TRACKING THE PROGRESS:AN ANALYSIS OF STUDENTS OF SIR SYED COLLEGE

Submitted to Kannur University

In the partial fulfilment for the award of the degree of

BACHELOR OF SCIENCE IN STATISTICS

Submitted by Name: .....MUHAMMED NADEEM K.....

Register number: .....SS21CSTR01.....

DEPARTMENT OF STATISTICS

SIR SYED COLLEGE, TALIPARAMBA

2021-24

CERTIFICATE

This is to certify that Mr/Ms.....MUHAMMED NADEEM K....., Register No .....SS21CSTR01....., Department of Statistics, Sir Syed College Taliparamba has done the project entitled **"Tracking the Progress: An Analysis of students of Sir Syed college**" Submitted to Kannur University for the partial fulfilment for the award of Degree of Bachelor of Science in statistics during the Academic year 2021-24.

Supervising Teachers: Head of the Department

1. Mansoor N K
2. Fathimathul Nithasha Beegam I
3. Thanseeha Nargees

External examiner:

ACKNOWLEDGEMENT

I am extremely grateful to Dr. Ismail Olayikkara, Principal, Sir Syed College Taliparamba for providing the necessary facilities.

I express my thankfulness and gratitude to Mr. Mansoor N K, Head of the Department of Statistics for her valuable advice and support.

It is my privilege to express my deep sense of thankfulness to supervising teachers Mr. Mansoor N K , Ms. Rizwana K , Ms.Fathimathul Nithasha Beegam I . for their valuable suggestions and guidance. And I deeply indebted to other faculties of Department of Statistics, Sir Syed College Taliparamba.

I honestly thank all the handloom workers in Kannur district for their cooperation and encouraging support for conducting the survey. Above all, I owe my gratitude to the almighty for showering abundant blessing upon me and I express my gratitude to my beloved parents and friends.

Taliparamba ,

CONTENTS

INTRODUCTION

Understanding the factors that influence student performance is crucial for educational institutions. This project investigated significant factors impacting college student performance using statistical techniques. Data was collected from Sir Syed college Taliparamba. The study aimed to identify the factors affecting student performance by analysing variables such as socio-economic factors, extra-curricular activities

We are taking a closer look at the lives of college students using data analysis. Our goal is to understand what factors affect their success and well-being. By looking at things like study habits, extracurricular activities, and backgrounds, we hope to find useful insights. These insights can help schools make better decisions about how to support their students and provide resources.

We are not just looking at the big picture; we want to understand the details of each student's experience. Using data analysis, we will find trends and patterns that show what college life is really like. Our aim is to give college suggestions on how to help their students in different ways.

Our project is all about contributing to the conversation on how students succeed. By combining education and data science, we want to show how colleges can create an environment that helps students do well both academically and personally.

AIM

To identify key factors that significantly influence the academic performance of 3rd year students of Sir Syed college.

OBJECTIVES

1) Profile students based on their gender, age, residential arrangements.
2) To examine if there any significant difference between CGPA of BA and BSC students.
3) To examine the association between category and CGPA of students.
4) To examine the relationship between average study time of students and CGPA.
5) To examine if there is any significant difference between CGPA of students and department.
6) To examine the utilisation of college resources by students.
7) To examine whether back papers, extracurricular activities, stress, have statistically significant relationships with course satisfaction.
8) To examine the association between plans after graduation and CGPA of students.
9) To examine the difference in CGPA based on category and hosteller/day scholar of students.
10) To examine if there is any significant difference between CGPA and gender of students.
11) To examine the relationship between CGPA and extracurricular activities of students.
12) To examine if there is any significant difference between CGPA and hosteller/day scholar

13) To examine the relationship between average mobile phone usage and CGPA.

MATERIALS AND METHOD

- ## Data and Methodology:

We analysed data on academic performance of 3rd year students in Sir Syed college from various sources such as surveys and institutional records. We identified the factors that influence academic performance, including study habit, extracurricular activities, demographics, and residential arrangements. We used quantitative research method, providing insights for improving academic performance of students in college.

