

GPA vs. Hours of Study

Is there a relation?

Introduction:

The purpose of this project is to discuss whether the length of time students studying in the library has any impact on their GPA. I always prefer to study in the library since it provides higher efficiency. I am very inefficient while study at home during the pandemic, which decreases my GPA. As a result, I conduct this study to see whether others experience the same situation as me. To start the study, I build a survey with 4 questions that included the essential data. The first two questions gather categorical variables and the rest gather qualitative variables:

What Year Are You ?

What Major ?

Your GPA ?

Average Hours/Week Spent in Lib

In the end, I gathered 32 responses, which satisfied our project requirement (at least 30 responses).

2 Categorical Variables:

YEAR: Freshman/Sophomore/Junior/Senior/Graduate

MAJOR: Determine what professional field students studying in.

2 Quantitative Variables:

HOURS: Determine the average time that a student spends in library every week.

GPA: Determine Students' current GPA.

Descriptive Statistics

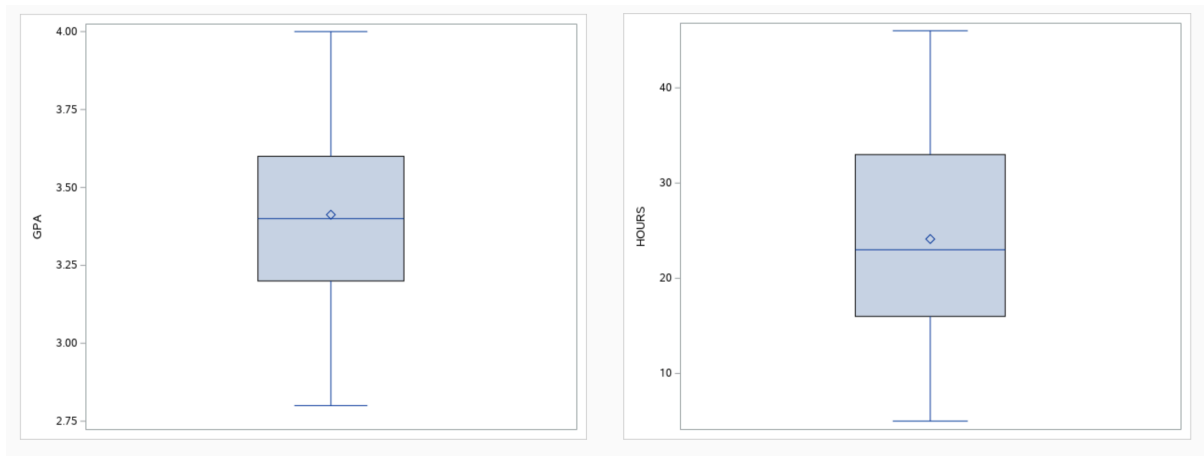
Quantitative Variables:

I use PROC MEAN to determine the five-number summary for our two quantitative variables (GPA/HOURS). From this summary, we can get an overview of our data distribution. Since both of our means for GPA and HOURS are larger than their corresponding medians, the overall data distribution is skewed to the right. However, the difference between means and their corresponding medians is small. Therefore, the data is barely skewed, and the actual skewness should not be easy to discover on the graph.

