

StudentSurvey_Week7_TatreauGillian

Gillian Tatreau

2022-10-13

i.

```
# covariance of time spent reading and time spent watching TV  
cov(x = survey$TimeReading, y = survey$TimeTV)
```

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

You would use this calculation to determine the direction of the relationship between the two variables- this value of -20.4 indicates that there is an inverse relationship between time spent reading and time spent watching TV. This means that, on average, when time spent reading is low, time spent watching TV is high and vice versa.

ii.

It appears that the measurement for time spent reading is hours and for time spent watching TV it is minutes. This could affect the covariance because the measurement for time spent watching TV is multitudes higher than that for time spent reading, exaggerating the result seen. This is not necessarily an issue if we are looking at only these two variables, but if we want to compare covariances between different pairings of variables within the dataset, there is no way to compare them unless the units of measurement are standardized. For example, if we were to change the measurement for time spent watching TV to be in hours, we get a value for the covariance that, while still negative, is much closer to zero at -0.34.

```
## [1] -0.3393939
```

iii.

The correlation test to be performed will be the the Pearson's correlation coefficient test. This is an appropriate test to use because Pearson's coefficient (as a significant value) assumes the sampling distributions of both variables are normal, which both variables (time spent watching TV and time spent reading) are. I predict the test will reveal a negative correlation between time spent reading and time spent watching TV, based on the negative value for the covariance.

```
##           TimeReading  
## nbr.val      11.0000  
## nbr.null      0.0000
```

```
## nbr.na          0.0000
## min             1.0000
## max             6.0000
## range           5.0000
## sum             40.0000
## median          4.0000
## mean            3.6364
## SE.mean         0.5270
## CI.mean.0.95    1.1741
## var             3.0545
## std.dev         1.7477
## coef.var        0.4806
## skewness        -0.0025
## skew.2SE        -0.0019
## kurtosis        -1.6422
## kurt.2SE        -0.6418
## normtest.W      0.9209
## normtest.p      0.3265
```

```
##               TimeTV_min
## nbr.val        11.000
## nbr.null        0.000
## nbr.na          0.000
## min            0.833
## max            1.583
## range           0.750
## sum            13.583
## median          1.250
## mean            1.235
## SE.mean         0.066
## CI.mean.0.95    0.148
## var             0.048
## std.dev         0.220
## coef.var        0.178
## skewness        -0.118
## skew.2SE        -0.090
## kurtosis        -1.038
## kurt.2SE        -0.406
## normtest.W      0.987
## normtest.p      0.992
```

iv.

1.

```
round(cor(survey), digits = 2)
```

```
##               TimeReading TimeTV Happiness Gender TimeTV_min
## TimeReading      1.00  -0.88      -0.43  -0.09      -0.88
## TimeTV           -0.88   1.00       0.64   0.01       1.00
```

```
## Happiness      -0.43   0.64   1.00   0.16   0.64
## Gender         -0.09   0.01   0.16   1.00   0.01
## TimeTV_min     -0.88   1.00   0.64   0.01   1.00
```

2.

```
# correlation for two variables (time watching TV and happiness)
cor(x = survey$TimeTV, y = survey$Happiness, method = "pearson")
```

```
## [1] 0.64
```

3.

```
##
## Pearson's product-moment correlation
##
## data: survey$TimeTV and survey$Happiness
## t = 2, df = 9, p-value = 0.04
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.059 0.895
## sample estimates:
## cor
## 0.64
```

```
##
## Pearson's product-moment correlation
##
## data: survey$TimeTV and survey$Happiness
## t = 2, df = 9, p-value = 0.04
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
## -0.16 0.93
## sample estimates:
## cor
## 0.64
```

At the 99% percentile, we cannot reject the null hypothesis and thus conclude that the true correlation between these two variables is 0.

4.

From the correlation matrix, we can infer that there is a strong negative relationship between time spent watching TV and time spent reading. It is likely that time spent reading and happiness have a negative relationship, and time spent watching TV and happiness have a positive relationship. There is nearly no relationship between time spent reading and gender, and time spent watching TV and gender. There may exist a very weak positive relationship between gender and happiness.

V.

```
##
## Pearson's product-moment correlation
##
## data: survey$TimeReading and survey$TimeTV
## t = -6, df = 9, p-value = 3e-04
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
## -0.97 -0.60
## sample estimates:
## cor
## -0.88

## [1] 0.78
```

There is a strong negative relationship, as predicted above, between time spent reading and time spent watching TV. The coefficient of determination (0.78) means time spent reading accounts for 78% of the variability in time spent watching TV.

vi.

From this data, we cannot conclude causality. We cannot conclude that watching TV caused students to read less. We can conclude the variables are related, and within the context of this data, we can infer that when students spend more time reading, they spend less time watching TV, and when they spend more time watching TV, they spend less time reading.

vii.

```
## $estimate
##               survey.TimeReading survey.TimeTV survey.Happiness
## survey.TimeReading             1.00         -0.87             0.35
## survey.TimeTV                  -0.87             1.00             0.60
## survey.Happiness                0.35             0.60             1.00
##
## $p.value
##               survey.TimeReading survey.TimeTV survey.Happiness
## survey.TimeReading             0.00000         0.00098         0.319
## survey.TimeTV                  0.00098         0.00000         0.068
## survey.Happiness               0.31906         0.06804         0.000
##
## $statistic
##               survey.TimeReading survey.TimeTV survey.Happiness
## survey.TimeReading             0.0          -5.1             1.1
## survey.TimeTV                  -5.1           0.0             2.1
## survey.Happiness               1.1           2.1             0.0
##
## $n
## [1] 11
##
```

```
## $gp
## [1] 1
##
## $method
## [1] "pearson"
```

If we control happiness, the partial correlation between time spent reading and time spent watching TV is -0.87, with a p-value of 0.00098. Compared to the result from earlier (-0.88) this is only a very small difference. The largest difference is when we control time spent watching TV, to have a partial correlation of 0.35 between time spent reading and happiness (compared to -0.43 from earlier), but due to the p-value of 0.32, this is not a significant value in actuality.