

Week7_Assignment

Akila Selvaraj

1/26/2022

Research question being investigated:

We have StudentSurvey dataset which has data about time spent by students in reading, time spent for TV, Happiness and Gender. In this project, We are going to investigate if there is a significant relationship between the amount of time spent reading and the time spent watching television.

i) Covariance of the Survey variables and Indication of Results

Covariance of TimeReading and TimeTV is -20.3636364

Covariance of all the Survey variables

```
##           TimeReading    TimeTV Happiness
## TimeReading    3.054545 -20.36364 -10.35009
## TimeTV        -20.363636 174.09091 114.37727
## Happiness     -10.350091 114.37727 185.45142
```

Covariance indicates the relationship between two variables whenever one variable changes. Covariance gives us a positive number if the variables are positively related. We would get a negative number if they are negatively related. A high covariance basically indicates there is a strong relationship between the variables. A low value means there is a weak relationship.

Covariance results of Survey variables says that there is a negative correlation between TimeReading and TimeTV which indicates that these variables are negatively related. Covariance between TimeReading and Happiness is negative whereas covariance between TimeTV and Happiness is positive which indicates watching TV increasing happiness than reading.

ii) Measurement being used for the variables

By looking at the data, TimeReading was measured in hours whereas TimeTV was measured in minutes, and Happiness was calculated in percentage.

Effect on covariance calculation by changing the measurement

Time variables in the data, TimeReading and TimeTV were measured differently. TimeReading was measured in hours gives less covariance. TimeTV measured in minutes yields greater covariance value. Using same measurement for both the variables change the result of covariance. Covariance result will be high for TimeReading if we convert the values into minutes.

Better alternative

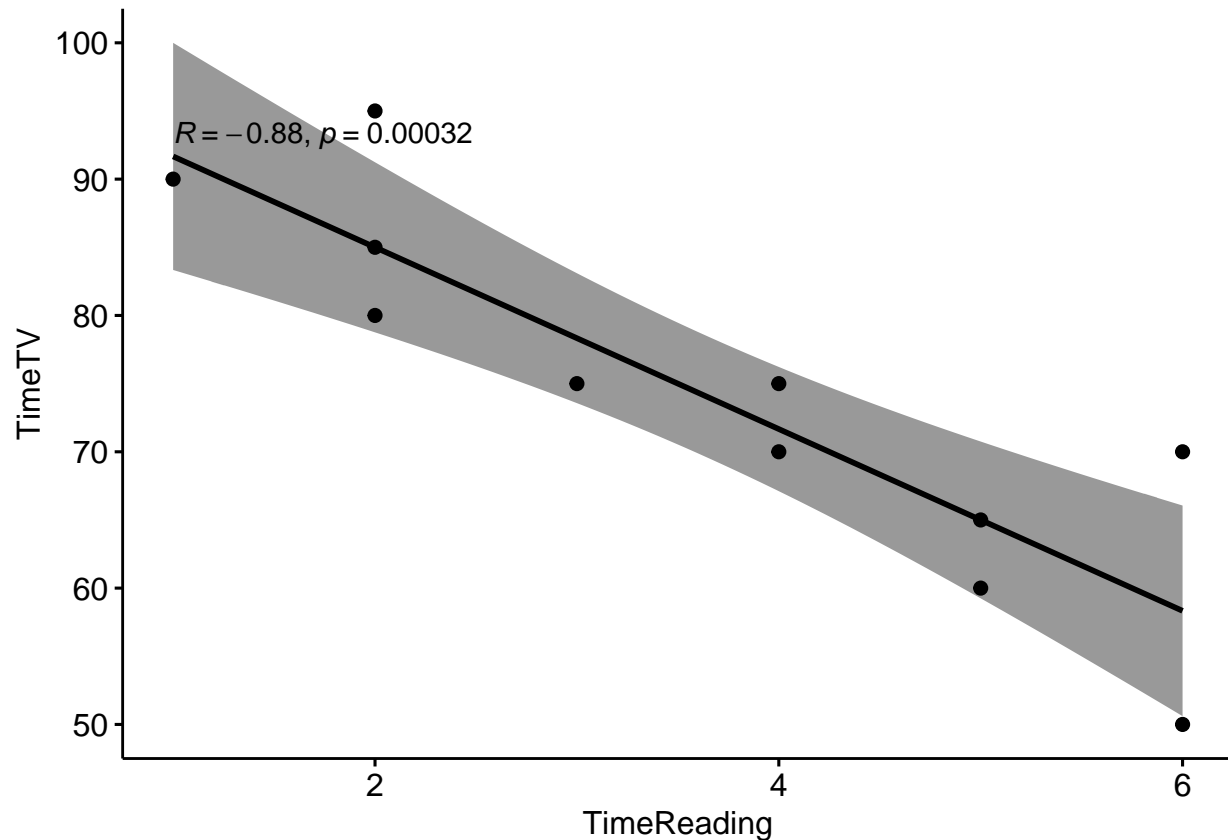
Using different measurements might be a problem especially when we compare these two variables with the third variable which is Happiness in this study. Covariance varies high between TimeReading and TimeTV

because of this different measurements being used. As an alternative, we can convert the measurements of one of the variables to be in sync with other variable for better analysis of results.

iii) Correlation test to perform

To determine the type of correlation to perform, let's visualize the data using scatter plot.