The MEANS Procedure

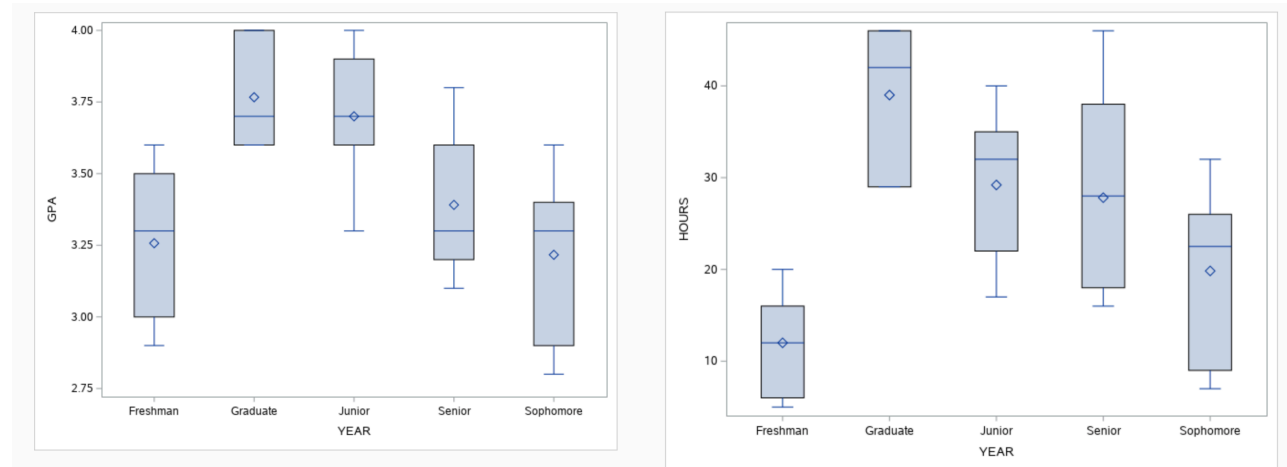
Variable	Minimum	Lower Quartile	Median	Upper Quartile	Maximum	Mean
GPA	2.80	3.20	3.40	3.60	4.00	3.41
HOURS	5.00	16.00	23.00	33.00	46.00	24.13

I use PROC SGPLOT to construct box plots for both variables. Overall, we see no outliers in our data. Both plots show characteristics as the five-number summary suggests: the data is barely skewed to the right (skewed to the top in our vertical plots). We can also observe that variable GPA has a smaller interquartile range than variable HOURS, which means that GPA has less variability than HOURS.

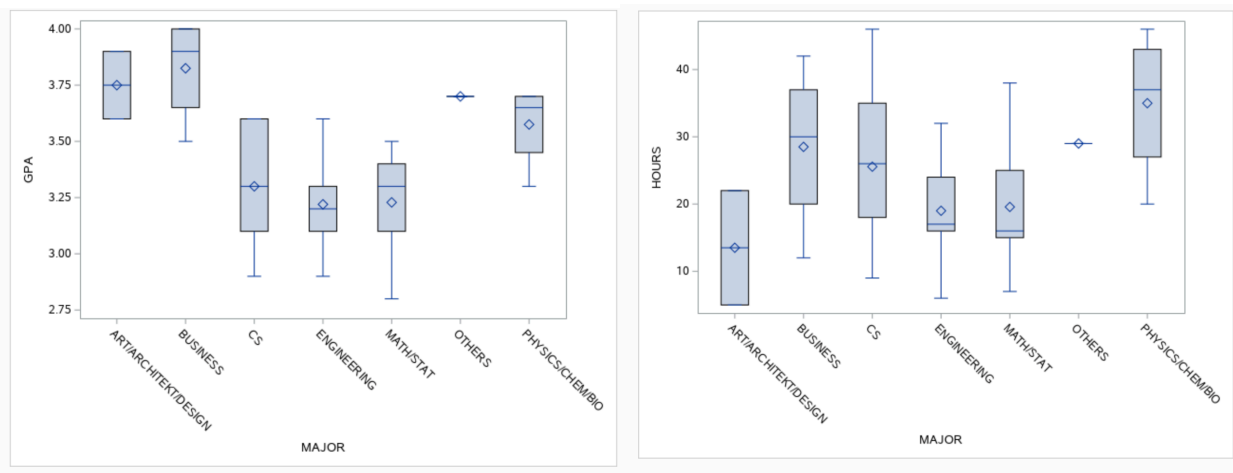


To have a more comprehensive look at the data, I also divided both box plots into different portions determined by YEAR and MAJOR. In this way, we can determine different study characteristics for different groups of students.

