week7_VayuvegulaSomaShekar

Soma Shekar Vayuvegula

2022-07-24

```
setwd("/Users/somashekarvayuvegula/Documents/Workspace/dsc520")
survey_df <- read.csv("completed/week7/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 TimeReading and TimeTV are negatively related.

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

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

The result is in negative and thus TimeReading and Happiness are negatively related.

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

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

The result is in negative and thus TimeReading and Gender are negatively related.

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

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

The result is in positive and thus TimeTV and Happiness are positively related.

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

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

The result is in positive and thus TimeTV 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 - Numeric 0-100 (O being lowest and 100 being highest) Gender - Binary 0 and 1 Except gender all other variables are in numeric I would change the gender from numeric to String as in current world people do not represent themselves using binary gender.They represent themselves using various forms which include but not limited to gender queer, agender etc., I would suggest avoid 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?

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

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

Answer: I have chosen to calculate correlation between Time Reading v Time TV as both are time related. Correlation is negative and closer to -1 which mean both are in inverse relationship. This suggest time spent on reading is inverse to time spent on TV. If a person spends more time in watching TV,will spend less time reading or vice-versa.

Perform a correlation analysis of:

```
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
```

```
cor(survey_df$TimeTV,survey_df$Happiness)
```

```
## [1] 0.636556
```

```
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 reading and Gender has negative correlation All above three comparison shows if one goes up other variable will go down, which means they have inverse relationship. Time TV and Happiness has positive correlation Time TV and Gender has positive correlation Happiness and Gender has positive correlation All above three comparison shows both variables have closer relationship.

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

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

```
##             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 strong correlation Time reading and Time TV has inverse correlation Time reading and Happiness has moderate correlation

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

Answer: Based on my analysis, Time reading and Time TV has inverse correlation, which means if watching TV goes up, reading time goes down. This proves that watching more tv caused students to read less.

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

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

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