

Exercise 7.2

Amelia Farrell

October 11th 2021

Student Survey

i. Covariance of Survey variables

Covariance shows the direction of a liner relationship between two variables. This can be used to see the direction the relationship between two variables on the same scale. Such as, time reading and time TV. Looking at the data below we can infer that as time spent reading goes up, time spent watching TV decreases. Same can be said for Happiness, however we cannot infer anything based off the absolute vaule since TimeReading and Happiness are on a differet scale than TimeTV variable.

##	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

ii. Examining the Survey data variables

As we can see below, each variable is measured on a different scale. This can cause issues with our Covariance measure.

##	TimeReading	TimeTV	Happiness	Gender
## Min.	:1.000	Min. :50.00	Min. :45.67	Min. :0.0000
## 1st Qu.:	:2.000	1st Qu.:67.50	1st Qu.:65.34	1st Qu.:0.0000
## Median	:4.000	Median :75.00	Median :75.92	Median :1.0000
## Mean	:3.636	Mean :74.09	Mean :73.31	Mean :0.5455
## 3rd Qu.:	:5.000	3rd Qu.:82.50	3rd Qu.:83.83	3rd Qu.:1.0000
## Max.	:6.000	Max. :95.00	Max. :89.52	Max. :1.0000

If we converted TimeReading to minuets, we would get a different Covariance regarding it's relationship to other variables. As you can see from the example below TimeReading and TimeReadingMins has a different Covariance due to the change in scale (from Hours to Mineunts). However, we can still infer the direction of the relationship by looking for positive or negative vaule. A solution to this would be to take it a step futher and look at the corrilation, not only the covariance.

```
TimeReadingMins <- (survey$TimeReading)*60
survey <- cbind(survey, TimeReadingMins)
cov(survey)
```

```
##           TimeReading      TimeTV  Happiness      Gender
## TimeReading      3.05454545 -2.036364e+01 -10.350091 -0.08181818
## TimeTV           -20.36363636  1.740909e+02  114.377273  0.04545455
## Happiness        -10.35009091  1.143773e+02  185.451422  1.11663636
## Gender           -0.08181818  4.545455e-02   1.116636  0.27272727
## TimeReadingMins  183.27272727 -1.221818e+03 -621.005455 -4.90909091
##           TimeReadingMins
## TimeReading      183.272727
## TimeTV           -1221.818182
## Happiness        -621.005455
## Gender           -4.909091
## TimeReadingMins  10996.363636
```

