

Student Survey Analysis

Reppeto, Brian

July 18, 2023

Student Survey Analysis

Questions

1.

Calculating the covariance matrix

##	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

The covariance matrix provides valuable insights into the relationships between variables. It measures how the variables change together. Based on these results the covariance for cov_time_read_gender is -0.08181818. This means there is a slightly negative relationship between the time spent reading and gender. The covariance for cov_time_tv_gender is 0.04545455. This indicates a positive relationship between the time spent watching TV and gender. The covariance for cov_happiness_gender is 1.116636, which indicates a relatively strong positive relationship between happiness and gender. All and all, A positive covariance indicates a positive relationship, and a negative covariance indicates a negative relationship.

2.

The Student Survey dataset consists of four variables: TimeReading: The amount of time spent reading (in minutes). TimeTV: The amount of time spent watching television (in minutes). Happiness: Subjective happiness rating. Gender: Gender of the student (0 for female, 1 for male).

Changing the measurement scale of variables can affect the covariance calculation. For example, if we change the measurement of TimeReading to hours instead of minutes, the covariance values would increase by a factor of 60 since the unit of measurement is larger. The same could be said about changing the measurement units of TimeTV, as it is measured in minutes as well.

A better alternative to address this issue is to standardize the variables before calculating the covariance or correlation. Standardizing the variables scales to have a zero mean and unit variance, would allowing for more meaningful comparison.

3.

A Pearson correlation test might be a good alternative with this dataset. Since all the variables are continuous (TimeReading, TimeTV, Happiness), and we are interested in measuring the linear relationship between them. Pearson correlation is appropriate for this scenario. Based on the Pearson correlation below, there

appears to be a positive correlation between TimeTV and Happiness, and a negative correlation between TimeReading and TimeTV.

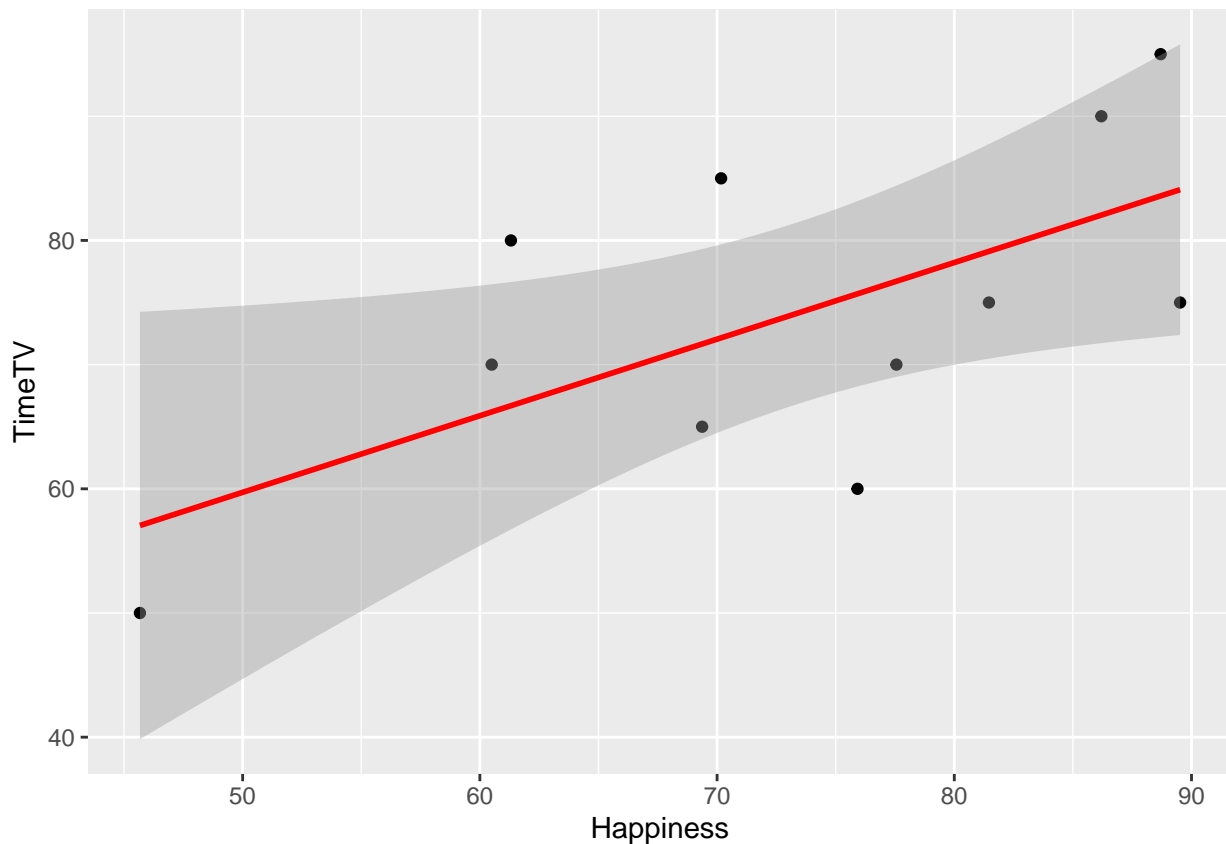
4.

Correlation analysis of all variables

```
##           TimeReading    TimeTV  Happiness
## TimeReading  1.0000000 -0.8830677 -0.4348663
## TimeTV      -0.8830677  1.0000000  0.6365560
## Happiness   -0.4348663  0.6365560  1.0000000
```

Correlation point graph on Happiness and TimeTv

```
## `geom_smooth()` using formula = 'y ~ x'
```



Single correlation between TimeTv and Happiness

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

Confidence correlation with interval at 99%

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

The correlation matrix suggests that there are significant relationships between the variables. Spending more time reading appears to be associated with less time spent watching TV and a lower happiness level. Additionally, spending more time watching TV is linked to less time spent reading and a slightly higher happiness levels.

5.

Pearson Correlation coefficient of TimeTV and Happiness

Coefficient of determination of TimeTV and Happiness

```
## [1] 0.636556
## [1] 0.4052035
```

The results suggest that there is a positive relationship between time and happiness, and time can explain about 40.52% of the variability in happiness levels.

6.

Based on the analysis, there is a correlations between TimeReading, TimeTV, and Happiness. However, remember that correlation does not imply causation. While I cannot conclude that watching more TV caused students to read less based on the analysis. Further more controlled experiments would be necessary to establish causation.

7.

Partial correlation between TimeReading and Happiness controlling for Gender

```
## [1] -0.02402155
```

The partial correlation between time reading and happiness, controlling for gender, indicates a weak negative association. Controlling for the gender variable allows us to understand the unique relationship between the other two variables and better understand their connection.