

## Week 7

Mithil Patel

2022-04-25

### Covariance

	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

- i) The covariance is an important factor to determine whether two variables contain a positive or negative relationship with each other. From the matrix above, there exists a strong positive relationship between time watching TV and happiness. On the contrary, there is a weak negative relationship between time spent reading and time watching TV along with time spent reading and happiness. The covariance values for gender versus the three other variables (TimeReading, timeTV, and Happiness) are approximately zero indicating no relationship between them.

### Level of Measurements

TimeReading(hour): Ratio

TimeTV(minute): Ratio

Happiness(percentage): Ratio

Gender(label): Nominal

	TimeReading	NewTimeTV	Happiness	Gender
TimeReading	3.05454545	-0.3393939394	-10.350091	-0.0818181818
NewTimeTV	-0.33939394	0.0483585859	1.906288	0.0007575758
Happiness	-10.35009091	1.9062878788	185.451422	1.1166363636
Gender	-0.08181818	0.0007575758	1.116636	0.2727272727

- ii) To evaluate the effect changing variables have on covariance, I will change the 'TimeTV' unit from minutes to hours and compare the covariance matrix. Based on the new matrix, the covariance value (-0.340) suggests that there exists no relationship between 'NewTimeTV' and 'TimeReading', which is different from the earlier conclusion – that there is a weak negative relationship. In essence, changing measurement for 'TimeTv' has resulted in no relationship with the other three variables; therefore, speculating a relationship based on covariance value can lead to many problems. Compare to covariance, the correlation coefficient tends to produce more promising results since it is not affected by the change in measurement.

## Correlation test

- iii) I will be performing a paired samples correlation test because we want to analyze the significant relationship between the amount of time reading and the time spent watching television. Based on the covariance value above, we can be certain that the correlation test will yield a negative correlation.

## Correlation

	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

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
```

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
```

- iv) Based on the calculated correlation matrix above, we can see that there is a strong inverse relationship between Time spent reading and watching TV. Additionally, we see that there is an inverse relationship between time spent reading and happiness and a positive relationship between time spent watching TV and happiness. There's a non-existing relationship between gender and other three variables; happiness, TimeReading, and TimeTV.

## Correlation Coefficient and Coefficient of Determination

	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

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

- v) The coefficient of determination (CoD) estimates a linear relationship between two variables based on the variability in one variable that can be attributed to the second variable. From the coefficient of determination matrix above, we get that the CoD ( $r^2 = 0.780$ ) between 'TimeTv and 'TimeReading' is the largest among the relationships. In fact, only one relationship has a CoD value above 0.60 and can considerate a "good" CoD to make a claim about a correlation between two variables. Therefore, it is safe to conclude that there is a strong correlation between them 'TimeTv and 'TimeReading'.
- vi) Based on the analysis the correlation coefficient and coefficient of determination (CoD) matrix above, we can conclude that watching more TV did indeed caused students to read less. The correlation coefficient ( $r = -0.8831$ ) shows, together with high CoD value ( $r^2 = 0.780$ ), that there is an inverse correlation between reading and watching TV.
- vii) To perform a partial correlation, we will evaluate a linear relationship between time watching TV and reading while keeping happiness as the controlled variable. As shown below, the estimated partial correlation is 0.873 (ignore the negative sign) with significantly small p-val and high statistical value, which suggest that happiness partially explain the casual relationship between watching TV and reading. In other words, there exist other possible variables (i.e. happiness) that could explain the inverse correlation between reading and watching TV. As a result, this certainly changes our interpretation of the results because we cannot surely claim that watching tv results in reading less.

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