

---

## R Markdown

Title: Exercise 9 Student Survey

Author: Carla Bradley

Date: 01/05/2021

## Analysis

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. Is there a significant relationship between the amount of time spent reading and the time spent watching television?

```
## Warning: package 'ggplot2' was built under R version 4.0.5
```

```
##   TimeReading TimeTV Happiness Gender
## 1           1     90      86.20      1
## 2           2     95      88.70      0
## 3           2     85      70.17      0
## 4           2     80      61.31      1
## 5           3     75      89.52      1
## 6           4     70      60.50      1
```

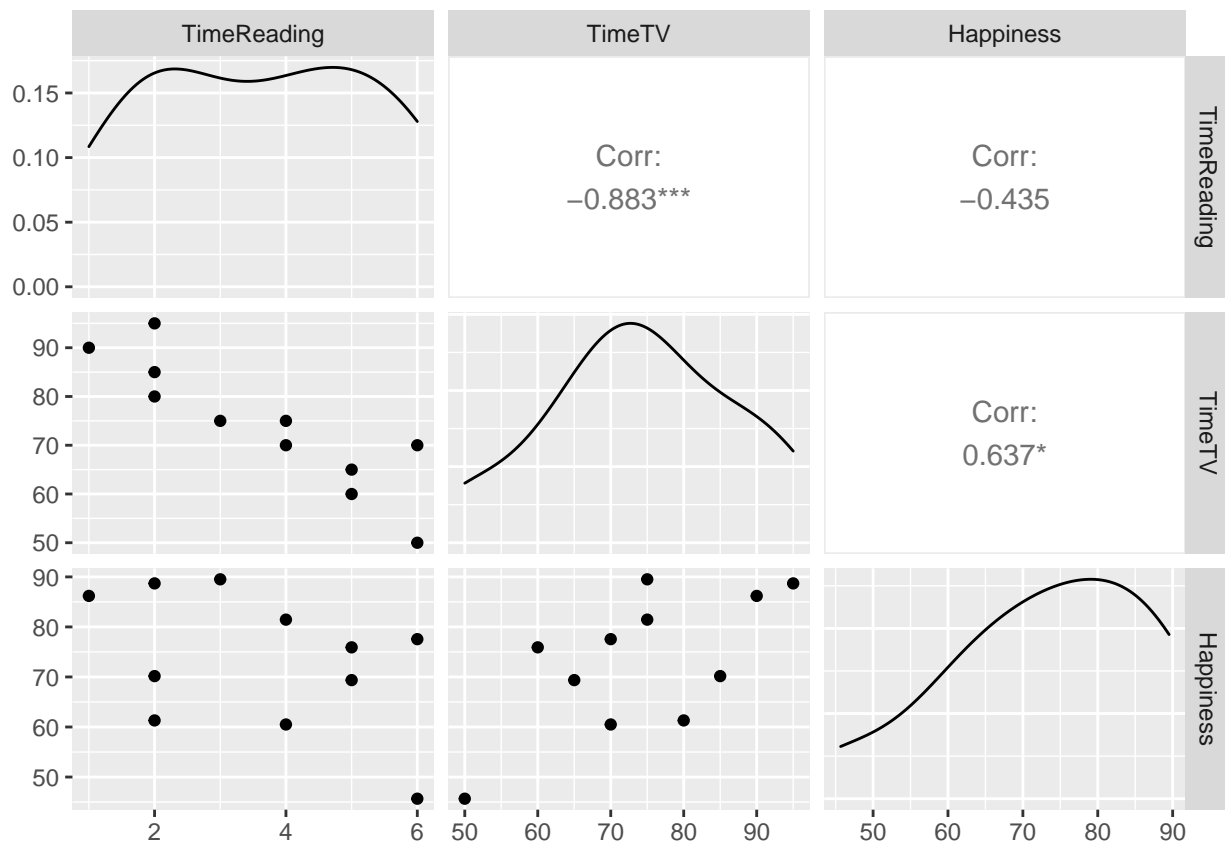
```
## [1] 0.636556
```

```
##   TimeReading TimeTV Happiness Gender
## 1           1     90      86.20      1
## 2           2     95      88.70      0
## 3           2     85      70.17      0
## 4           2     80      61.31      1
## 5           3     75      89.52      1
## 6           4     70      60.50      1
```

