

Week 7 Assignment Part 2

Tabbalat, Abed

2021-04-30

Part I

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.

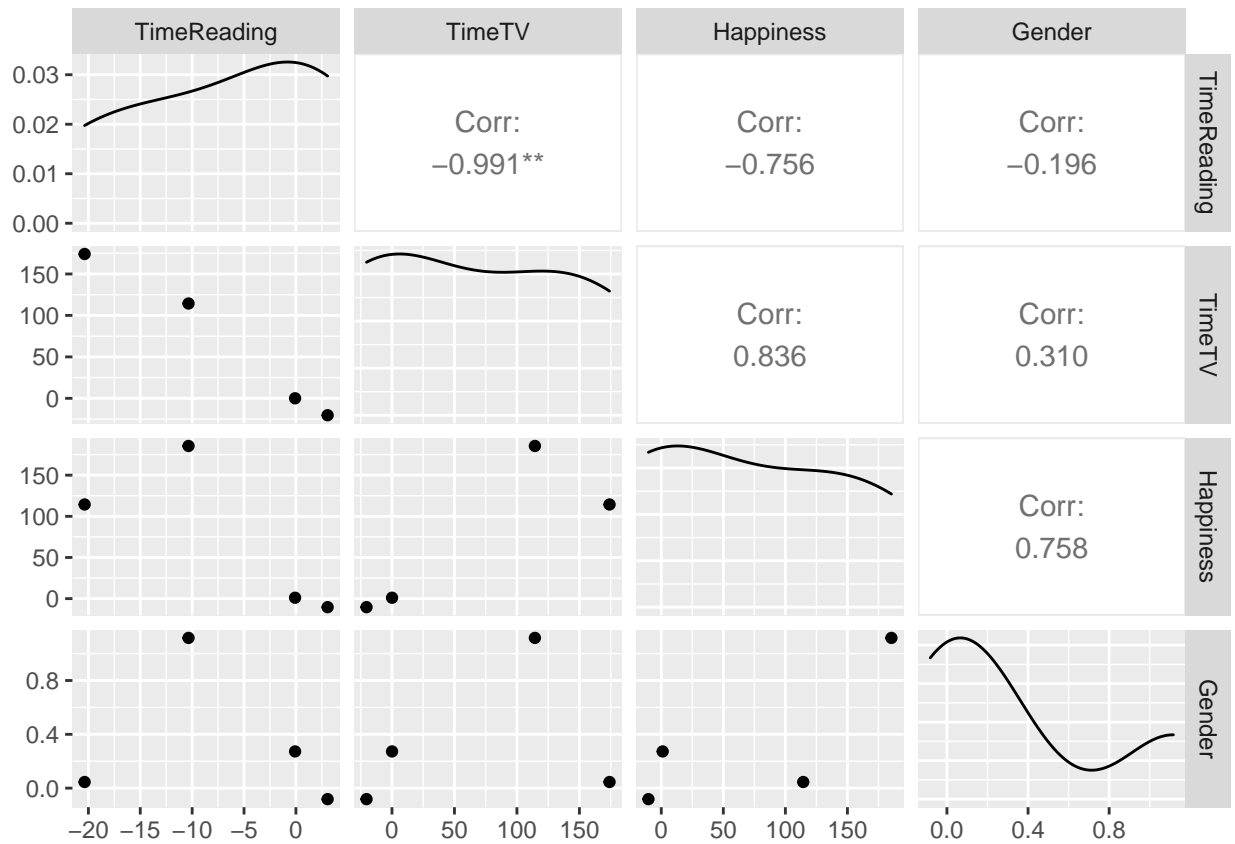


Table 1: Pearson Covariance For Student Dataset

| | TimeReading | TimeTV | Happiness | Gender |
|-------------|-------------|-------------|------------|------------|
| TimeReading | 3.0545455 | -20.3636364 | -10.350091 | -0.0818182 |
| TimeTV | -20.3636364 | 174.0909091 | 114.377273 | 0.0454545 |
| Happiness | -10.3500909 | 114.3772727 | 185.451422 | 1.1166364 |
| Gender | -0.0818182 | 0.0454545 | 1.116636 | 0.2727273 |

Using the pearson method would be the answer here as the data seem to be at a normal scale for each variable, the results indicate the following:

1.**TimeReading:** TimeTV -> negative covariance, Happiness -> negative covariance, Gender -> negative covariance (minimal, close to zero)

2.**TimeTV:** TimeReading -> negative covariance, Happiness -> positive covariance, Gender -> positive covariance (minimal, close to zero)

3.**Happiness:** TimeReading -> negative covariance, TimeTV -> positive covariance, Gender -> positive covariance

4.**Gender:** TimeReading -> negative covariance (minimal, close to zero), TimeTV -> positive covariance (minimal, close to zero), Happiness -> positive covariance

Part II

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.