- ## Study period:

Actual survey was started on December and completed on March 2023

- ## Schedule:

The academic performance of 3rd year students. This information was used to track and analyse the survey questionnaire covered various aspects related to student academic performance from Sir Syed college, including demographic information, residential arrangements, reasons for dropout, impact of study habits on CGPA, and challenges faced during academic year.

DATA EXPLORATION

For the data analysis, we used the following options in R

1. Univariate analysis

- **Pie diagram:**

A circle is constructed and then is sliced up into distinct sectors, one for each different data values. The area of each sector is meant to represent the relative frequency of the values.

R COMMAND: pie ()

- **Bar Diagram:**

Bar chart is popular type of graph used to display a frequency distribution for nominal or ordinal data. In a bar chart the various categories in which the observations fall are represented along a horizontal axis. A vertical bar is drawn above each category within that class. A bar should be equal width and separated from one another so as not to imply continuity.

R COMMAND: barplot()

2. Bivariate analysis

- **Cross tabulation and chi-square test:**

The cross-tabulation approach is especially used when the data are in the nominal form. Cross tabs generate contingency from nominal or ordinal data under which we classify each variable in these categories. Chi-square yields the linear-by-linear association test.

R COMMAND: chisq.test()

- **chi-squared Test**

The chi-squared test is a statistical test used to determine whether there is a significant association between two categorical variables. It is used to test the null hypothesis that there is no association between the variables.

R command: chisq.test()

- **Binary logistic regression**

Binary logistic regression is a statistical method used to model the relationship between a binary dependent variable (also known as the outcome or response variable) and one or more independent variables (also known as predictor variables or covariates). The dependent variable can only have two possible outcomes, typically coded as 0 or 1.

The logistic regression equation can be expressed as follows:

$\text{logit}(p) = \beta 0 + \beta 1X1 + \beta 2X2 + ... + \beta kXk$

R COMMAND: glm()

- **Regression fitting**

Regression fitting refers to the process of finding the best-fitting line or curve tha describes the relationship between two or more variables. The objective is to find a mathematical equation that can predict the value of one variable (dependent variable) based on the values of one or more other variables (independent variables).

R COMMAND: lm()

lm stands for linear model

- **Wilcoxon test**

A Wilcoxon test is a non-parametric statistical hypothesis test used to compare two related samples. It's particularly useful when the data are not normally distributed or when the assumptions of a parametric test like the paired t-test are not met. The test assesses whether the median of the differences between paired observations differs significantly from zero.

R COMMAND: wilcox.test()

- **Kruskal-Wallis test**

Kruskal-Wallis test is a non-parametric statistical test used to compare the median ranks of two or more independent groups. The test assesses whether the samples originate from the same population or from populations with the same distribution. It's often used when the assumptions of parametric tests like the ANOVA are violated, particularly when the data are not normally distributed or when there are outliers.

R COMMAND: kruskal.test()

- **Spearman's rank correlation coefficient**

Spearman correlation is a non-parametric measure of correlation between two variables. It assesses the strength and direction of association between variables, particularly when the relationship is not linear or when the data are ranked rather than continuous or non-normal. Instead of measuring Pearson correlation, Spearman correlation assesses how well the relationship between variables can be described by a monotonic function.

R COMMAND: cor.test(x,y,method="spearman")

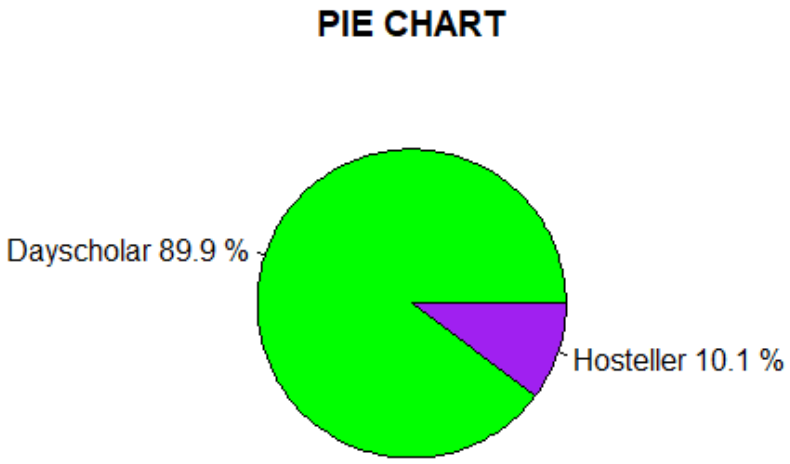
ANALYSIS OF DATA

1) Profile students based on their gender, age and residential arrangements.

