Ahmedabad University

**Winter Semester 2023**

**MAT502 Advanced Statistics**

**Section 1**

**Faculty: Shashi Prabh**

# Study on Extracurricular Activities and Academic Performance of a College Student

**Group Members:**

| Enrollment Number | Name |
|---|---|
| AU2040097 | Samarth Chauhan |
| AU2040253 | Ushmay Patel |

## Motivation

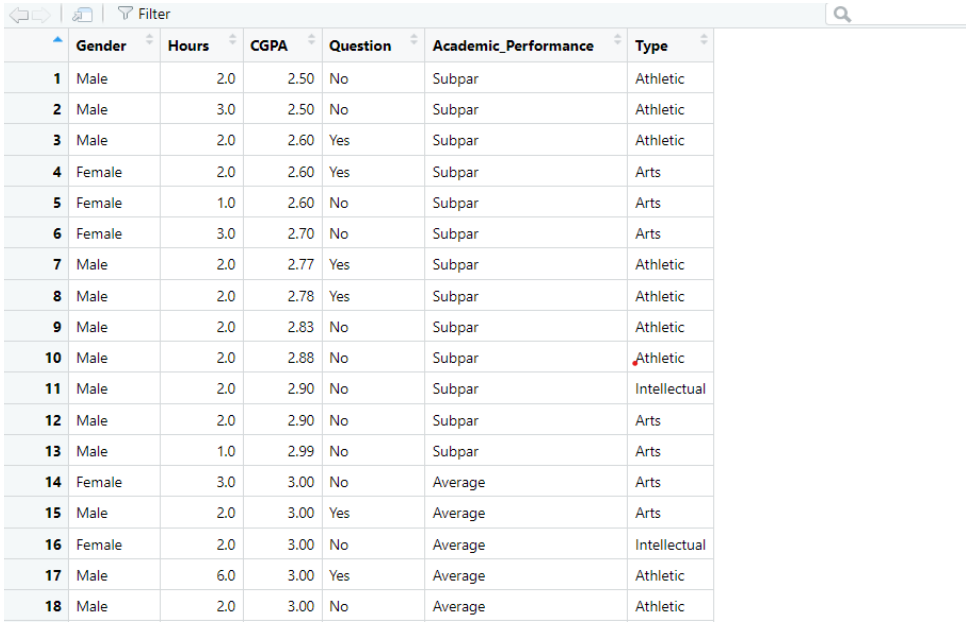
To identify the types of extracurricular activities that have the most significant impact: Different types of extracurricular activities may have different effects on academic performance. For instance, sports activities may improve physical health and self-esteem, while arts activities may promote creativity and critical thinking. By studying the impact of different extracurricular activities on academic performance, educators can identify the most effective activities for enhancing academic success.

## Objective

Our objective is to use statistical tests and methods to determine relation between extracurricular activities and academic performance of a college student. We want to test two hypotheses related to this case study.

## Data Collection

We conducted a survey of 81 college students and asked them some questions and prepared an excel file for analysis purposes. It was a method of convenience sampling. The questions were based on extracurricular activities and their academic performance. Most of the students didn't think there is an effect of extracurricular activities on their academic performance in any way. We have used ***RStudio*** for statistical analysis and Hypothesis testing.


Filter

| | Gender | Hours | CGPA | Question | Academic_Performance | Type |
|---|---|---|---|---|---|---|
| 1 | Male | 2.0 | 2.50 | No | Subpar | Athletic |
| 2 | Male | 3.0 | 2.50 | No | Subpar | Athletic |
| 3 | Male | 2.0 | 2.60 | Yes | Subpar | Athletic |
| 4 | Female | 2.0 | 2.60 | Yes | Subpar | Arts |
| 5 | Female | 1.0 | 2.60 | No | Subpar | Arts |
| 6 | Female | 3.0 | 2.70 | No | Subpar | Arts |
| 7 | Male | 2.0 | 2.77 | Yes | Subpar | Athletic |
| 8 | Male | 2.0 | 2.78 | Yes | Subpar | Athletic |
| 9 | Male | 2.0 | 2.83 | No | Subpar | Athletic |
| 10 | Male | 2.0 | 2.88 | No | Subpar | Athletic |
| 11 | Male | 2.0 | 2.90 | No | Subpar | Intellectual |
| 12 | Male | 2.0 | 2.90 | No | Subpar | Arts |
| 13 | Male | 1.0 | 2.99 | No | Subpar | Arts |
| 14 | Female | 3.0 | 3.00 | No | Average | Arts |
| 15 | Male | 2.0 | 3.00 | Yes | Average | Arts |
| 16 | Female | 2.0 | 3.00 | No | Average | Intellectual |
| 17 | Male | 6.0 | 3.00 | Yes | Average | Athletic |
| 18 | Male | 2.0 | 3.00 | No | Average | Athletic |