Time is the measurement used for reading and watching TV, Happiness seems to be a percentage metric or a range from 0-100, gender is a boolean of 1 and 0 (doesn't determine which is which). If we add an additional variable that determines the type of books and TV shows, this will impact the covariance calculation, as TV shows could be done for leisure while books could be class materials instead of entertainment books.

Part III

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?

Using pearson method will determine if we have positive or negative correlation between the variables, I believe reading vs tv will have a negative correlation.

Part IV

Perform a correlation analysis of:

All variables

Table 2: Pearson Correlation For All Variables

| | TimeReading | TimeTV | Happiness | Gender |
|-------------|-------------|------------|------------|------------|
| TimeReading | 1.0000000 | -0.8830677 | -0.4348663 | -0.0896421 |
| TimeTV | -0.8830677 | 1.0000000 | 0.6365560 | 0.0065967 |
| Happiness | -0.4348663 | 0.6365560 | 1.0000000 | 0.1570118 |
| Gender | -0.0896421 | 0.0065967 | 0.1570118 | 1.0000000 |

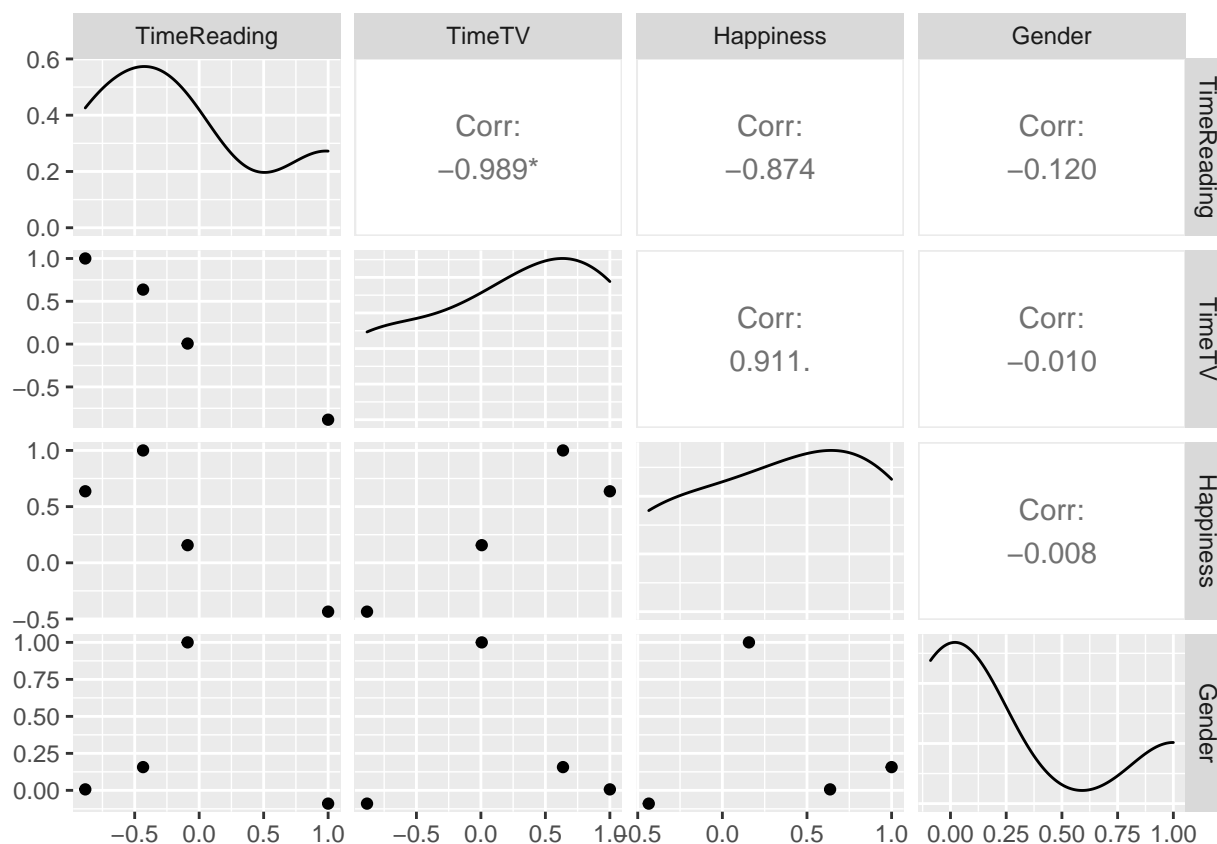
A single correlation between two a pair of the variables

Single correlation between Happiness and TimeTV is **0.636556**.

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

Single correlation between Happiness and TimeTV at 99% confidence intervall is $c(t = 2.47613128686031)$, $c(df = 9)$, 0.0352142508399742 , $c(cor = 0.636555986095781)$, $c(correlation = 0)$, two.sided, Pearson's product-moment correlation, $student_dfHappinessandstudent_dfTimeTV$, $c(-0.157021243016985, 0.930627539051343)$.