a) **Pie chart of proportion of hostellers and day scholars**

- **Rcode & Output**

```
values=c(248,28)`

labels=c("Dayscholar","Hosteller")

percentage=round(100*values/sum(values),1)

labelswithpercentage=paste(labels,percentage,"%")

pie(main="PIE CHART",col=c("green","purple"),values,labels=labelswithpercentage)
```



Conclusion:

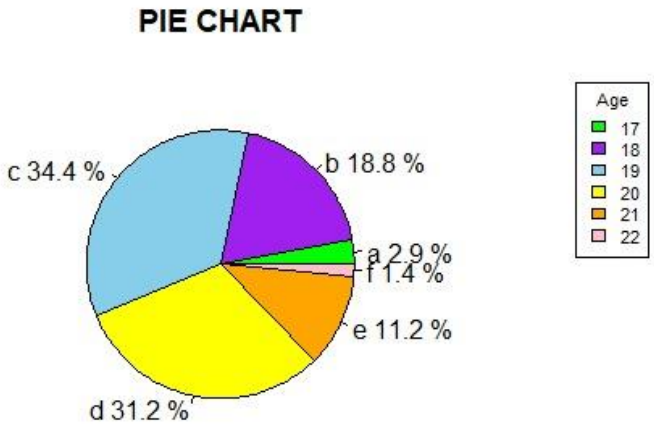
hostellers and day scholars are in the proportion of 9:1

b) **Pie chart of age distribution**

a. **Rcode & Output**

```
data1$Age
```

```
stay_count=table(data1$Age)

print(stay_count)

values=c(8,52,95,86,31,4)

percentage=round(100*values/sum(values),1)

labels=c("a","b","c","d","e","f")

labelswithpercentage=paste(labels,percentage,"%")

pie(main="PIE CHART",col=c("green", "purple", "skyblue", "yellow",
"orange","pink","blue"),values,labels=labelswithpercentage)

my_colors=c("green", "purple", "skyblue", "yellow", "orange","pink","blue")

legend("topright",names(stay_count), cex = 0.7, fill = my_colors,title = "Age")
```

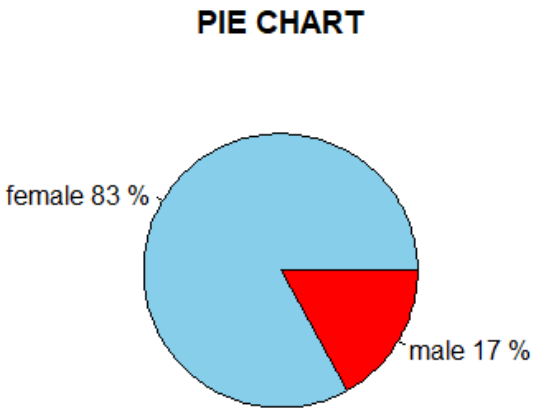


c) **Pie chart of ratio of male and female**

a. <u>Rcode and output</u>

```
values=c(229,47)

labels=c("female","male")

percentage=round(100*values/sum(values),1)
```

```
labelswithpercentage=paste(labels,percentage,"%")
pie(main="PIE CHART",col=c("skyblue","red"),values,labels=labelswithpercentage)
```



Conclusion:

Female and male are in the ratio of 83:17

2) To examine if there are any significant differences between CGPA of BA and BSC students.

H_0 **: There is no significant difference between CGPA of BA and BSC students**

- **Rcode & Output**

```
shapiro.test(BSC)

shapiro.test(BA)

wilcox.test(BA,BSC)
```

```
        Shapiro-Wilk normality test

data:  BSC
p-value < 2.2e-16
```

```
        Shapiro-Wilk normality test

data:  BA
p-value = 2.114e-12
```

```
        Wilcoxon rank sum test with continuity correction

data:  BA and BSC
p-value = 1.21e-05
```

Conclusion

Here p value is less than 0.05 . Hence we reject H_0

There is significant difference between CGPA of BA and BSC students

3) To examine the association between category and CGPA of students.

