RMarkdown Assignment Template

Sarah Qiang

May 01 2022

assignment 05

Set the working directory to the root of your DSC 520 directory setwd("/Users/sarah/documents/hello_world")

Load the `data/r4ds/heights.csv` to heights_df <- read.csv("/Users/sarah/documents/hello_world/data/r4ds/heights.csv")

```
Using 'cor()' compute correclation coefficients for
cor(heights_df$height, heights_df$earn, method = "pearson")
[1] 0.2418481
```

age vs. earn cor(heights_df$age, heights_df$earn, method = "pearson") [1] 0.08100297

```
ed vs. earn
cor(heights_df$ed,heights_df$earn, method = "pearson")
[1] 0.3399765
```

Spurious correlation The following is data on US spending on science, space, and technology in millions of today's dollars and Suicides by hanging strangulation and suffocation for the years 1999 to 2009

```
Compute the correlation between these variables
tech_spending <- c(18079, 18594, 19753, 20734, 20831, 23029, 23597, 23584, 25525, 27731, 29449)
suicides <- c(5427, 5688, 6198, 6462, 6635, 7336, 7248, 7491, 8161, 8578, 9000)
cor(tech_spending,suicides,method = "pearson")
```

#Student-survey result

survey_df <- read.csv("/Users/sarah/documents/hello_world/data/r4ds/student-survey.csv")

```
I. covariance
happiness = survey_df$Happiness
tv = survey_df$TimeTV
timereading = survey_df$TimeReading
cov(happiness, tv)
[1] 114.3773
```

```
The covariance of eruption duration and waiting time is about 114. It indicates a positive linear relat
```

II. What measurement is being used for the variables? timereading is an ordinal variable happinese is an ordinal variable timetv is an ordinal variable gender is a nominal variable

III. Choose the type of correlation test to perform, explain why you chose this test, and make a predict
I chose the pearson method because I believe there's a linear correlation between the two quantitative

IV. Perform a correlation analysis of: A single correlation between two a pair of the variables cor(happiness, tv, method = "pearson") [1] 0.636556 Repeat your correlation test in step 2 but set the confidence interval at 99% cor(happiness, tv, method = "pearson",level =0.99)

Describe what the calculations in the correlation matrix suggest about the relationship between the variables. Be specific with your explanation. The correlation coefficient of happiness and TV is 0.636556 Since it is rather close to 1, we can conclude that the variables are positively linearly related.

```
V. Calculate the correlation coefficient and the coefficient of determination, describe what you conclu
head(survey_df)
happiness.lm = lm(happiness ~ tv, data=survey_df)
summary(happiness.lm)$r.squared
[1] 0.4052035
The coefficient of determination of the simple linear regression model for the data set faithful is 0.4
```

Vi. Based on your analysis can you say that watching more TV caused students to read less? Explain. cor(tv, timereading, method = "pearson") [1] -0.8830677 cov(tv, timereading) [1] -20.36364

the correlation between the time student spent watching TV and reading is -0.8830677, it creates a neative linear colrreation, the more time students spent on watching TV, the less they read.

```
VII. Partial colrreation
install.packages("ppcor")
library(ppcor)
pcor(survey_df)

$estimate
            TimeReading     TimeTV Happiness      Gender
TimeReading   1.0000000 -0.8827973 0.4013124 -0.2706036
TimeTV       -0.8827973  1.0000000 0.6311611 -0.2943135
Happiness     0.4013124  0.6311611 1.0000000  0.2833152
Gender       -0.2706036 -0.2943135 0.2833152  1.0000000

$p.value
            TimeReading      TimeTV  Happiness     Gender
TimeReading 0.000000000 0.001615344 0.28437887 0.4812716
TimeTV      0.001615344 0.000000000 0.06832112 0.4420392
Happiness   0.284378868 0.068321119 0.00000000 0.4600603
Gender      0.481271572 0.442039185 0.46006033 0.0000000

$statistic
            TimeReading     TimeTV Happiness      Gender
TimeReading   0.0000000 -4.9720962 1.1592148 -0.7436966
TimeTV       -4.9720962  0.0000000 2.1528933 -0.8147673
Happiness     1.1592148  2.1528933 0.0000000  0.7816064
Gender       -0.7436966 -0.8147673 0.7816064  0.0000000

$n
[1] 11
```

```
$gp
[1] 2

$method
[1] "pearson"
```