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



1. Time Reading: It is determined that TimeReading has a negative correlation with TimeTV and Happiness, Gender has a very small negative correlation that can be determined as no correlation. Meaning, the more students read the less happy they are, and watching less tv regardless of gender.

2. Time TV: It is determined that TimeTV has a negative correlation with reading, a positive correlation with happiness, and a very small negative correlation that can be determined as no correlation. Meaning, the more tv students watch the happier they are, and less time to read regardless of the gender.

3. Happiness: It is determined that the more time spent on reading their happiness drops, and the more time they spend watching tv the happier they get regardless of gender.

Part V

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

Correlation Coefficient:

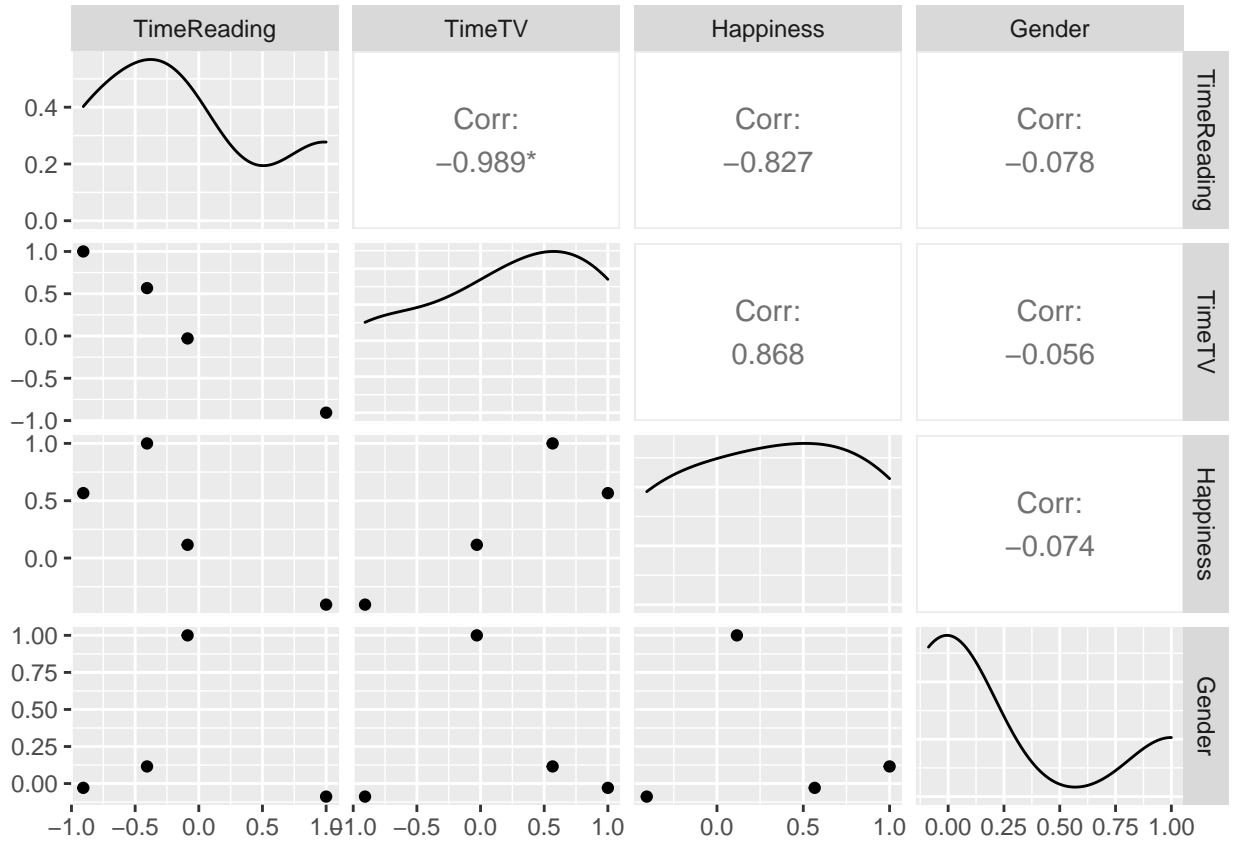


Table 3: Correlation Coefficient using Spearman

| | TimeReading | TimeTV | Happiness | Gender |
|-------------|-------------|------------|------------|------------|
| TimeReading | 1.0000000 | -0.9072536 | -0.4065196 | -0.0880141 |
| TimeTV | -0.9072536 | 1.0000000 | 0.5662159 | -0.0289996 |
| Happiness | -0.4065196 | 0.5662159 | 1.0000000 | 0.1154701 |
| Gender | -0.0880141 | -0.0289996 | 0.1154701 | 1.0000000 |

Applying the coefficient results in showing the correlation without a confidence interval, the highest result shows TimeReading vs TimeTV at -0.9.