H_0 : There is no association between category and CGPA

- **Rcode & Output**

```
x=matrix(c(25,38,28,7,17,32,73,16,3,5,8,4,4,15,1,0),nrow = 4,ncol = 4)
rownames(x)=c("below 5","5-7.5","7.5-9","above 9")
colnames(x)=c("BA","BSC","BCOM","BMMC")
chisq.test(x)
```

```
        BA BSC BCOM BMMC
below 5 25  17    3    4
5-7.5   38  32    5   15
7.5-9   28  73    8    1
above 9  7  16    4    0


Pearson's Chi-squared test
data:  x
X-squared = 42.405, df = 9, p-value = 2.772e-06
```

Conclusion

Here p value is less than 0.05 . Hence we reject H_0

There is association between category and CGPA

4) To examine the relationship between average study time of students and CGPA.

H_0 **: There is no relation between average study time of students and CGPA**

i. CORRELATION

- **Rcode & Output**

```
x=data1$`Avg Studying time (in hrs)`
y=data1$CGPA
shapiro.test(y)
correlation_coefficient=cor.test(x,y,method = 'spearman')
print(correlation_coefficient)
```

```
	Shapiro-Wilk normality test

data:  y
p-value < 2.2e-16


	Spearman's rank correlation rho

data:  x and y
S = 2431632, p-value = 2.144e-07
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.3060511
```

Conclusion

The data are not normally distributed .

There exist a less correlation between average study time of students and

ii. SIMPLE LINEAR REGRESSION

- **Rcode & Output**

```
lm_model <- lm(academic_success ~ study_hours, data = data)
summary(lm_model)
```

```
Call:
lm(formula = academic_success ~ study_hours, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0544 -0.2619  0.9556  1.8164  4.5268

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2128     0.4236   9.944  < 2e-16 ***
study_hours   1.9208     0.3410   5.633 4.38e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.92 on 274 degrees of freedom
Multiple R-squared:  0.1038,     Adjusted R-squared:  0.1005
F-statistic: 31.73 on 1 and 274 DF,  p-value: 4.383e-08
```

Conclusion

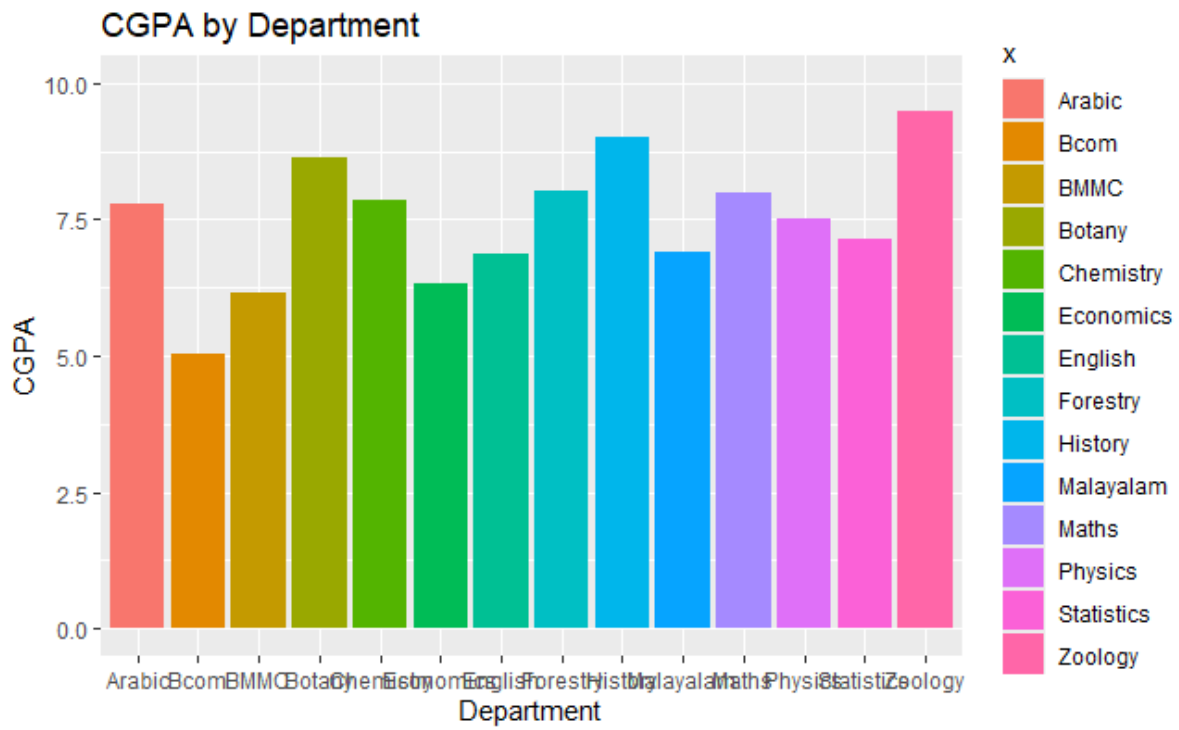
The estimated intercept is approx. 4.21 and estimated coefficient for study_hour is approx. 1.92.For each additional hour of study, we expect an average increase of 1.92 units in academic_success. Since p-value is less than 0.05 , the model as a whole is statistically significant.