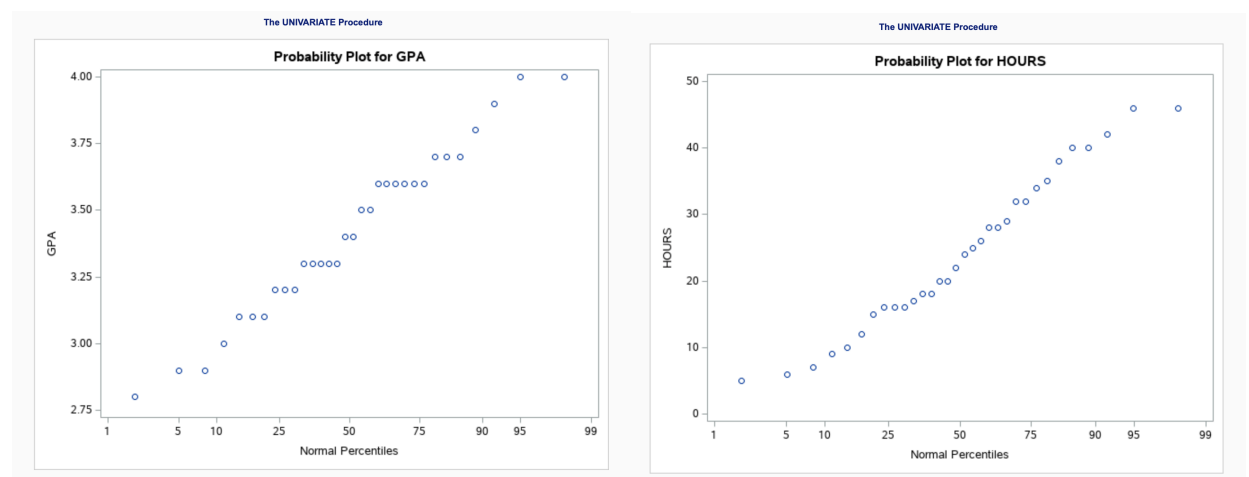
Base on YEAR, we can see some positive linear relationship between GPA and HOURS. I conclude that graduate students have a high concentration in their studies while undergraduate students are less concentrated in studying during their freshman year (lowest GPA with lowest HOURS). Undergraduate students' concentration increase in the next few years, reaching the summit during their junior year, and then falls back a small portion during their last year in college. This conclusion base on the box-plot is very reasonable considering our habit: Partying hard during the first year, catching up during the second year, reaching the top during the third year since we are about to find a job or applying graduate school, and falls back a little bit during the last semester.



Base on MAJOR, we can also see linear positive relation. However, the linearity is weaker in this situation, especially for students major in BUSINESS and ART/ARCHITECT/DESIGN. Those students have higher than usual GPA, and students major in ART/ARCHITECT/DESIGN even have a lower than usual HOURS compare to others. I can only assume that there are bias in my respondents, or these majors are easier compare with others.



I use PROC UNIVARIATE to construct probability plots for both variables. Overall, we see similar distributions for both variables. The only difference been that data for HOURS are more linearly distributed while data for GPA has several same values. This is because the range for our GPA is smaller compare to HOURS, so there is a larger possibility to record same values.



I use PPROC CORR to calculate the correlation between GPA and HOURS. We find a Pearson Correlation of 0.68964 between these two variables, meaning they have a relatively strong positive correlation. When a student has high GPA, he/she also spend more time studying in the library.

The CORR Procedure

1 With Variables:	HOURS
1 Variables:	GPA

Simple Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
HOURS	32	24.12500	11.88588	772.00000	5.00000	46.00000
GPA	32	3.41250	0.31699	109.20000	2.80000	4.00000

Pearson Correlation Coefficients, N = 32 Prob > r under H0: Rho=0		
	GPA	
HOURS	0.68964	<.0001

Categorical Variables:

I use PPROC FREQ and TABLE to build the frequency table for YEAR and MAJOR. Overall, most of our respondents are Seniors who major in Computer Science.

The FREQ Procedure

YEAR	Frequency	Percent
Freshman	7	21.88
Graduate	3	9.38
Junior	5	15.63
Senior	11	34.38
Sophomore	6	18.75

MAJOR	Frequency	Percent
ART/ARCHITEKT/DESIGN	2	6.25
BUSINESS	4	12.50
CS	9	28.13
ENGINEERING	5	15.63
MATH/STAT	7	21.88
OTHERS	1	3.13
PHYSICS/CHEM/BIO	4	12.50

To have a more concrete understanding on our population, I use PROC GCHART and PIE to build pie charts for both variables.

