Assignment 05 Joshua Burden Student Survey

Joshua Burden

2022/05/01

```
ss_df <- read.csv(url('http://content.bellevue.edu/cst/dsc/520/id/resources/student-survey.csv'))
ss_df
```

```
##    TimeReading TimeTV Happiness Gender
## 1            1     90     86.20      1
## 2            2     95     88.70      0
## 3            2     85     70.17      0
## 4            2     80     61.31      1
## 5            3     75     89.52      1
## 6            4     70     60.50      1
## 7            4     75     81.46      0
## 8            5     60     75.92      1
## 9            5     65     69.37      0
## 10           6     50     45.67      0
## 11           6     70     77.56      1
```

```
head(ss_df)
```

```
##   TimeReading TimeTV Happiness Gender
## 1           1     90     86.20      1
## 2           2     95     88.70      0
## 3           2     85     70.17      0
## 4           2     80     61.31      1
## 5           3     75     89.52      1
## 6           4     70     60.50      1
```

Use R to calculate the covariance of the Survey variables and provide an explanation of why you would use this calculation and what the results indicate.

```
cov(ss_df$TimeReading, ss_df$TimeTV)
```

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

The result is negative TimeReading and TimeTV are negatively related.

```
cov(ss_df$TimeReading, ss_df$Happiness)
```

```
## [1] -10.35009
```

The result is negative TimeReading and Happiness are negatively related.

```
cov(ss_df$TimeTV, ss_df$Happiness)
```

```
## [1] 114.3773
```

The result is in positive TimeTV and Happiness are positively related.

```
cov(ss_df$TimeReading, ss_df$Gender)
```

```
## [1] -0.08181818
```

The result is negative TimeReading and Gender are negatively related.

```
cov(ss_df$TimeTV, ss_df$Gender)
```

```
## [1] 0.04545455
```

The result is positive TimeTV and Gender are positively related.

```
cov(ss_df$Happiness, ss_df$Gender)
```

```
## [1] 1.116636
```

The result is positive Happiness and Gender are positively related.

Examine the Survey data variables. What measurement is being used for the variables? Explain what effect changing the measurement being used for the variables would have on the covariance calculation. Would this be a problem? Explain and provide a better alternative if needed.

TimeReading - In hours TimeTV - In minutes Happiness - In Scale 0-100 Gender - Binary 0 and 1

Choose the type of correlation test to perform, explain why you chose this test, and make a prediction if the test yields a positive or negative correlation?

```
cov(ss_df$TimeReading, ss_df$TimeTV)
```

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

Both variables are related to time They are negatively related as a person who watches more tv they reading time will go down and vise versa

Perform a correlation analysis of:

All variables

```
cor(ss_df, use = "complete.obs", method = "pearson")
```

```
##             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
```

A single correlation between two a pair of the variables

```
cor(ss_df$TimeReading, ss_df$Happiness)
```

```
## [1] -0.4348663
```

Repeat your correlation test in step 2 but set the confidence interval at 99%

```
cor.test(ss_df$TimeReading, ss_df$Happiness, conf.level = .99)
```

```
##
##  Pearson's product-moment correlation
##
## data:  ss_df$TimeReading and ss_df$Happiness
## t = -1.4488, df = 9, p-value = 0.1813
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
##  -0.8801821  0.4176242
## sample estimates:
##        cor
## -0.4348663
```

Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

TimeReading and timeTV have a negative correlation TimeReading and Happiness have a negative correlation TimeTV and Happiness have a negative correlation All above three comparison shows that if one variable goes up the opposite variable will go down.

Calculate the correlation coefficient and the coefficient of determination, describe what you conclude about the results.

```
ss_df1 <- ss_df[, c("TimeReading", "TimeTV", "Happiness")]
cor(ss_df1)
```

```
##             TimeReading     TimeTV  Happiness
## TimeReading   1.0000000 -0.8830677 -0.4348663
## TimeTV       -0.8830677  1.0000000  0.6365560
## Happiness    -0.4348663  0.6365560  1.0000000
```

TimeTV and Happiness have a moderate correlation TimeReading and TimeTV have a strong correlation

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

Based on the above correlation, there is a strong evidence that more reading leads to less watching TV.

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.

```
install.packages("ggm", repos = "http://cran.us.r-project.org")
```

```
## Warning: dependency 'graph' is not available
```

```
##
## The downloaded binary packages are in
##  /var/folders/ct/l8y_jb751b1ghqz7j4qhxv580000gn/T//RtmpbX4HCa/downloaded_packages
```

```
library(ggm)
partial_correlation <- pcor(c("TimeReading", "TimeTV", "Happiness"), var(ss_df1))
partial_correlation^2
```

```
## [1] 0.762033
```

Happiness is the variable that is controllable as when TV time and time reading are changed, Happiness is effected.