Assignment_07.2 Students survey

KarthikeyanChellamuthu

30/01/2022

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.

```
str(students)
```

```
## 'data.frame':    11 obs. of  4 variables:
##  $ TimeReading: int  1 2 2 2 3 4 4 5 5 6 ...
##  $ TimeTV     : int  90 95 85 80 75 70 75 60 65 50 ...
##  $ Happiness  : num  86.2 88.7 70.2 61.3 89.5 ...
##  $ Gender     : int  1 0 0 1 1 1 0 1 0 0 ...
```

```
head(students)
```

```
##   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
```

```
cov(students)
```

```
##              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
```

I can consider covariance for TimeReading vs. Time TV since they are same variable as related to time. Since the units between the two are not identical the interpretation becomes fairly difficult to understand, but at the very least, the covariance output indicates that the two have a negative covariance. This means that as you spend more time reading you will spend less time watching TV and vice versa which makes sense in this case.

ii.Examine the Survey data variables. What measurement is being used for the variables? Explain what effect changing the measurement

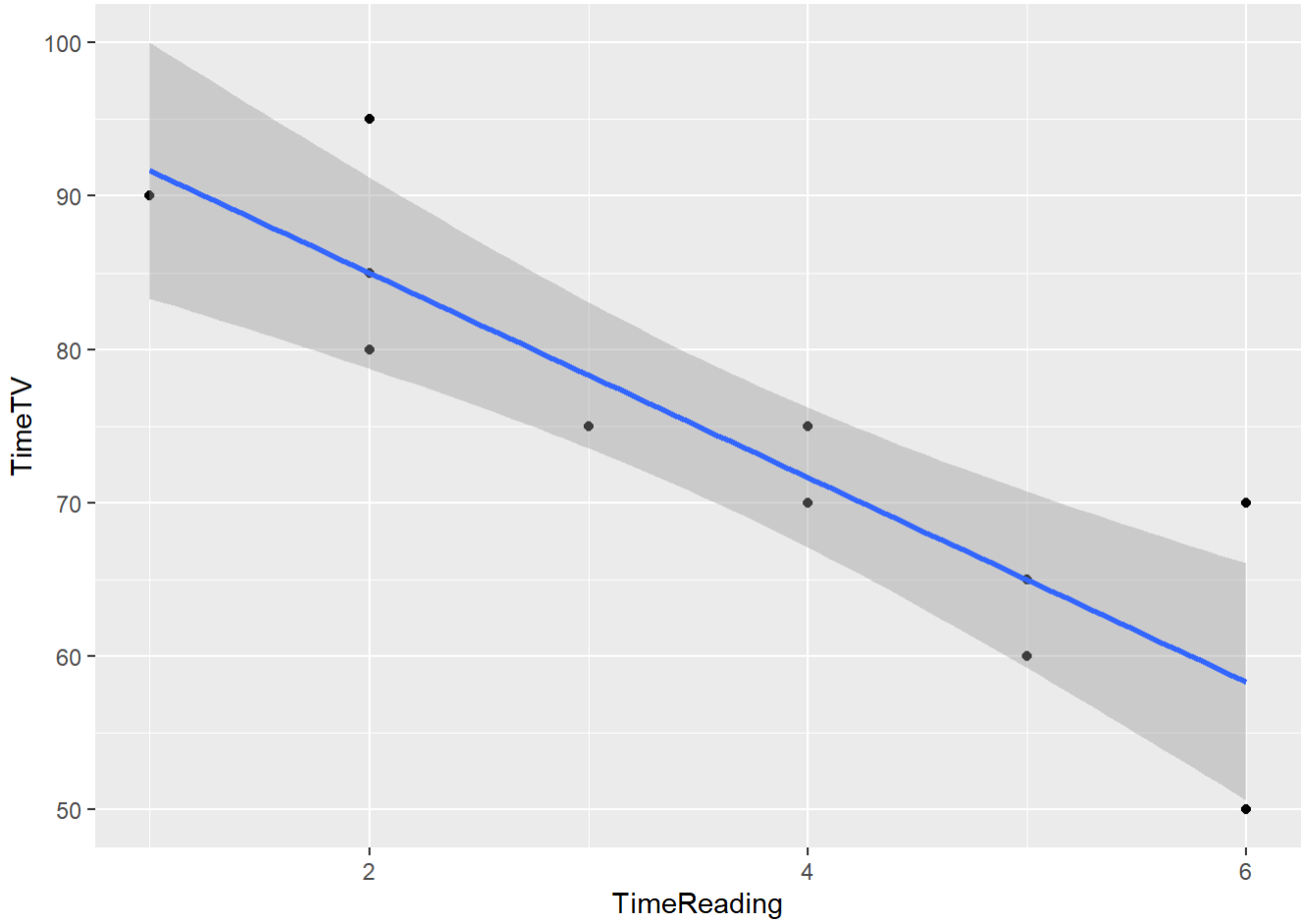
being used for the variables would have on the covariance calculation. Would this be a problem? Explain and provide a better alternative if needed.