iii. Correlation Test

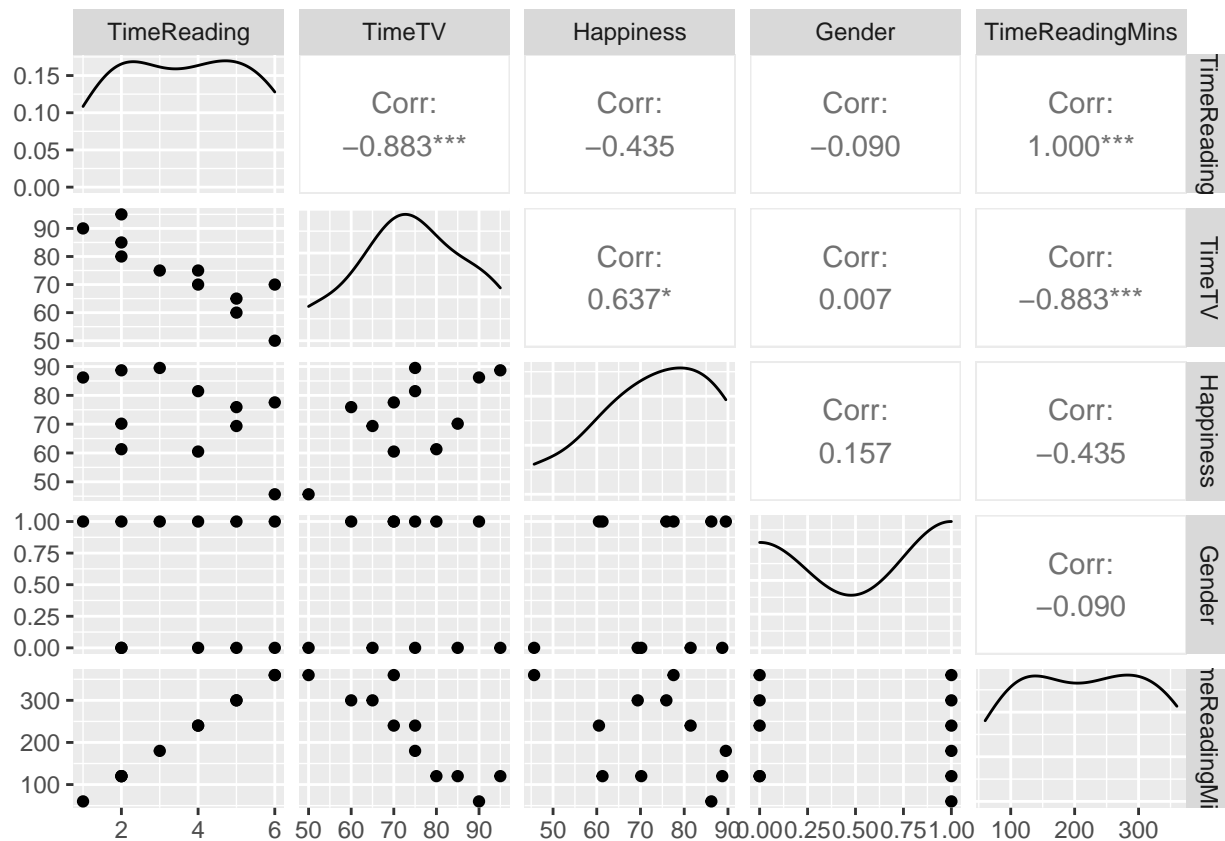
We will be using a Pearson correlation to test the correlation between the variables since they are continuous or binary variables. Based on the covariance calculation, we can predict that the positive and negatives will remain the same and only the absolute value should change.

iv. Correlation Analysis

1. All variables - After running a Pearson correlation test on all variables, you can see the correlation coefficient did not change when the scale did (i.e. TimeReading vs TimeReadingMins), further displaying the importance of scale in regards to covariance.

```
##           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
## TimeReadingMins  1.00000000 -0.883067681 -0.4348663 -0.089642146
##           TimeReadingMins
## TimeReading      1.00000000
## TimeTV           -0.88306768
## Happiness        -0.43486633
## Gender           -0.08964215
## TimeReadingMins  1.00000000
```

We can also visualize the correlation using the ggpairs function from ggplot2.



- Single correlation - Since other variables can effect the correlation, lets run a correlation between TimeTV and Happiness and TimeReading and Happiness independently.

TimeTV and Happiness

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

TimeReading and Happiness

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

We can see above that the single correlation is slightly different than the correlation coefficient that included all variables. However, it still indicates a moderate positive correlation between TimeTV and Happiness and a low negative correlation between TimeTV and Happiness.

- Confidence interval at 99% - Now lets look at the same variables but set the Confidence interval to 99%.

TimeTV and Happiness

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

```
## data: survey$TimeTV and survey$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
```

TimeReading and Happiness

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

As you can see above, even though TimeTV and Happiness had a stronger correlation (0.6) the correlation actually had no significance ($p=0.181$). Whereas TimeReading and Happiness has statistical significance ($p=0.035$)

4. Correlation matrix meaning - The above correlation matrix also tells us that there is a strong negative correlation (or relationship) (-0.88) between time spent reading and time spent watching TV. Meaning that as one variable increases the other decreases. There also appears to be a moderate positive correlation between time spent watching TV and reported happiness (0.63). As well as a low negative correlation between time spent reading and reported happiness (-0.43). When only looking at these correlations, one may think that watching TV increases happiness. However, we must be aware that anything under 0.5/-0.5 is a low correlation and Correlation Coefficients do not indicate causation.

v. Correlation coefficient and the coefficient of determination

Now let's look at how TimeTV and TimeReading are related more closely.

