Student Survey Analysis

Julie Campbell

October 15, 2022

Background

This report details the research to investigate the question of "Is there a significant relationship between the amount of time spent reading and the time spent watching television?". Additional analysis was also done to determine if additional relationships were involved.

Survey Data Covariance

Covariance is the measurement of the relationship between two variables to determine if the change in one variable is equal to the other. This allows us to evaluate the positive or negative value to describe the relationship. The covariance of the survey results are presented below.

```
##              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 relationship between the sets of variables below shows a positive covariance in that they deviate in the same direction.

- TimeTV and Happiness
- TimeTV and Gender
- Happiness and Gender

The relationship between the sets of variables below shows a negative covariance in that they deviate in different directions.

- TimeTV and Time Reading
- TimeReading and Happiness
- TimeReading and Gender

Data Measurement

The survey data variables hold numeric values to represent their value. TimeReading and TimeTV are measured in the number of hours of completing the activity. Happiness is measured on a 1-100 scale to represent the emotion. Gender allows a 0 or 1 value for female or male participants. Covariance is not a standardized measure and cannot be compared in an objective way. This requires both data sets to be measured in the same units. In order to express covariance in a standard unit of measurement, you can divide the correlation coefficient by the standard deviation.

Survey Data Correlation

For this data set, the Pearson's correlation coefficient test was used since the time watching tv and the time reading distributions are approximately normal. Without additional data, I would suspect a negative correlation between time watching tv and time reading since leisure time is limited on an average day.

The correlation analysis 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
```

The correlation between time spent watching TV and reading:

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

The correlation between time spent watching TV and reading with the confidence level at 99%:

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

Using the calculations from the correlations above, a negative correlation is found between the time watching tv and time spent reading variables. Comparing the 95% and 99% confidence level data, the ranges do not cross 0 which validates the likelihood that it has a strong negative relationship. It also show a p value < .001.

Survey Data Coefficient

The following is the survey data correlation coefficients:

```
##              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
```

The following is the survey data coefficient of determination:

```
##              TimeReading        TimeTV  Happiness        Gender
## TimeReading 1.000000000 0.7798085292 0.18910873 0.0080357143
## TimeTV      0.779808529 1.0000000000 0.40520352 0.0000435161
## Happiness   0.189108726 0.4052035234 1.00000000 0.0246527174
## Gender      0.008035714 0.0000435161 0.02465272 1.0000000000
```

Based on the data above, it can be concluded that 18.9% of time reading and 40.5% of time watching tv contributed to the value of happiness score recorded by the survey. Gender appears to have a very small contribution the happiness score. The percentages found does not consider any additional external factors.

Data Analysis

The data analysis above shows a negative correlation between participants reading and watching TV. It can be concluded that watching tv could be a factor in time spent reading.

Partial Correlation

For the partial correlation, I took into consideration the relationship between watching tv and reading with the control happiness. Below is the results:

```
## [1] -0.872945
```

While using happiness as a control, the time spent watching tv and the time spent reading still have a strong negative correlation. This demonstrates that level of happiness is mostly independent of how time is utilized.