## 1. Chi-square Test of Independence

A Chi-square test of independence is used to determine if there is a significant relationship between two categorical variables.

We need a Chi-square test to determine if there is any association between the type of extracurricular activity and the academic performance of a college student. So we first performed Chi-square test where the hypotheses were as follows:

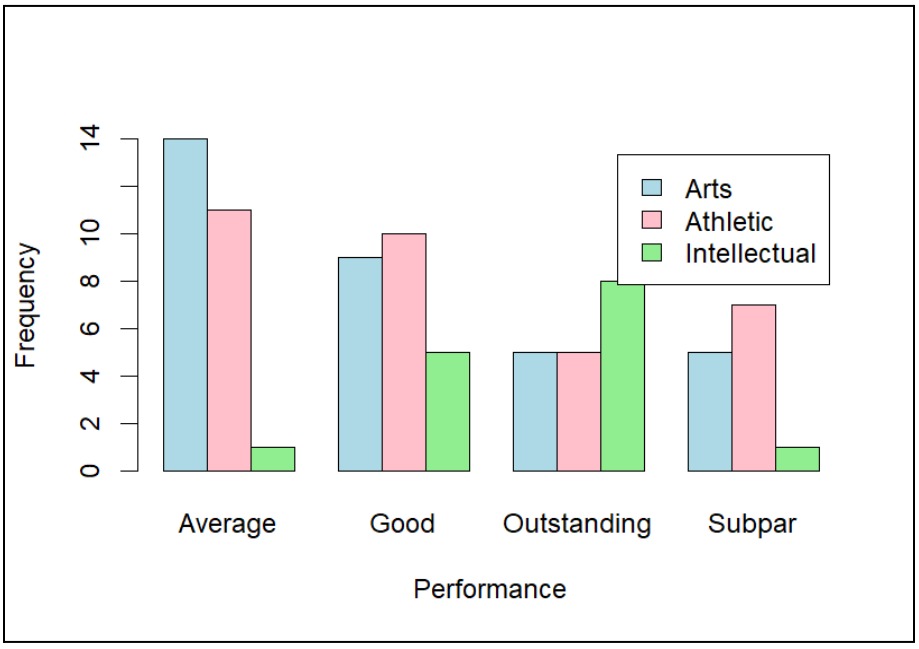
***H0:*** *There is no relation between Type of extracurricular activity and Academic performance of a college student.*

***H1:*** *There is a relation between Type of extracurricular activity and Academic performance of a college student.*

It is necessary to make sure that we have enough data with at least five observations for 80% of the cells in the Observed table.

Since only 2 cells have less than five observations we can perform the Chi-square test.

| | Average | Good | Outstanding | Subpar |
|---|---|---|---|---|
| Arts | 14 | 9 | 5 | 5 |
| Athletic | 11 | 10 | 5 | 7 |
| Intellectual | 1 | 5 | 8 | 1 |



```r
```{r,warning=FALSE}
#chi-square test
chisqTest <- chisq.test(Chisq_table)
print(chisqTest)