5)To examine if there is any significant difference between CGPA of students and department.

i) Comparison by using Bar Diagram

- **Rcode & Output**

```
original_data=data.frame(x,y)
ggplot(original_data, aes(x,y, fill = x))+
geom_bar(stat="identity")+
labs(title = "CGPA by Department",
x = "Department", y = "CGPA") +
scale_y_continuous(limits = c(0,10,1))
theme_minimal()
```



Conclusion

Zoology stands out with the highest CGPA, represented by the pink bar.

Bcom stands out with the lowest CGPA, represented by the orange bar.

ii) KRUSKAL-WALLIS TEST

H_0 **: There is no significant difference between CGPA of students and department.**

```
shapiro.test(x1)
model=kruskal.test(a~b)
```

```
Shapiro-Wilk normality test

data:  x1
W = 0.71592, p-value = 8.658e-05
```

```
        Kruskal-Wallis rank sum test

data:  a by b
Kruskal-Wallis chi-squared = 43.458, df = 13, p-value = 3.773e-05
```

Conclusion

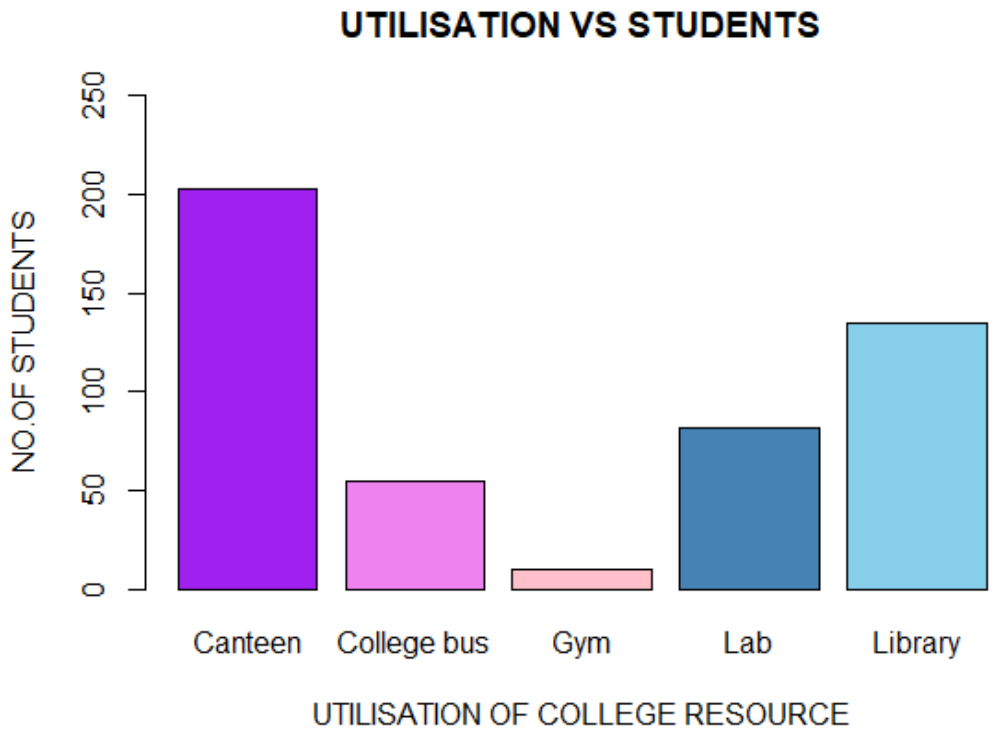
Here p value is less than 0.05 . Hence we reject H_0