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

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

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2
```



Answer: There is a significant inverse relationship between TimeTV and TimeReading variables.

## 2. Is there any other significant relationships that can be discovered?

Answer: Another significant relationship on this calculation denotes a close relationship between the TimeTV and Happiness variables.

### A. 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?

Answer: The covariance between variables "TimeTV" and TimeReading resulted in a covariance of -20.36364 Denoting a negative relationship between the two variables. I have to denote that the order in cov command doesn't matter cov(student\_srv\$ TimeReading,student\_srvTimeTV)and cov(student\_s rvTimeTV,student\_srv\$TimeReading) both these will give the same result.

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

Answer: I calculated the covariance TimeTV and Happiness resulting in a covariance of 114.3773 Denoting a positive relationship between the two variables. I have to denote that the order in cov command doesn't matter cov(student\_srv\$ Happiness,student\_srvTimeTV) and cov(student\_s rvTimeTV,student\_s rvHappiness) both these will give the same result. I choose this calculation because it allows to examine the relationship between two variables.

```
## Warning in data("student_srv"): data set 'student_srv' not found
```

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

### B. Examine the Survey data variables. What measurement is being used for the variables?

Answer: The type of data measurement used is quantitative and qualitative Categorical variables.

**C. 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.**

Answer: The type of data measurement use as variables are Nominal Scale and Ratio scale.

Both of the time variables are continuous variables since time is a continuous datatype. Covariance calculation works best with continuous variables and therefore changing measurements would not correct results.

A change in the units of measurement can change the covariance Thus, covariance is only useful to find the direction of the relationship between two variables and not the magnitude.

I dont believe there is a problem if units of measurement are change as long the TimeTV and TimeReading variables has the same type of unit measurement since this will determine the type of techniques to be used for statistical analysis.

```
## Gender TimeTV
## 1 0 305.0000
## 2 1 104.1667
```

##Perform a correlation analysis of

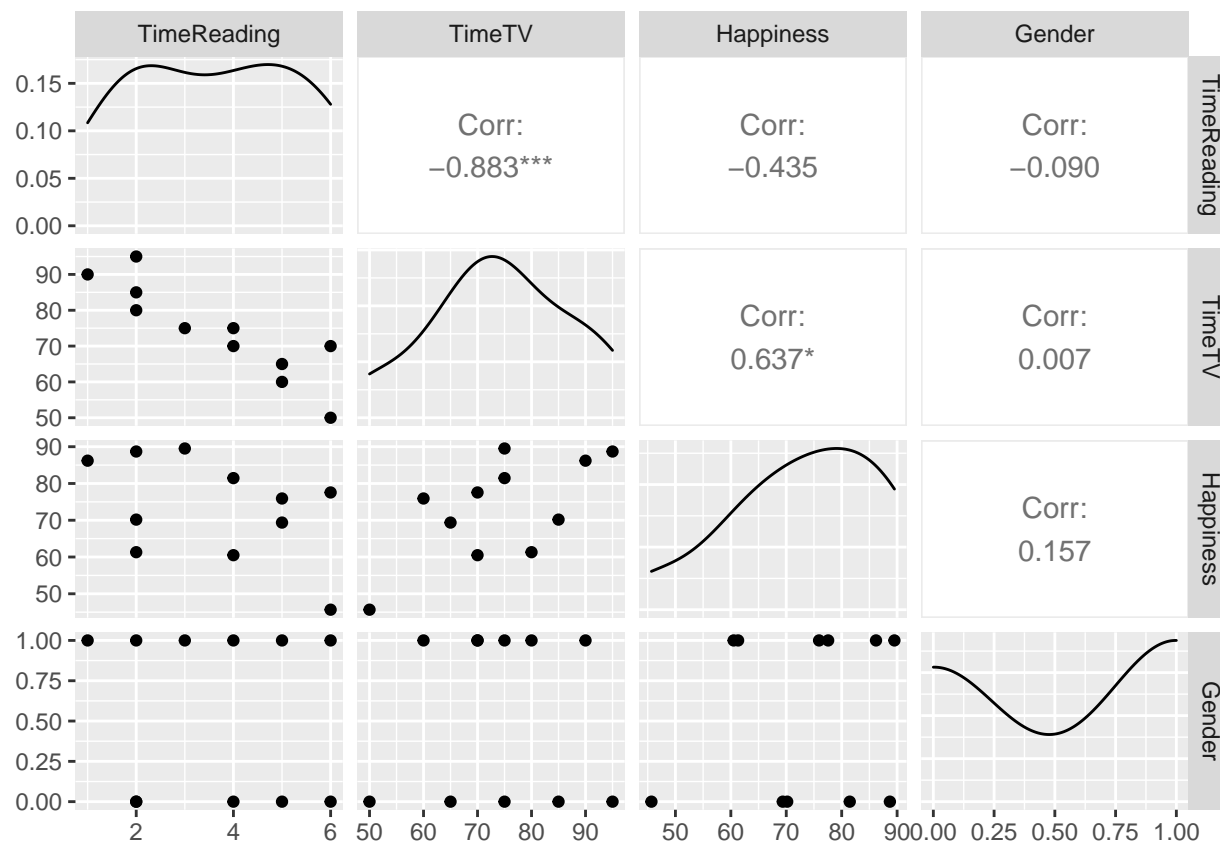
1. All variables

