RMarkdown Assignment Template

Marina MacDonald

May 1, 2022

I. 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.

Covariance tells you how much two random variables vary together. But what is important is that they are on the same scale/unit of measure. So for the homework this week, doing the covariance between time reading and time tv watching - isn't going to tell you much unless you convert them to the same units, and the covariance between time reading and happiness really isn't going to tell you much because they are on completely different units. But let's say that time reading and time watching tv were both in minutes, if the covariance is large, it could mean a strong relationship (but you also have to keep in mind that the larger the values, the larger the covariance. Why it really matters is that you need it to find the correlation, which is going to tell you the real strength of the relationship. Correlation also doesn't need the unit of measure/scale to be the same.

```
student_survey <- read.csv("data/student-survey.csv")
# Not converting the measurements
cov(student_survey)
```

```
##              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
```

```
# converting measurement time tv to hours
student_survey$TimeTV = student_survey$TimeTV / 60
cov(student_survey)
```

```
##              TimeReading        TimeTV  Happiness        Gender
## TimeReading   3.05454545 -0.3393939394 -10.350091 -0.0818181818
## TimeTV       -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
```

I calculated the covariance to assess whether two variables are related to each other.

The results indicate: Time Reading is negatively related to Time TV (-.033), Happiness (-10.35) and Gender (-0.08). Time TV is negatively related to Time reading (-0.33), positively related to happiness (1.90) and no relation to gender (0.0) Happiness is negatively related to Time Reading (-10.35), positively related to Time TV (1.90) and Gender (1.11)

As time reading decreases, time TV and happiness increases. As time TV decreases, time reading increases. As time TV increases, happiness increases. As happiness decreases, time reading increases. As happiness increases, time TV increases.

II. 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

```
head(student_survey)
```

```
##   TimeReading   TimeTV Happiness Gender
## 1           1 1.500000     86.20      1
## 2           2 1.583333     88.70      0
## 3           2 1.416667     70.17      0
## 4           2 1.333333     61.31      1
## 5           3 1.250000     89.52      1
## 6           4 1.166667     60.50      1
```

Currently, the data is presented in different measurements. For Time Reading, it seems to be in hours, for Time TV it seems to be in minutes, happiness (which I don't know how someone would measure), I believe is in % and gender is binary. I did the test and ran the covariance as the data was presented and then converting the TV time from minutes to hours to match the reading time and the results were the same. The covariance tells if there is a relationship between the 2 variables where the unit is the same. Happiness and gender wouldn't be possible to convert to match the first two variables.

III. 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?

```
cor(student_survey, 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
```

I used the Pearson method to know if there is relationship between watching TV and reading. I predict the correlation is negative, meaning that the more time student spend watching TV, less time to read. The test proved right with a negative correlation (-0.88).

IV. Perform a correlation analysis of:

1. All variables
2. A single correlation between two a pair of the variables
3. Repeat your correlation test in step 2 but set the confidence interval at 99%
4. Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.

```
cor(student_survey, 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
```

```
cor.test(student_survey$TimeReading, student_survey$TimeTV, method = "pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  student_survey$TimeReading and student_survey$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
```

```
cor.test(student_survey$TimeReading, student_survey$TimeTV, method = "pearson", conf.level = 0.99)
```

```
##
##  Pearson's product-moment correlation
##
## data:  student_survey$TimeReading and student_survey$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
```

The results reiterates that the Pearson correlation between time reading and time TV was -0.88, but tells us that this was highly significantly different from zero, t(101) = -5.64, p < .001. Most important, the 95% confidence (default) ranged from -0.96 to -0.60 and the 99%, -0.98 to -0.44, which does not cross zero. This tells us that the population or actual value of the correlation is negative, so the 2 variables are negatively related.

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

```
cor(student_survey$TimeReading, student_survey$TimeTV)^2 *100
```

```
## [1] 77.98085
```

This tells us how much of variability in one variable is shared by the other. Time TV shared 78% of the variability in time reading. This leaves the other 22% to be accounted by other variables.

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

No. The correlation is high, however, correlation does not equal causation. We cant assume if watching TV caused to read less or reading more caused to watch less TV.

VII. 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.

```
library(ggm)
pc<-pcor(c("TimeReading", "TimeTV", "Gender"), var(student_survey))
pc
```

```
## [1] -0.8860628
```

```
pc^2
```

```
## [1] 0.7851073
```

```
pcor.test(pc, 1, 11)
```

```
## $tval
## [1] -5.406281
##
## $df
## [1] 8
##
```

```
## $pvalue
## [1] 0.0006411949
```

The 3rd variable is the one I will control (Gender). This is a partial correlation for the variables reading and tv time but controlling for the effect of gender. The partial correlation is -0.88 which is the same when gender is not considered. The tv time accounts for 78% of the variance in reading time, the same as above. My conclusion is that gender does not have an effect on time reading or watching tv.