week7_KalaikkovanVasanthakumar

Vasanthakumar Kalaikkovan

30/04/2021

```
setwd("E:/Repos/StatisticsR/DSC520-Statistics/week7")
survey_df <- read.csv("student-survey.csv")
head(survey_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(survey_df$TimeReading,survey_df$TimeTV)
```

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

The result is in negative and thus Time-Reading and Time-TV are negatively related.

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

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

The result is in negative and thus Time-Reading and Happiness are negatively related.

The result is in positive and thus Time-TV and Happiness are positively related.

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

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

The result is in negative and thus Time-Reading and Gender are negatively related.

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

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

The result is in positive and thus Time-TV and Gender are positively related.

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

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

The result is in positive and thus 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.

Answer: Time-Reading - In hours Time-TV - In minutes Happiness - In Scale 0-100 Gender - Binary 0 and 1 Except gender all other variables are in numeric And I would just omit computing Covariance between Gender and the others since its not really relevant to the research question.

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(survey_df$TimeReading,survey_df$TimeTV)
```

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

Answer Is there relationship between time spent reading and watching TV? I choose this because both variables are time related Its negatively related and the prediction is if someone spends more time in reading their time watching TV goes down.

Perform a correlation analysis of: All variables

```
cor(survey_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(survey_df$TimeReading, survey_df$Happiness)
```

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

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

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

```
##
##  Pearson's product-moment correlation
##
## data:  survey_df$TimeReading and survey_df$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. Answer Time Reading and time TV has a negative correlation Time reading and Happiness has negative correlation Time TV and Happiness has negative correlation All above three comparison shows if one goes up other variable will go down.

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

```
survey_df1 <- survey_df[, c("TimeReading", "TimeTV", "Happiness")]
cor(survey_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
```

Time TV and Happiness has moderate correlation Time reading and Time TV has strong correlation

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

```
library(ggm)
```

```
## Warning: package 'ggm' was built under R version 4.0.5
```

```
partial_correlation <- pcor(c("TimeReading", "TimeTV", "Happiness"), var(survey_df1))
partial_correlation^2
```

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

In this dataset, Happiness is being controlled here because Time TV accounts Time Reading and thus Happiness is controlled.