StudentSurvey_YoungStephen.R

young

2022-05-01

```r
# Assignment: Student Survey
# Name: Young, Stephen
# Date: 2022-05-1

## Set the working directory to the root of your DSC 520 directory
setwd("C:/Users/young/Desktop/Classes/DSC520/GIT")

## Load the `data/r4ds/heights.csv` to
student_df <- read.csv("data/student-survey.csv")

#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.

#the covariance is -10.35009, meaning the relationship between the two is negative
cov(student_df$TimeReading, student_df$Happiness)
```

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

```r
#the covariance is 114.3773 so there is a positive relationship
cov(student_df$TimeTV, student_df$Happiness)
```

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

```r
#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
#Would this be a problem? Explain and provide a better alternative if needed.

#the measurement of TimeReading and TimeTV are both in terms of hours, Happiness is a percentage and Ge
#Changing the measurement being used would possibly effect how they relate to each other. If the time v
#minutes, then it would be different because the units are. The magnitude of the covariance is arbitrar
#the number does not necessarily show the strength of the relationship

#iii. Choose the type of correlation test to perform, explain why you chose this test, and make a predi
#a positive or negative correlation?

#I would chose the pearson correlation test to be between timeTV and timeReading and say that there is
#variables and say it would be a possitive correlation between them both
cor(student_df$TimeTV, student_df$TimeReading) #-0.8830677 is negatively correlated so as one increases
```

```
## [1] -0.8830677
```

```r
#iv. Perform a correlation analysis of:
#a. All variables
cor(student_df)
```

```
##             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
```

```r
#b. A single correlation between two a pair of the variables
cor(student_df$TimeTV, student_df$Gender)
```

```
## [1] 0.006596673
```

```r
#.006596673, very low to no relationship between the two

#c. Repeat your correlation test in step 2 but set the confidence interval at 99%
cor.test(student_df$TimeTV, student_df$Gender, conf.level = 0.99)
```

```
##
##  Pearson's product-moment correlation
##
## data:  student_df$TimeTV and student_df$Gender
## t = 0.01979, df = 9, p-value = 0.9846
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
##  -0.7182866  0.7246128
## sample estimates:
##         cor
## 0.006596673
```

```r
#d. Describe what the calculations in the correlation matrix suggest about the relationship between the
#There seems to be correlation between TimeTV and Happiness. Everything else is a weak
#relationship (Time TV and Gender .006596673 and Happiness and Gender .157011838) or negative relations
#(TimeReading to TimeTV -0.8831, TimeReading to Happiness -0.4349,TimeReading to Gender -0.0896, Happin

#v. Calculate the correlation coefficient and the coefficient of determination, describe what you concl
#correlation coefficient
cor(student_df)
```

```
##             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
```

```r
#coefficient of determination
student.lm = lm(formula = Happiness ~ TimeReading + TimeTV+Gender, data = student_df)
summary(student.lm)$r.squared
```

```
## [1] 0.5205882
```

```r
#The coefficient of determination is .5205882 which means that there is a 52.06% chance of explaining t
#between the variables so only 52.06% of the relation between happiness, time reading, Time TV and Gend
#rather low amount of explaination between the variables


#vi. Based on your analysis can you say that watching more TV caused students to read less? Explain.
# the relation between TV and Reading was a negative coefficient of -0.883067681 which means there is a
#of negative relation between the two, so as one increases the other one decreases and vice versa so if
#rise, this would cause reading to go down.

#vii. Pick three variables and perform a partial correlation, documenting which variable you are "contr
#Explain how this changes your interpretation and explanation of the results
library(ppcor)
```

```
## Loading required package: MASS
```

```r
pcor.test(student_df$TimeReading, student_df$TimeTV, student_df$Happiness)
```

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

```r
#the estimate is -0.872945 which is a negative correlation between the two variables, the p.value is .0
#means that the it is not statistically significant and it is controlling for happiness.
```