6) To examine the utilisation of college resources by students.

- **Rcode & Output**

```
barplot(data_count, xlab = "UTILISATION OF COLLEGE RESOURCE", ylab =
"NO.OF STUDENTS", main = "UTILISATION VS STUDENTS",
    col=c("purple","violet","pink","steelblue","skyblue"),ylim=c(0,250),)
```

```
Canteen       College bus       Gym         Lab      Library
    203            55             10          82       135
```



Conclusion

Canteen is utilised the most by the students and gym is the least utilised.

7) To examine whether back papers, extracurricular activities, stress, have statistically significant relationships with course satisfaction.

H_0 **: There is no statistically significant relationship between the predictors**

- **Rcode & Output**

```
z=data3$`Are you satisfied with your course`

x1=data3$`Have any back papers`

y1=data3$`Participated in any extra curricular activities conducted in our college`
```

```
z1=data3$`Do you feel anxious/stressed during exam days`
data=data.frame(x1,y1,z1,z)
logistic_model=glm(z~y1+x1+z1,data=data,family = binomial)
print(logistic_model)
summary(logistic_model)
```

```
Call:  glm(formula = z ~ y1 + x1 + z1, family = binomial, data =
data)

Coefficients:
(Intercept)           y1           x1           z1
     2.5612      -0.3814      -0.4886      -0.4321

Degrees of Freedom: 275 Total (i.e. Null);  272 Residual
Null Deviance:        198
Residual Deviance: 195.3      AIC: 203.3

Call:
glm(formula = z ~ y1 + x1 + z1, family = binomial, data = data)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.5612     0.4356   5.879 4.12e-09 ***
y1           -0.3814     0.4127  -0.924    0.355
x1           -0.4886     0.4450  -1.098    0.272
z1           -0.4321     0.4517  -0.956    0.339
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 198.04  on 275  degrees of freedom
Residual deviance: 195.27  on 272  degrees of freedom
AIC: 203.27

Number of Fisher Scoring iterations: 5
```

Conclusion

- The coefficient estimate for back papers is approximately **1.13**.The p-value (**0.00859**) indicates that this predictor affect course satisfaction.
- The coefficient estimate for extracurricular activity is also approximately **1.13.**The p-value (**0.00859**) indicates that this predictor affect course satisfaction.
- The coefficient estimate for stress is very close to zero. The p-value is high (**1.00000**), indicating that this predictor does not significantly affect course satisfaction.

Hence we can conclude that,
Back papers and participation in extracurricular activities positively impact course satisfaction. Feelings of stress during exam days do not significantly affect course satisfaction.

8) To examine the association between plans after graduation and CGPA of students.

H_0 **: There is no association between plans after graduation and CGPA of students.**

- **Rcode & Output**

```
x=matrix(c(12,13,19,4,34,73,90,21,1,0,1,0,1,4,0,0,1,2,1,0),nrow = 4,ncol = 5)

rownames(x)=c("below 5","5-7.5","7.5-9","above 9")

colnames(x)=c("Employment","Higher studies","Not decided","Marriage","Others")

chisq.test(x)
```

```
         Employment Higher studies Not decided Marriage Others
below 5          12             34           1        1      1
5-7.5            13             73           0        4      2
7.5-9            19             90           1        0      1
above 9           4             21           0        0      0

Warning: Chi-squared approximation may be incorrect

        Pearson's Chi-squared test

data:  x
X-squared = 11.623, df = 12, p-value = 0.4764
```

Conclusion

Here p value is greater than 0.05 . Hence we accept H_0

There is no association between plans after graduation and CGPA of students.