Correlation coefficient of TimeTV and TimeReading

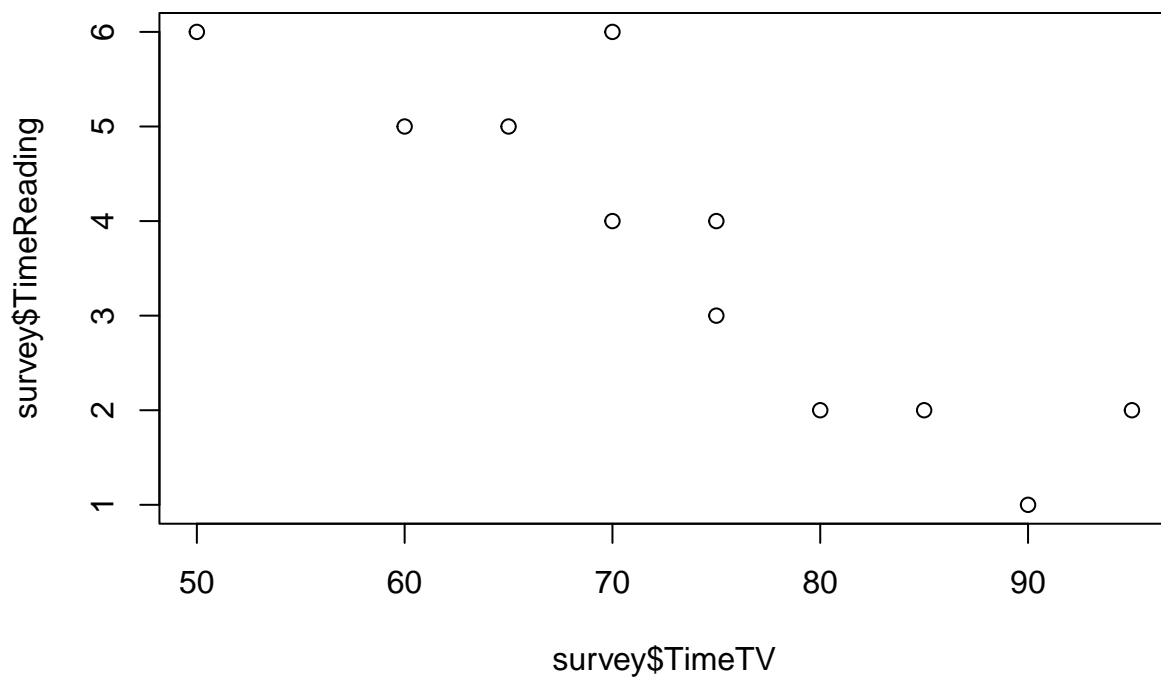
```
##
## Pearson's product-moment correlation
##
## data: survey$TimeTV and survey$TimeReading
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## -0.9801052 -0.4453124
## sample estimates:
## cor
## -0.8830677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

We can see that these variables are both strongly correlated (-.883) as well as statistically significant ($p=0.0003$).

Coefficient of determination

```
## [1] 0.7798085
```

Based off the R-squared value (Coefficient of determination) above and the significance of the Correlation coefficient, we can infer that the two variables are related. The relationship being, as TV time increases Reading Time decreases. Lets visualize it in the scatter plot below:



vi. Does watching more TV caused students to read less?

We can not state that watching more TV caused students to read less. We can only state the relationship between the two variables.

vii. Partial correlation

Lets look into this a bit further and use only three of the variables for a Partial correlation. We will be looking at TimeTV and Happiness controlling for Gender using the `pcor.test` function. In doing so we get the results shown below.

```
##      estimate    p.value statistic    n gp Method
## 1 0.6435158 0.04469059  2.377919 11  1 pearson
```

Based on the above we can determine that Gender has an effect on the correlation between TV Time and Happiness. We can assume this due to the fact that TimeTV and Happiness had a slightly smaller correlation within the overall matrix and the Partial correlation has a p value less than 0.05.

References

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

Lander, J. P. 2014. R for Everyone: Advanced Analytics and Graphics. Addison-Wesley Data and Analytics Series. Addison-Wesley. <https://books.google.com/books?id=3eBVAgAAQBAJ>.