

```

---
title: "ASSIGNMENT 7"
author: "Saurabh Shrestha"
date: '2021-05-03'
output:
  word_document: default
  html_document: default
  pdf_document: default
bibliography: C:/Users/Saurabh/Desktop/DSC 520/Week2/dsc520
---

```

```

## Set the working directory to the root of your DSC 520 directory
```{r}
setwd("C:/Users/Saurabh/Desktop/DSC 520")
```

```

```

## Load the `data/r4ds/heights.csv` to
```{r}
student_df <- read.csv("C:/Users/Saurabh/Desktop/DSC
520/Week2/dsc520/data/student-survey.csv")
head(student_df)
```

```

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."

```

```{r}
cov(student_df)
```

```

"The reason is it shows how time reading with time spent on tv and happiness is interrelated. Higher time reading means less time spent on tv but happiness is negative whereas higher time spent on tv lead to higher level of happiness."

"ii. Examine the Survey data...."

"Time reading in hours  
Time on TV is on minutes  
Happiness is on a scale of 1-100  
Gender is binary which is 0 and 1  
A different measurement applicable or which would be used is genre of books and time of the day and another measurement would be changing the time:since  
both time spent on tv and time spend reading book would have been converted into minutes. This would give us approx exact minute to minute variance of association between the variables."

"iii. Choose the type of correlation test to perform."

"To get single corelation coefficient I used only the variables because by

default this function produces Pearson's r and a 95% confidence interval  
95% confidence ranged from -.969 to -.602 which does not  
cross zero which tells us that in time reading and time on tv is  
negatively  
related."

```
```{r}  
cor.test(student_df$TimeReading, student_df$TimeTV)  
```
```

#iv.Perform a correlation analysis of...

```
#1.  
```{r}  
cor(student_df)  
l11  
#2.  
```{r}  
cor(student_df$TimeReading, student_df$TimeTV, use = "complete.obs",  
method = "pearson")  
```  
#3.  
```{r}  
cor.test(student_df$TimeReading, student_df$TimeTV, use = "complete.obs",  
method = "pearson", conf.level = 0.95)  
```
```

#4.  
The relationship between variables based on the student survey matrix  
suggests that  
students who spend more on reading tend to spend less time watching tv  
and higher time spent on tv lead to increase in happiness."

```
#v.  
# correaltion coefficient  
```{r}  
cor(student_df, use = "complete.obs", method = "pearson")  
```  
# coefficient of determination  
```{r}  
cor(student_df)^2  
```
```

Above we squared the student survey variables by usin ^2. This way we can  
see  
that how much of the variability in Time reading is shared by happiness  
and how much variability in Time spend on tv is shared by happiness.

```
#vi.  
cor(student_df$TimeReading, student_df$TimeTV)^2  
"Based on correlation of determination, it shows 0.77 which converted  
to percentage is 77% of variability  
between students time reading with time spent on tv."
```

```
#vii. Pick three variables and perform a partial correlation..."
```{r}
install.packages("ggm")
library(ggm)
partial_correlation <- pcor(c("TimeReading", "TimeTV", "Happiness"),
var(student_df))
partial_correlation^2g, suicides)
```
```