Will students who have high GPA in high school get high GPA in the first-year college?

## Data description

## The FirstYearGPA data includes details from a study of 219 first year students at a midwestern college. In this topic, we will look into two set of data which are first-year college GPA and high school GPA. Those two sets of data reflects academic performance of students in the first year of their college and high school respectively and we will examine them accordingly.

## Analysis description

As we are going to investigate if there is a relationship between the academic performance of students in high school and college, the most appropriate analysis to answer this question is Pearson’s correlation coefficient. Pearson’s correlation coefficient is a measure of the degree of linear dependence between two quantitative variables. We can see how strong variables are correlated by checking the value of Pearson’s correlation coefficient.

Interpretation

From the analysis that we have done below, it is clear to see that the Pearson’s correlation coefficient is equal to 0.4468873 which is less than 0.5. This indicates that the high school GPA has a weak positive linear relationship with the first-year college GPA. Thus, student who perform worse in high school academically will get the low GPA in first year of college but students who have high GPA in high school doesn’t imply that they will get high GPA in the first year of college. As we can see from graph in figure 1, there are certain number of students who has high GPA(>3.3) in high school but scoring below 3.0 in their first year college. Therefore, student who perform well in high school can’t be promised to have same performance in the first year of their college.

Conclusion

There are actually sort of relationship between students’ academic performance in first year college and high school but it is a weak relationship.

## Implementation

#set up  
install.packages("ggplot2")  
library(ggplot2)  
#load the data  
student\_info<-read.csv("FirstYearGPA.csv",header=T)  
#set up variables  
x<-student\_info$HSGPA  
y<-student\_info$GPA  
#Graph plotting  
graph\_GPA<-ggplot(student\_info,aes(x=HSGPA,y=GPA))+geom\_point(size=0.35,color="forestgreen") +geom\_smooth(size=0.4,method="lm",color="red",se=F)+labs(x="High school GPA",y="First year college GPA",title="First year college GPA versus High school GPA")+theme(axis.text=element\_text( size=11),plot.title=element\_text(size=11,hjust = 0.5))  
graph\_GPA

## `geom\_smooth()` using formula 'y ~ x'

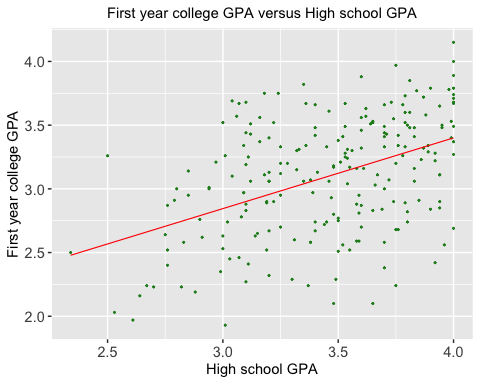


Figure 1

#Pearson’s Correlation Coefficient r  
size\_of\_sample<-length(x)  
mean\_x<-mean(x)  
mean\_y<-mean(y)  
numerator<-sum((x-mean\_x)\*(y-mean\_y))  
denominator\_1<-(sqrt(sum(((x-mean\_x)^2))))  
denominator\_2<-(sqrt(sum(((y-mean\_y)^2))))  
r<-numerator/(denominator\_1\*denominator\_2)  
r

## [1] 0.4468873