```

I looked at a partial correlation between the students happiness and the amount of time reading with gender as a control. There is a moderate negative partial relationship between the amount they read and how happy they were. Not including the control, 18.3% of the variance in students reading amount can account for their happiness.(Field, Miles, and Field 2012)

References

Field, A., J. Miles, and Z. Field. 2012. *Discovering Statistics Using R*. SAGE Publications. https://books.google.com/books?id=wd2K2zC3swIC.