#degree of freedom
df <- chisqTest$parameter

critical_value <- qchisq(1 - alpha, df)
cat("Critical value: ",critical_value)

```
```

```
        Pearson's Chi-squared test

data:  Chisq_table
X-squared = 13.676, df = 6, p-value = 0.03347

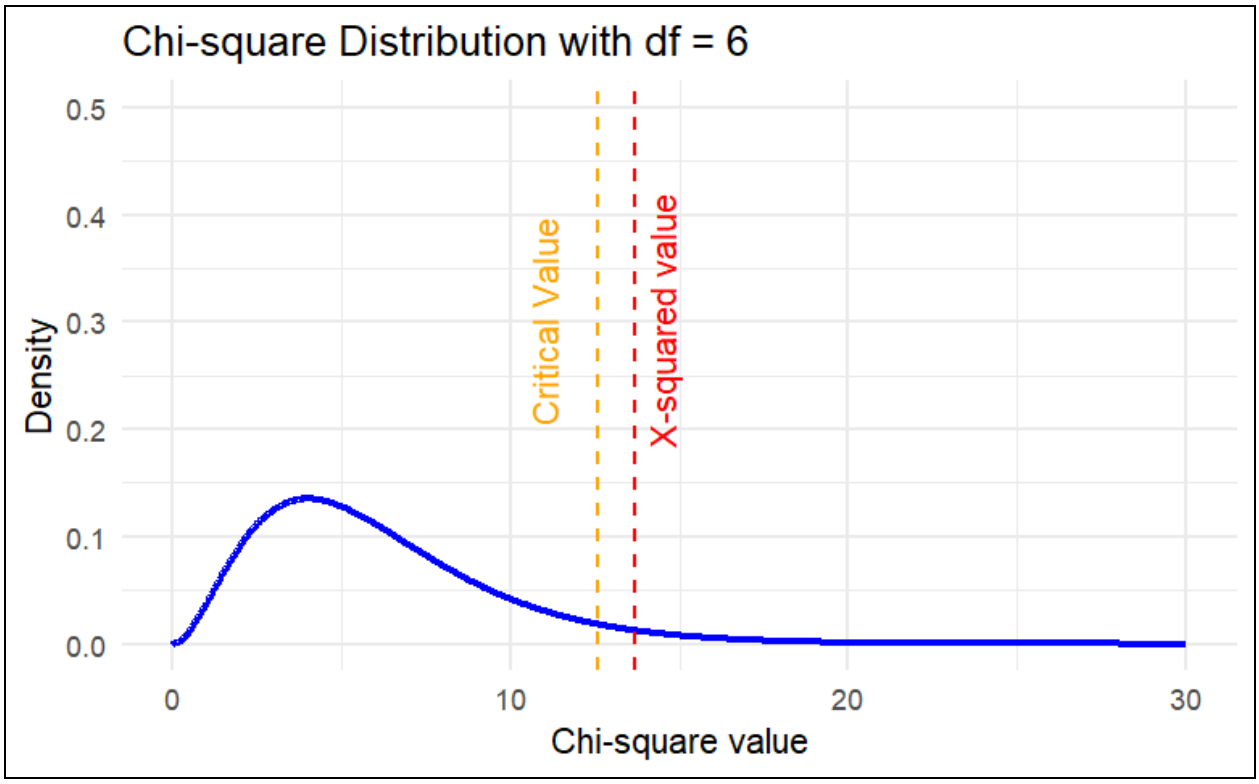
Critical value:  12.59159
```

The Chi-squared test was performed using the inbuilt function of R Studio. Degrees of freedom are calculated by the formula :

$$\text{Degree of freedom} = (R - 1) \times (C - 1)$$

Here, R = number of rows in Chi-square contingency table and C = number of columns in Chi-square contingency table.

Below is the chi-square distribution plotted in RStudio with clear indication of the critical value and X-squared lines for 6 degrees of freedom.




The critical value is used to calculate the p-value, which is the probability of observing a test statistic as extreme as the one calculated from the sample, assuming the null hypothesis is true. Since the calculated p-value is less than the chosen significance level (0.05), the null hypothesis (H0) gets rejected and it is concluded that there is significant evidence to support the alternative hypothesis (H1).

The X-squared value obtained from the test measures the degree of association between the categorical variables. Since the X-squared value is greater than the critical value the association between variables is strong.

