

RMarkdown Assignment Week7 Student survey

Koppula Veera

March 23rd 2018

As a data science intern with newly learned knowledge in skills in statistical correlation and R programming, you will analyze the results of a survey recently given to college students. You learn that the research question being investigated is: “Is there a significant relationship between the amount of time spent reading and the time spent watching television?” You are also interested if there are other significant relationships that can be discovered? The survey data is located in this StudentSurvey.csv file.

1. 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.

```
## [1] -20.36364
```

```
## [1] -10.35009
```

```
## [1] -0.08181818
```

```
## [1] 114.3773
```

```
## [1] 0.04545455
```

```
## [1] 1.116636
```

I am using these variable pairs, without considering Gender as it is binary a binary Variable. The results indicate that Time Reading and Time Tv are inversely correlated. This makes sense since if you're reading it's harder to spend time on watching tv. Time reading and Happiness also appear inversely correlated. This one is interesting and a bit surprising. Maybe if watching tv and reading are not balanced, makes one less happier. Lastly, Time Tv and Happiness are strongly correlated. This makes sense as maybe people are more happier watching tv that makes people release immediate happy hormones.

2. 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 reading looks to be in hours, Time Tv looks to be in minutes, Happiness looks like on scale 0-100, and gender is binary, 0 or 1. The first 3 are numerical values. Time reading is small compared to Time Tv, exaggerating the differences. Same applies to Time Reading and Happiness. Time Tv and Happiness are similar in numbers. All the variables are hard to compare with gender since it is binary. However, scaling Time TV would not result in a big difference in covariance direction, it just scales it by a factor of 60 (Min to hours). It shouldn't cause any problem, and I would just omit computing Covariance between Gender and the others since its not very relevant to the research in question. We could remove Gender from the whole dataset for evaluation.

3. 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?

```
## [1] -20.36364
```

The research question i am looking at is -“Is there a significant relationship between the amount of time spent reading and the time spent watching television?”I chose this one because data shows a direct correlation between Time Reading and Time Tv. The negative correlation suggests as one variable goes up the other varibale goes down. My prediction is if someone spends more time reading their time watching tv goes down (expected as there is limited time in a day).

##4. Perform a correlation analysis of: #a. 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
```

#b. A single correlation between two a pair of the variables

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

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

```
##
## 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
```

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

Time reading and time tv have a negative correlation, suggesting when one goes up the other tends to goes down. Time reading and Happiness have negative correlation, suggesting again when one variable goes up the other variable goes down. The relationship is opposite for happiness and time tv as well. Gender is harder to examine since it's a binary variable and could be difficult to asceratain correltion with other variables.

##5. Calculate the correlation coefficient and the coefficient of determenation, describe what you conclude about the results.

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

```
##           TimeReading      TimeTV  Happiness
## TimeReading  1.0000000  0.7798085  0.1891087
## TimeTV      0.7798085  1.0000000  0.4052035
## Happiness   0.1891087  0.4052035  1.0000000
```

I conclude that Time Reading and Time TV have strong correlation and Time TV and Happiness have moderate correlation.

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

Based on above point 5, about 78% of the variability of Time Reading can be explained by correlation with Time TV. The other 22% is explained by other variability. I would conclude there is strong evidence that more time reading leads to less time watching tv.

##7. 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.

[1] 0.762033

I am controlling “Happiness” as a variable here. This doesn’t really change my interpretation by much. It shows that TimeTv still accounts TimeReading about the same as before. Results appear in the same direction with “Happiness” as a controlled variable.