Correlation of Determination:

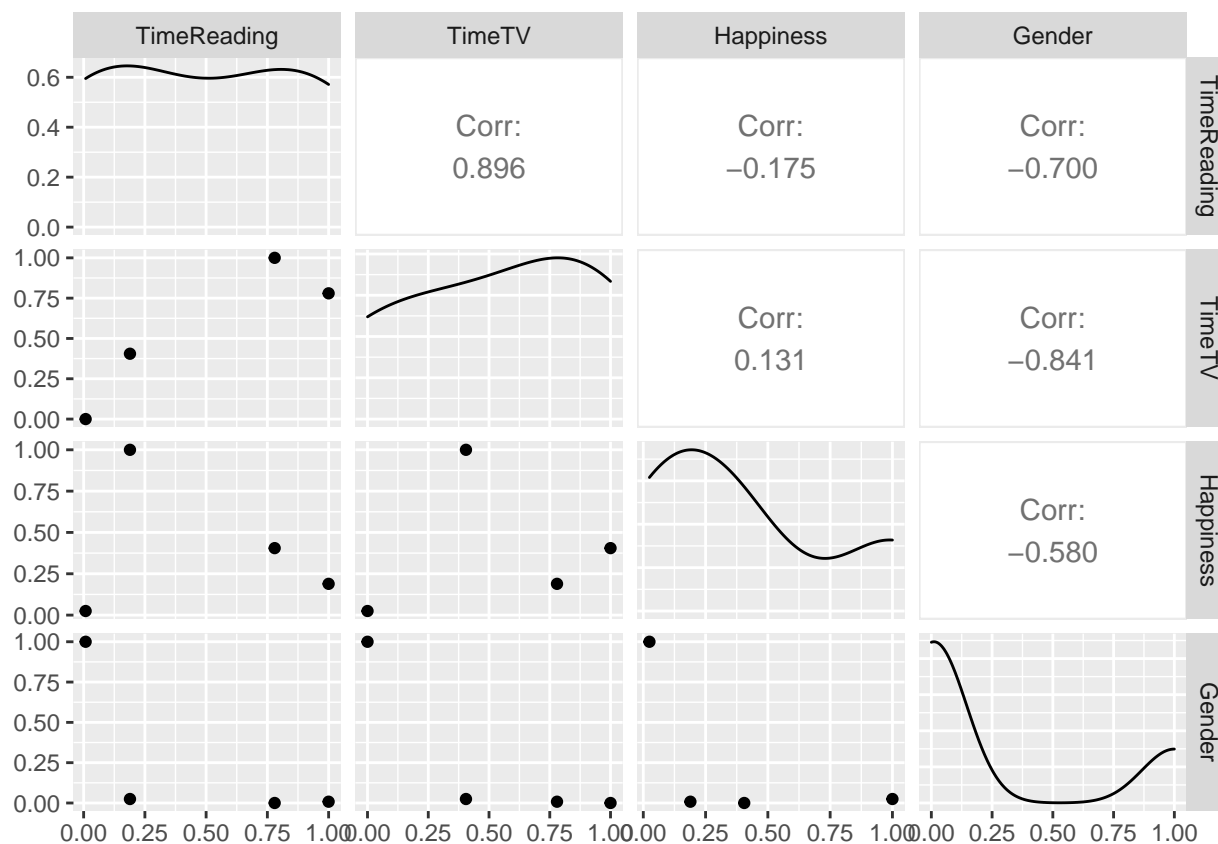


Table 4: Correlation of Determination

| | TimeReading | TimeTV | Happiness | Gender |
|-------------|-------------|-----------|-----------|-----------|
| TimeReading | 1.0000000 | 0.7798085 | 0.1891087 | 0.0080357 |
| TimeTV | 0.7798085 | 1.0000000 | 0.4052035 | 0.0000435 |
| Happiness | 0.1891087 | 0.4052035 | 1.0000000 | 0.0246527 |
| Gender | 0.0080357 | 0.0000435 | 0.0246527 | 1.0000000 |

Applying the correlation of determination results in determining the variability of the correlation, the highest is between TimeReading and TimeTV at 0.7789 (77.89% of students are impacted by the correlation).

Part VI

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

Yes, there is a clear negative correlation between watching tv and reading, watching more tv will result in less reading.

Part VII

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.

Partial Correlation between TimeReading and TimeTV holding Happiness Constant is **-0.872945**.

In the above analysis on watching tv vs reading correlation, happiness is a factor as it determines students are doing what makes them happy. Placing happiness as the controlled variable will remain a negative correlation between watching tv and reading, as there are so many hours during the day and it is working as a substitute.