Exercise 9: Student Survey

Nosky, Christopher

1/17/2021

As a data science intern with newly learned knowledge in skills in statistical correlation and R programming, you will analyze the results of a survey recently given to college students. You learn that the research question being investigated is: “Is there a significant relationship between the amount of time spent reading and the time spent watching television?” You are also interested if there are other significant relationships that can be discovered? The survey data is located in this StudentSurvey.csv file.

* 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.
  + We are using the covariance function cov() to measure each variable’s linear relationship. In the examples above we can see that the covariance of TimeReading with each of the other variables is negative indicating no linear relationship. With TimeTV we can see a positive linear relation ship between it and Happiness indicating that when one changes the other changes with it.

## 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

* 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.
  + No measurement scale was given with regards to the variables nor was any length of time the students were reporting for, after examining the data I might suggest that TimeReading and TimeTV were measured in hours while Happiness measured as a Percentage.
  + Additionally, the data doesn’t suggest when the students happiness was measured, was it measured immediately after watching tv or reading?
  + An alternative would be to use either Kendall or Spearman methods.
* 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?
  + For this test I’m going to test with the method for Pearson’s r. The reason for this choice is due to the positive linear relationship demonstrated with the covariance test. I predict there will be a positive correlation between TimeTV and Happiness.
* Perform a correlation analysis of:
  + All variables

## 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

## [1] 0.636556

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

##   
## Pearson's product-moment correlation  
##   
## data: student\_survey\_df$TimeTV and student\_survey\_df$Happiness  
## t = 2.4761, df = 9, p-value = 0.03521  
## alternative hypothesis: true correlation is not equal to 0  
## 99 percent confidence interval:  
## -0.1570212 0.9306275  
## sample estimates:  
## cor   
## 0.636556

* Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation.
* Calculate the correlation coefficient and the coefficient of determination, describe what you conclude about the results.
* Based on your analysis can you say that watching more TV caused students to read less? Explain. \*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.