9) To examine the difference in CGPA based on category and hosteller/day scholar of students.

H01**: There is no interaction between CGPA and category**

H02**: there is no interaction between CGPA and hosteller and dayscholar**

H03 : **there is no combined effect of category and hosteller/dayscholar on CGPA**

- **Rcode & Output**

```
a=data1$CGPA
b=data1$Category
c=data1$`Hosteller/Day scholar`
my_data=data.frame(a,b,c)
model=aov(a~b*c,data = my_data)
summary(model)
```

```
             Df Sum Sq Mean Sq F value  Pr(>F)
b             3  127.4   42.48   4.852 0.00264 **
c             1    1.6    1.58   0.181 0.67097
b:c           3  132.0   44.01   5.026 0.00209 **
Residuals   268 2346.4    8.76
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Conclusion

Here p value is less than 0.05 . Hence we reject H**01**

Hence there is interaction between CGPA and category

Here p value is greater than 0.05 . Hence we accept H**02**

Hence there is no interaction between CGPA and hosteller and dayscholar

Here p value is less than 0.05 . Hence we reject H**03**

Hence there is a combined effect of category and hosteller/dayscholar on CGPA

10) To examine if there is any significant difference between CGPA and gender of students

H_0 : There is no significance difference between CGPA and gender of students

- **Rcode & Output**

```
x=data3$Gender

y=data1$CGPA

data_a=data.frame(x,y)

shapiro.test(x)

wilcox.test(x,y)
```

```
Shapiro-Wilk normality test


data:  x

W = 0.45397, p-value < 2.2e-16


Wilcoxon rank sum test with continuity correction


data:  x and y

W = 7913.5, p-value < 2.2e-16

alternative hypothesis: true location shift is not equal to 0
```

Conclusion

The given data is not normally distributed. Hence we use non-parametric Wilcoxon test. Here p value is less than 0.05 . Hence we reject H_0

There is significant difference between CGPA and gender of students

11) To examine the relationship between CGPA and extracurricular activities of students

H_0**: There is no relationship between CGPA** and extracurricular activities of students

- **Rcode & Output**

```
x=data1$CGPA
y=data3$`Participated in any extracurricular activities conducted in our college`
shapiro.test(x)
cor.test(x,y,method = 'spearman')
```

```
        Shapiro-Wilk normality test

data:  x
W = 0.7007, p-value < 2.2e-16


data:  x and y
S = 3083340, p-value = 0.04628
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1200639
```

Conclusion

The calculated Spearman's rho for the given data is approximately 0. 1201.This value indicates a very weak positive correlation between the variables. Here p value is less than 0.05. Hence, we reject H_0. There is relationship between CGPA and extracurricular activities of students

12.To examine if there is any significant difference between CGPA and hosteller/day scholar of students

H_0 **:There is no significance difference** between CGPA and hosteller/day scholar of students

- **Rcode & Output**

```
hosteller=data3$Hosteller

dayscholar=data3$Dayscholar

shapiro.test(hosteller)

wilcox.test(hosteller,dayscholar)
```

```
Shapiro-Wilk normality test


data:  hosteller

W = 0.6588, p-value = 7.846e-07


Wilcoxon rank sum test with continuity correction


data:  hosteller and dayscholar

W = 3388.5, p-value = 0.8353

alternative hypothesis: true location shift is not equal to 0
```

Conclusion

Here data is not normally distributed. Hence, we use Wilcoxon test. Since p value is greater than 0.05, we accept H_0. There is no significance difference between CGPA and hosteller/day scholar of students

13)To examine the relationship between average mobile phone usage and CGPA.

Ho: There is no relation between mobile phone usage and CGPA

- **Rcode & Output**