```
## 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
```

2. A single correlation between two a pair of the variables

```
## Warning in data(student_srv, package = "reshape2"): data set 'student_srv' not
## found
```

```
## TimeReading TimeTV Happiness Gender
## 1 1 90 86.20 1
## 2 2 95 88.70 0
## 3 2 85 70.17 0
## 4 2 80 61.31 1
## 5 3 75 89.52 1
## 6 4 70 60.50 1
```



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

Answer According to this test, I can conclude that watching tv and reading has a remarkable difference. While all this statistical simple rule of thumb would be to see if the two means are within two standard deviations of each other and the mean according with this sample display a distant standard deviation.

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

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

```
##
## Welch Two Sample t-test
##
```

```
## data: student_srv$TimeTV and student_srv$Happiness
## t = 0.13707, df = 19.98, p-value = 0.8923
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -15.48525 17.05252
## sample estimates:
## mean of x mean of y
## 74.09091 73.30727

##
## Welch Two Sample t-test
##
## data: student_srv$TimeReading and student_srv$Happiness
## t = -16.83, df = 10.329, p-value = 7.643e-09
## alternative hypothesis: true difference in means is not equal to 0
## 99 percent confidence interval:
## -82.69751 -56.64431
## sample estimates:
## mean of x mean of y
## 3.636364 73.307273
```

Answer Calculations in the correlation matrix suggest that both positive and negative correlations between variables exist. Time spent on TV has a moderate positive correlation to happiness while time spent reading has a low negative correlation to happiness.

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

#Correlation Coefficient

Correlation Coefficient is -0.883% causing no effect -0.883% of the total variance.

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

## Coefficient of Determination,

This correlation, is represented as a value of 0.7798085 indicating a very close fit to the model and suggests that 78% of the dependent variable is predicted by the independent variable.

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

Answer: The coefficient of determination of the simple linear regression model for the data set is 0.7798085 indicating a very close fit to the model and suggests that 78% of the dependent variable is predicted by the independent variable.

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

Based on my analysis student are more inclined to watch tv than reading.I also denote a noticeable high level of happiness by watching tv.

g. 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.

```
## Loading required package: MASS
```

```
##      estimate      p.value statistic  n gp Method
## 1 -0.872945 0.0009753126 -5.061434 11  1 pearson
```

```
““
```

```
#The controlling variable in this case is “Happiness” from the result, for students who are happy, watching TV has strong negative correlation to reading.
```