Student_Survey_RamirezKyleRM

Kyle Ramirez

1/16/2022

library(magrittr) library(dplyr)

Set the working directory to the root of your DSC 520 directory

setwd("/Users/Kyle/Documents/GitHub/KR/Ramirez_Kyle_DSC510/dsc520")

Load the `data/r4ds/student-survey` to

studentsurvey <- read.csv("data/r4ds/student-survey.csv")

Using `cov()` compute correclation coefficients for

TimeReading vs. Gender

cov(studentsurvey$TimeReading$, $studentsurvey$Gender) # Analysis thought: Small negative relation; will move opposite of each other as one variable changes

TimeReading vs. Happiness

cov(studentsurvey$TimeReading$, $studentsurvey$Happiness) # Analysis thought: Large negative relation; will move opposite of each other as one variable changes

TimeTV vs. Gender

cov(studentsurvey$TimeTV$, $studentsurvey$Gender) # Analysis thought: Small positive relation; will move together along side each other as one variable changes

TimeTV vs. Happiness

cov(studentsurvey$TimeTV$, $studentsurvey$Happiness) # Analysis thought: Large positive relation; will move together along side each other as one variable changes

Happiness vs. Gender

cov(studentsurvey$Happiness$, $studentsurvey$Gender) # Analysis thought: Small positive relation; will move together along side each other as one variable changes

TimeTV vs. TimeReading

cov(studentsurvey$TimeTV$, $studentsurvey$TimeReading) # Analysis thought: Large negative relation; will move opposite of each other as one variable changes

studentsurvey %>% dplyr::select(TimeTV, TimeReading, Gender, Happiness) %>% cov()

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

cor.test(studentsurvey$TimeTV$, $studentsurvey$Happiness, alternative = c("two.sided", "less", "greater"), method = c("pearson", "kendall", "spearman"), exact = NULL, conf.level = 0.99, continuity = FALSE)

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

TimeTV and Happiness have the significantly highest covariance

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

TimeTV vs. Happiness

cor(studentsurvey$TimeTV$, $studentsurvey$Happiness)

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

Looking at the data's covariance. There shows to be a strong negative relation. When one variable moves. The other will move in the opposite direction.

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.

I dont think that the TimeTV and TimeReading can be matched together very well. Even though that you can see that there is more time spent watching tv than reading. They seem to be too independent of variables to try to get a correlation with them.

TimeTV and gender is another one that I don't think would be good since it's close to 0 and having no correlation at all.

TimeReading and happiness seem to be interesting. As it seems when you compare Time Reading and TimeTV and happiness. There is a higher correlation to people being happier watching TV than reading.