Covariance cannot be compared over data sets with different scales. Large covariance means the relationship between two variables is strong. One problem with interpretation is with wide range of results that that makes it hard to interpret. It can be see that the covariance produced for Happiness and TimeTV is higher than TimeReading and Gender.

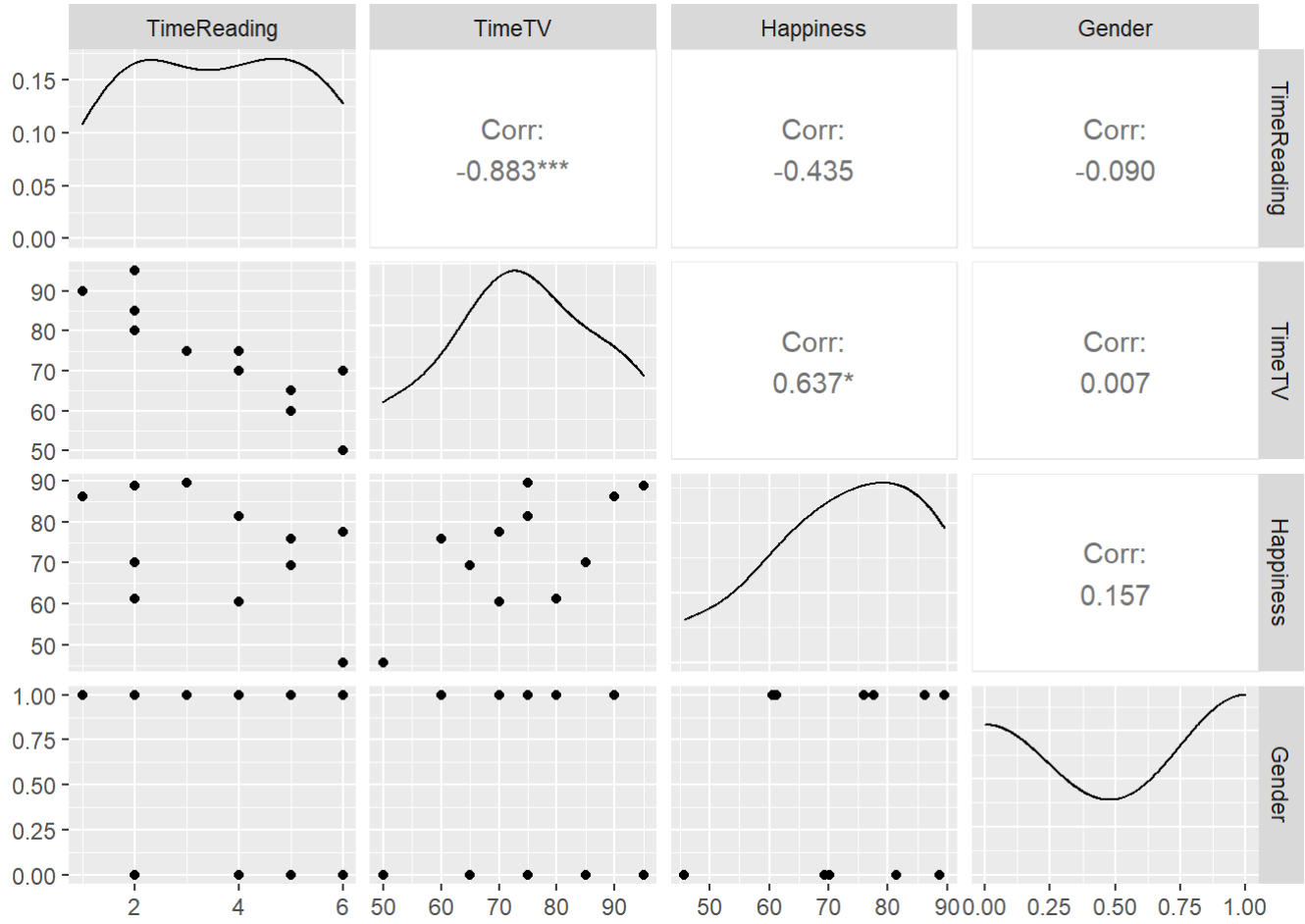
iii.Choose the type of correlation test to perform, explain why you chose this test, and make a prediction if the test yields a positive or negative correlation?