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



From the plot, p value is 0.00032 which is less than 5% shows that correlation between TimeReading and TimeTV is significant. As we can say that both populations may come from normal distributions, it's recommended to use parametric correlation method Pearson. From the plot, we can predict that correlation test of TimeTV and TimeReading would yield negative correlation.

Correlation Test of TimeReading and TimeTV

```
##
## Pearson's product-moment correlation
##
## data: Survey_df$TimeTV and Survey_df$TimeReading
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9694145 -0.6021920
## sample estimates:
##      cor
## -0.8830677
```

iv) Correlation analysis

1. Correlation of 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. Single correlation between two pair of the variables.

Correlation test of TimeReading and TimeTV is

```
##
## Pearson's product-moment correlation
##
## data: Survey_df$TimeReading and Survey_df$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.9694145 -0.6021920
## sample estimates:
##           cor
## -0.8830677
```

Correlation test of TimeReading and Happiness is

```
##
## Pearson's product-moment correlation
##
## data: Survey_df$TimeReading and Survey_df$Happiness
## t = -1.4488, df = 9, p-value = 0.1813
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.8206596  0.2232458
## sample estimates:
##           cor
## -0.4348663
```

Correlation test of TimeTV and Happiness is

```
##
## Pearson's product-moment correlation
##
## data: Survey_df$TimeTV and Survey_df$Happiness
## t = 2.4761, df = 9, p-value = 0.03521
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.05934031 0.89476238
## sample estimates:
##           cor
##  0.636556
```

3.Repeating correlation test in step 2 by setting the confidence interval at 99%

Correlation test of TimeReading and TimeTV is

```
##
## Pearson's product-moment correlation
##
## data: Survey_df$TimeReading and Survey_df$TimeTV
## 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.Relationship between the variables as per the correlation matrix

The correlation coefficient between the two vectors TimeReading and TimeTV turns out to be -0.8830677. The test statistic turns out to be -5.6457 and the corresponding p-value is 0.0003153. Since this value is less than .05, we have sufficient evidence to say that the correlation between the two variables is statistically significant.

v) Calculation of correlation coefficient and the coefficient of determination

```
##
## Call:
## lm(formula = Survey_df$TimeReading ~ Survey_df$TimeTV, data = Survey_df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.9452 -0.4922 -0.2846  0.3851  1.8851
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   12.30287    1.55704   7.901 2.44e-05 ***
## Survey_df$TimeTV -0.11697    0.02072  -5.646 0.000315 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8645 on 9 degrees of freedom
## Multiple R-squared:  0.7798, Adjusted R-squared:  0.7553
## F-statistic: 31.87 on 1 and 9 DF, p-value: 0.0003153
```

Conclusion from the results of correlation coefficient and coefficient of determination

The correlation coefficient of TimeTV and Happiness is 0.63655. Since it is little close to 1, we can say that the variables are positively linearly related. The correlation coefficient of TimeReading and Happiness is -0.4348663. Since it is negative, we can say that the variables are negatively linearly related.

Co-efficient of determination(R-squared) of the model is 0.7798. This means that 77.98% of the variation in the TimeReading can be explained by the number of hours spent in watching TV.

vi) Based on the analysis, we can say that watching more TV makes the students to read less. Explain?

Yes, based on the analysis we did, we can say that watching more TV caused students to read less. Negative correlation coefficient of TimeReading and TimeTV indicates that more the time spent in TV , less the time students spend in reading. From the linear regression model, r-squared of 77% reveals that 77% of the data fit the regression model which in turns reiterates that the students who spending more time in TV reading less.

vii) Partial Correlation

Partial correlation between TimeReading and Happiness by controlling TimeTV

```
## Loading required package: MASS

##      estimate  p.value statistic  n gp Method
## 1 0.3516355 0.319059  1.062425 11  1 pearson
```

Partial correlation between TimeTV and Happiness by controlling TimeReading

```
##      estimate    p.value statistic  n gp Method
## 1 0.5976513 0.06804372  2.108388 11  1 pearson
```

Partial correlation between TimeReading and TimeTV by controlling Happiness

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

Partial correlation of time reading and time TV is negative -0.87294. As the TimeTV increases, time spent on reading decreases. In other words, there is a significant negative correlation between the time students spend on TV and reading. Partial correlation of timeTV and Happiness is more than partial correlation of timereading and happiness which indicates spending more time in TV yields more happiness than spending more time in reading.