```
x=data3$Avg mobile phone usage
y=data3$CGPA 1
shapiro.test(x)
correlation_coefficient=cor.test(x,y,method = 'spearman')
print(correlation_coefficient)
Spearman's rank correlation rho

data:  x and y
S = 5880006, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
     rho
-0.67806
```

Conclusion

rho is approx. -0.678. So there is a strong negative correlation between mobile phone usage and CGPA.

CGPA and mobile phone usage are inversely proportional

CONCLUSION

- There is a significant difference between CGPA of BA and BSc students
- There is a association between category of departments and CGPA of students
- There is a relation between average study time and CGPA of students
- There is significant difference between CGPA and department of students
- There is a relation between average mobile phone usage and CGPA of students
- There is no association between plans after graduation and CGPA of students
- There is interaction between CGPA and category
- There is no interaction between CGPA and hosteller and day scholar
- There is a combined effect of category and hosteller/day scholar on CGPA
- There is significant difference between CGPA and gender of students
- There is relationship between CGPA and extracurricular activities of students

FINDINGS

- BA and BSc students exhibit distinct CGPA patterns.
- The study found that students' CGPA varies significantly across different academic departments.
- Time spent studying significantly influences students' CGPA.
- Students who used smartphones had lower academic performance (CGPA)
- The relationship between CGPA and category is not random chance; there is evidence of an association.
- This lack of interaction suggests that the effect of being a hosteller or day scholar does not significantly impact CGPA.
- Considering both category and hosteller/day scholar status together, they jointly influence CGPA.

REFERENCES

1. Introduction To Linear Regression Analysis, D C. Montgomery,E A Peak And G G Vining , Wiley 2003.

2. Fundamentals Of Mathematical Statisics-By S. C. Gupta, V. K. Kapoor

3. Naked Statistics: Stripping The Dread From The Data. 1st Edition

4. Fundamentals Of Applied Statistics. - By S. C. Gupta, V. K. Kapoor

5. The Visual Display Of Quantitative Information, 2nd Ed. 2nd Edition By Edward Tufte

6. Sudha G.Purohit,Sharad D.Gore And Shailaja R.Deshmukh Statistics Using R

7. An Introduction To Statistical Learning With Applications In R.- Gareth James, Daniela Witten, Trevor Hastie, And Robert Tibshirani (Springer Texts In Statistics) . 2017 Edition

8. Malayali Statistician (Youtube channel)

9. Introduction to Linear Regression analysis, D C . Montgomery, E A Peak and G G Vining , Wiley

10. Applied Logistic Regression :David W. Hosmer, Stanley Lemeshow

11. An Introduction to Categorical Data Analysis, 3rd Edition Alan Agresti.

12. Statistics and Probability with Applications for Engineers and Scientists Using MINITAB, R and JMP. Wiley.

13. Statistical Analysis with Missing Data Roderick J. A. Little (Author), Donald B. Rubin

APPENDIX

SCHEDULE

Tracking the progress: an analysis of students of Sir Syed college

1) Name
2) Age
3) Gender
4) Rural/Urban
5) Department
    - Maths
    - Statistics
    - Physics
    - Chemistry
    - Botany
    - Zoology
    - Forestry
    - Arabic
    - Economics
    - English
    - History
    - Malayalam
    - BMMC
    - BCOM
6) Hosteller or day scholar
7) Hours of study per day
8) CGPA
9) Supply in previous paper
    i. Yes
    ii. No
10) Total number of back papers
11) Satisfaction in course
    i. Yes
    ii. No
12) Challenges faced in course
    i. Challenging course and high expectation
    ii. Balancing multiple assignments exams and projects
    iii. Time management
    iv. College expenses
    v. Anxiety and depression
    vi. Social media
13) Plan after graduation
    i. Higher studies
    ii. Employment
    iii. Marriage
    iv. Other

14) Use of college resources
    i. College bus
    ii. Gym
    iii. Lab
    iv. Library
    v. Canteen
15) Participation in extracurricular activities
    i. Yes
    ii. No
16) Average mobile time usage in a day
17) Areas required to improve
    i. Time management
    ii. Study habits
    iii. Communication skills
    iv. Stress management