I would choose pearson correlation as there is linear relationship between TimeReading vs TimeTv. Also, it can be see that correlation is negative.

```
## `geom_smooth()` using formula 'y ~ x'
```



##iv - 1. Correlation analysis for all variables Firstly, here is a correlation table for all the variables. All correlations for the gender variable will be ingored as this table output is generated using the pearson correlation method. And here are the results of the point-biserial correlation tests.



iv-2. A single correlation between two of the variables

```
cor.test(students$TimeReading, students$TimeTV)
```

```
##
##  Pearson's product-moment correlation
##
## data:  students$TimeReading and students$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.9694145 -0.6021920
## sample estimates:
##        cor
## -0.8830677
```

iv-3. Repeat your correlation test in step 2 but set the confidence interval at 99%

```
##
##  Pearson's product-moment correlation
##
## data:  students$TimeReading and students$TimeTV
## t = -5.6457, df = 9, p-value = 0.0003153
## alternative hypothesis: true correlation is not equal to 0
## 99 percent confidence interval:
##  -0.9801052 -0.4453124
## sample estimates:
##        cor
## -0.8830677
```

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

4. t value is -5.6457, degree of freedom(df) if 9, p-value is 0.0003153, confidence interval is 99%. Alpha = 0.05. Since the p-value is less than alpha, TimeReading and TimeTV are significantly correlated.

v. Calculate the correlation coefficient(Done in the previous step) and the coefficient of determination, describe what you conclude about the results.

```
##             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
```

```
##             TimeReading       TimeTV  Happiness       Gender
## TimeReading 1.000000000 0.7798085292 0.18910873 0.0080357143
## TimeTV      0.779808529 1.0000000000 0.40520352 0.0000435161
## Happiness   0.189108726 0.4052035234 1.00000000 0.0246527174
## Gender      0.008035714 0.0000435161 0.02465272 1.0000000000
```

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

Yes, watching more TV has caused students to read less. This can be validated using below. Correlation coefficient which is -0.88. Negative correlation indicates means TimeReading to TimeTV linear relationship in opposite direction. Also, coefficient of determination is 77% which validates the above point.

vii. 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 chose to look at the correlation between time spent watching tv (TimeTV) and Happiness, while controlling for time spent reading. Since time spent reading drastically influenced time spent watching TV I wanted to see how much watching TV correlated to happiness while controlling for time spent reading.

```
# Partial Correlation of TimeTv vs. Happiness, while controlling for "TimeReading"
pcr1
```

```
## [1] 0.5976513
```

```
# R^2 of TimeTV vs. Happiness, while controlling for "TimeReading"
pcr1 * pcr1
```

```
## [1] 0.3571871
```

##Output shows the partial correlation between TimeTV and Happiness with Gender was control parameter. It can be see that partial correlation is 0.59. pc^2 results in 0.35 which mean Happiness accounts for 40% of TimeTV. Inclusion of Gender has impact on Happiness on TimeTV. Also, P-value(0.04) is less than 0.5.