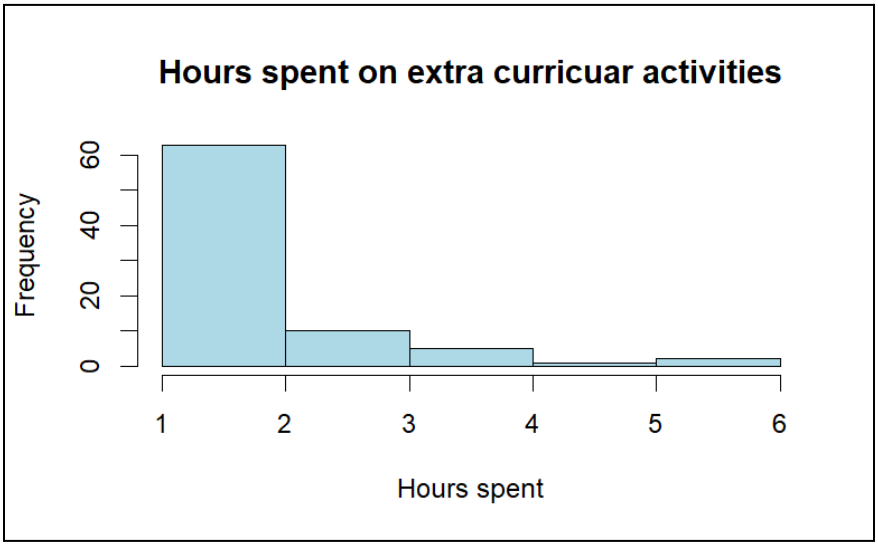
**Decision:** We reject the null hypothesis (H0), therefore there is indeed a relation between Type of extracurricular activity and Academic performance of a college student. The result is significant since p-value is less than 0.05.

## 2. Test for correlation with Fisher's z-transformation

We use the z-test to test the significance of Pearson's correlation coefficient. In this case, the distribution of the Pearson correlation coefficient can be approximated by a normal distribution using Fisher's z-transformation.

***H0:*** *There is no significant correlation between the CGPA and hours spent on extracurricular activity of a student.*

***H1:*** *There is a significant correlation between the CGPA and Hours spent on extracurricular activity of a student.*




We calculated the z-statistic using the Fisher's transformation formula $z = 0.5 * \log((1+r)/(1-r))$, where r is the Pearson correlation coefficient. The critical value is obtained using the q-norm function, which calculates the z-score corresponding to the given probability level. Given below is the output in R-code.

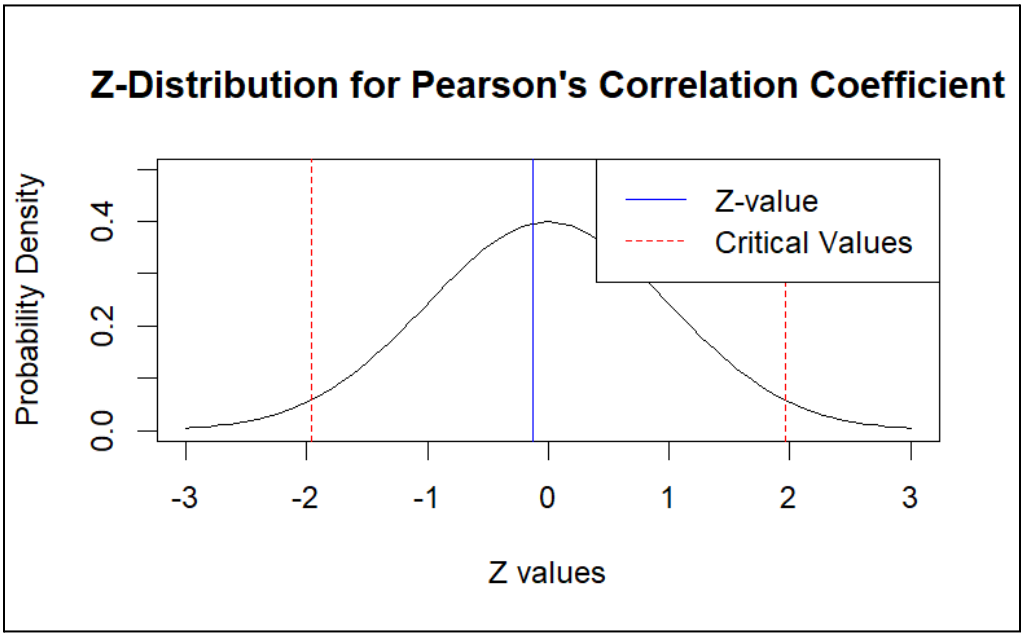
```r
cor_res <- cor.test(data$CGPA, data$Hours)
r <- cor_res$estimate
p_value <- cor_res$p.value

# Calculate z-statistic and critical value
z <- 0.5 * log((1+r)/(1-r))
critical_value <- qnorm(1 - alpha/2)

# Test the null hypothesis
if (abs(z) > critical_value) {
  print("Reject null hypothesis")
} else {
  print("Accept null hypothesis")
  print(paste("P-value: ", p_value))
}
cat("P-value: ", p_value,"\n")
cat("z-Critical value: ",critical_value,"\n")
cat("z-Statistic value: ",z)

