

Student Survey

Myranda Swartzwelter

1/25/2022

Student Survey Covariance

Covariance is the averaged sum of combined deviations. A positive covariance indicates that as one variable deviates from the mean, the other variable also deviates from the mean in the same direction. A negative covariance indicates that as one variable deviates from the mean, the other variable deviates from the mean in the other direction. The magnitude of difference is dependent on the scale of the data being measured.

Covariance of Student Survey variables

	TimeReading	TimeTV	Happiness	Gender
TimeReading	3.0545455	-20.3636364	-10.350091	-0.0818182
TimeTV	-20.3636364	174.0909091	114.377273	0.0454545
Happiness	-10.3500909	114.3772727	185.451422	1.1166364
Gender	-0.0818182	0.0454545	1.116636	0.2727273

These results indicate that:

- TimeReading is inversely related to TimeTV, Happiness and Gender and vice versa.
- Time TV is positively related to Happiness and Gender and vice versa.
- Happiness is positively related to Gender and vice versa.

Examining Variables

Example of Student Survey Variables

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

Looking at the variables in the student survey table, we find that all of the variables are using different scales of measurement. For example, it appears that TimeReading is likely measured in hours whereas TimeTV is likely measured in minutes. Although if corrected for this by converting TimeReading to minutes or TimeTV to hours, it would not change the

sign of the covariance (so the relationship between variables remains positive or inverse), it would change the magnitude of the covariance between variables

Covariance of Student Survey variables

	TimeReading	TimeReading_min	TimeTV	Happiness	Gender
TimeReading	3.0545455	183.272727	-20.3636364	-10.350091	-0.0818182
TimeReading_min	183.2727273	10996.363636	-1221.8181818	-621.005455	-4.9090909
TimeTV	-20.3636364	-1221.818182	174.0909091	114.3772727	0.0454545
Happiness	-10.3500909	-621.005455	114.3772727	185.451422	1.1166364
Gender	-0.0818182	-4.909091	0.0454545	1.116636	0.2727273

Happiness is likely on a scale from 1-100 but we have no concept of a scale for this metric. It is unable to be changed into a minute or hour measure like the previous variables were.

Gender is an indicator function, indicating a student is on gender if they are labeled as '1' and the other gender if labeled as '0'.

If you swap the gender indicators (i.e. the 1's become 0's and the 0's become 1's), you also change the direction of the relationship between Gender and the other variables, but maintain the magnitude. See table for examples.

Covariance of Student Survey variables

	TimeReading	TimeTV	Happiness	Gender	Gender_Swap
TimeReading	3.0545455	-20.3636364	-10.350091	-0.0818182	0.0818182
TimeTV	-20.3636364	174.0909091	114.377273	0.0454545	-0.0454545
Happiness	-10.3500909	114.3772727	185.451422	1.1166364	-1.1166364
Gender	-0.0818182	0.0454545	1.116636	0.2727273	-0.2727273
Gender_Swap	0.0818182	-0.0454545	-1.116636	-0.2727273	0.2727273

TimeTV, TimeReading and Happiness are ratio variables, where as gender is binary.

Type of Correlation Analysis

First we want to look at the type of variables. For TimeTV, TimeReading and Happiness, the variables are ration and linear, so we can do a Pearson's R test. However for Gender, we

must do a biserial correlation. This test is the same in R as the Pearson's test so we'll use Pearson's for all variables. Because of the values we saw in the covariances, I believe:

- TimeReading will be negatively correlated with TimeTV, Happiness and Gender.
- Time TV will be positively correlated to Happiness and Gender.
- Happiness will be positively correlated to Gender.

Correlation Analysis

Correlations of all variables. For this analysis we'll change the variables of time reading to be in minutes to match tv time.

Pearson R correlation of Student Survey variables

	TimeReading	TimeTV	Happiness	Gender
TimeReading	1.0000000	-0.8830677	-0.4348663	-0.0896421
TimeTV	-0.8830677	1.0000000	0.6365560	0.0065967
Happiness	-0.4348663	0.6365560	1.0000000	0.1570118
Gender	-0.0896421	0.0065967	0.1570118	1.0000000

Correlation of two variables TimeTV and TimeReading:

```
##
##  Pearson's product-moment correlation
##
## data:  student_survey_df$TimeReading and student_survey_df$TimeTV
## t = -5.6457, df = 9, p-value = 0.0001577
## alternative hypothesis: true correlation is less than 0
## 95 percent confidence interval:
##  -1.0000000 -0.6684786
## sample estimates:
##          cor
## -0.8830677
```

Correlation of two variables TimeTV and TimeReading with 99% Confidence Interval:

```
##
##  Pearson's product-moment correlation
##
## data:  student_survey_df$TimeReading and student_survey_df$TimeTV
## t = -5.6457, df = 9, p-value = 0.0001577
## alternative hypothesis: true correlation is less than 0
## 99 percent confidence interval:
##  -1.0000000 -0.5131843
## sample estimates:
##          cor
## -0.8830677
```

Looking at the correlation matrix we can say:

- TimeTV and TimeReading are strongly negatively correlated (p value calculated < 0.05)
- TimeTV and Happiness are potentially positively correlated
- TimeTV and Gender are not correlated
- TimeReading and Happiness are potentially negatively correlated
- TimeReading and Gender are not correlated
- Happiness and Gender may be slightly positively correlated

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

Yes, when doing a correlation test, we see that the null hypothesis is rejected with a p value of 0.000157 so TimeReading and TimeTV are negatively correlated, or the more someone watches TV, the less time they spend reading.

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.

I am choosing to perform a partial correlation for TimeTV and TimeReading while controlling for Happiness.

```
## [1] -0.872945
## $tval
## [1] -5.061434
##
## $df
## [1] 8
##
## $pvalue
## [1] 0.0009753126
```

When controlling for happiness, the partial correlation for TVTime and TVReading goes from -0.88 to -0.87. This is still a strong correlation. The p value for the partial correlation is 0.000975 which is still significantly below 0.05. This means that even when controlling for happiness, there is a strong negative correlation between TimeTV and TimeReading.