```

```
[1] "Accept null hypothesis"
[1] "P-value:  0.245257140017728"
P-value:  0.2452571
z-Critical value:  1.959964
z-Statistic value:  -0.131331
```



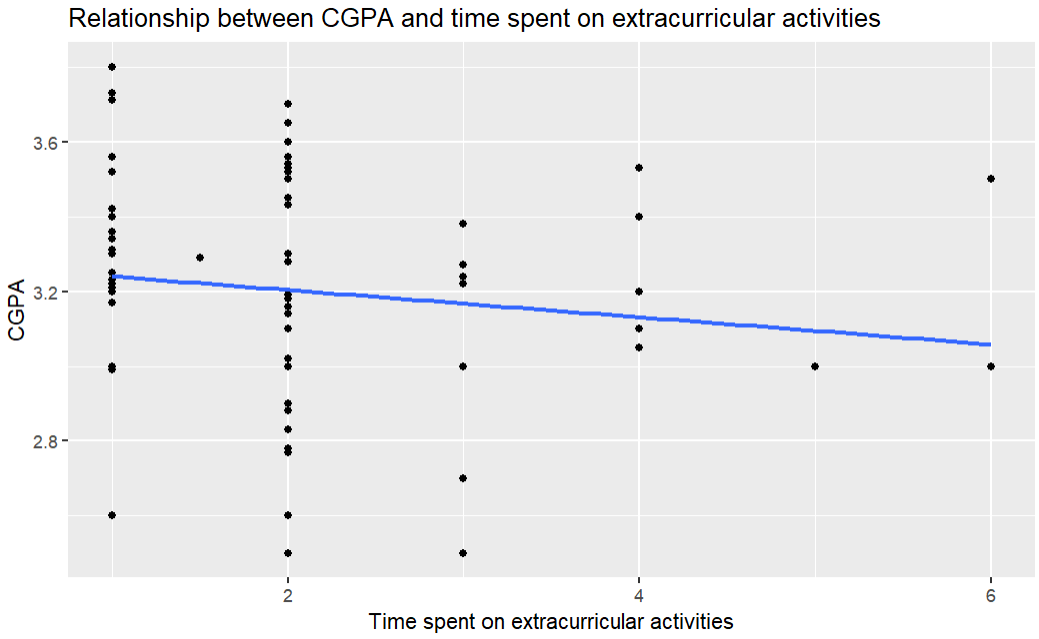

Since the absolute value of z-statistic is less than z-critical value we accept the null hypothesis. Even the p-value is greater than the significance level (0.05).

**Decision:** We accept the null hypothesis (H0), therefore there is no correlation between the CGPA and hours spent on extracurricular activity of a college student.

## 3. Linear Regression analysis

Regression analysis is a statistical method used to study the relationship between a dependent variable and one or more independent variables. The goal of regression analysis is to build a mathematical model that can predict the value of the dependent variable based on the values of the independent variables.

For the regression plot, CGPA variable is the response variable and Hours is the predictor variable. Given below is the scatter plot of 81 data points and a regression line is fitted on it.




It is a simple and effective way to visualize the relationship between the two variables using linear regression analysis in R.

## Conclusion

Based on the above results, we can infer that academic performance of the students does not depend on the hours spent by them on extra curricular activities, but it surely depends on the type of activity they are involved in. Chi-square is a good test for categorical variables and hypothesis testing. Linear Regression analysis is not useful in capturing complex relationships. Fisher's z-transformation is a good analysis for hypothesis testing. Deciding an accurate significance level plays the most important role for testing hypotheses.

## REFERENCES

[1] Stephanie. (2020, September 21). *Fisher Z-Transformation - Statistics How To*. Statistics How To. https://www.statisticshowto.com/fisher-z/

[2] *The Chi squared tests | The BMJ*. (2021, April 12). The BMJ | the BMJ: Leading General Medical Journal. Research. Education. Comment. https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one/8-chi-squared-tests

## CODE AND DATASET LINK

https://github.com/ushmay4802/MAT502-Project.git