A STUDY OF DIFFERENT FACTORS AFFECTING STUDENTS' RESULTS

A PROJECT WORK SUBMITTED TO THE

DEPARTMENT OF STATISTICS
MAULANA AZAD COLLEGE
Affiliated to the
UNIVERSITY OF CALCUTTA

BY

Name : Santarpan Pal
Semester : VI
CU Regn. No. : 146-1111-0507-21
CU Roll No. : 213146-21-0042

D E C L A R A T I O N

I, **Santarpan Pal**, a student of Semester 6, entitled to the programme B.Sc. Statistics Hons. at Maulana Azad College, University of Calcutta with **Regn. No. 146-1111-0507-21** and **Roll No. 213146-21-0042**, solemnly declare that the project entitled "**A study of different factors affecting students' results**" is a genuine and original work undertaken by me under the supervision of **Prof. Anup Kumar Giri** as a part of my Bachelor's programme.

I further affirm that:

- All the data collected for this project is authentic and sourced from reliable and verifiable sources.
- The analysis, interpretations, and conclusions presented in this project are the results of my own research and statistical analysis.
- This project has not been submitted for any other academic degree or assessment at any institution.
- I understand the academic integrity requirements of my university, and I have adhered to all guidelines and regulations regarding plagiarism and research ethics.
- In case of any unintentional oversight or error, I take full responsibility for the same and will cooperate with the necessary revisions or corrections as advised by my faculty or supervisor.
- I am aware of the consequences of academic misconduct, and I affirm that this project is entirely free from any act of dishonesty or deception.

I am sincerely committed to the successful completion of this project.

Signature: _____

Date: _____

Place: Kolkata, India

_____

Countersigned by The Supervisor

Date: _____

TABLE OF CONTENTS

1 INTRODUCTION

Scenario: Four years back the whole world stood still due to an unfortunate threat. A pandemic named **Covid** barged into our life. Many lives lost, economies destroyed. Now after four years everything seems normal. But is it really? In the time of pandemic, we adapted few habits; rather say we were forced to adapt those habits due to the circumstances. And now even after four years, when everything seems normal, those **"pandemic time habits"** remained in our lives. Use of smart phone is one of those habits. It is not that people didn't use smart phone before covid, but the use increased from the time of pandemic. Smart phone was just a communication device before pandemic, but now it became a part of our lives. Even the education now very much depends upon mobile phone. It all started from the **"online class era"** of the pandemic. The classes were arranged through various cloud meeting applications; books, notes had been sent to students as pdf files. And this practice remained even after covid is gone. It is not only about education, people who are involved in extracurricular activities or sports also tend to depend on smart phone. On the other hand, in the time of pandemic, people were frustrated by their monotonic lifestyle. So, people started using social media more than they did before. And by the time passed the usage of social media increased more and more; especially the students. And here lies the origin of this project.

Objective: The present generation of students tends to spend time at different things rather than studying. It is readily seen that students use mobile phones for different things. Even students use mobile phones for study purposes also. Is this actually good for students? Or using mobile phone creating a negative impact on students' results? Well, this project aims to analyze this whole situation. Not only impact of usage of mobile phone on students' results, this project deals with different factors which might affect students' results, like involvement in extracurricular activities, involvement in sports, etc. **In one word this project analyzes whether different factors has a significant impact on students' results or not.**

How to reach our goal?: To analyze whether different factors has a significant impact on students' results or not; firstly, we have to collect data from different students who are currently reading in 3rd year of their Bachelor's degree. In order to do that, **a questionnaire should be created and distributed**. After sorting the collected data, we should check whether there is association between different factors and students' results. **We may use Goodman-Kruskal's gamma to check the association**. After that the factors which would have no association should be ignored and we should focus on the factors with association for our further work. Then **we may construct a regression model to predict students' results of their next semester**. Now we should move on to **perform ANOVA between some of the main factors** like study hour spend on mobile, involvement in sports, involvement in extracurricular activities etc. After performing ANOVA **we should test for normality to check if the ANOVA assumptions are valid**.

2 METHODOLOGY

Collection of data: The very first and foremost work of this project is to collect data. But we can't collect data randomly. We should collect data in the following way:

- **Questionnaire:** The very first step for collecting data should be construction and distribution of a questionnaire. The term questionnaire means a list of certain systematically arranged questions pertaining to the subject of enquiry. We have constructed a questionnaire in Google forms and distributed it to the students who are currently studying in 3rd year of their Bachelor's degree through the help of social media. The questionnaire is attached in the "Appendix" of this project.

- **Scrutiny of Data:** After collection of data we have scrutinized them thoroughly to see if they are correct. This is very important, because however sophisticated statistical techniques may be used in analyzing the data, the results will be misleading if the data were erroneous.

Representation of data: It is quite natural that one cannot easily understand the significance of the collected data at a glance. To bring into focus the salient features of the data, it is necessary to present them in a neat systematic form. In our case, most of our data are categorical. So we have used **"pie-chart"** to represent the data. Also we have used **"bar diagram"** to represent our rating-type data.

Goodman-Kruskal's gamma: In statistics, Goodman and Kruskal's gamma is a measure of rank correlation, i.e., the similarity of the orderings of the data when ranked by each of the quantities. It measures the strength of association of the cross tabulated data when both variables are measured at the ordinal level. It makes no adjustment for either table size or ties. Values range from −1 (100% negative association, or perfect inversion) to +1 (100% positive association, or perfect agreement). A value of zero indicates the absence of association.

The estimate of gamma, $\boldsymbol{G}$, depends on two quantities:

- $\boldsymbol{N_c}$, the number of pairs of cases ranked in the same order on both variables (number of concordant pairs),
- $\boldsymbol{N_d}$, the number of pairs of cases ranked in reversed order on both variables (number of reversed pairs),

where "ties" (cases where either of the two variables in the pair are equal) are dropped. Then

$$\boldsymbol{G = \frac{N_c - N_d}{N_c + N_d}}$$

Critical values for the gamma statistic are sometimes found by using an approximation, whereby a transformed value, *t* of the statistic is referred to Student t distribution, where

$$\boldsymbol{T = G\sqrt{\frac{(N_C + N_D)}{n(1 - G^2)}}}$$

Where, n= Total number of observations.

Now, we should perform the Goodman-Kruskal test and then calculate the p-value.
Here, our null hypothesis is $H_0: G = 0$. If **p-value<0.05**,we reject $H_0: G = 0$ **at level 0.05 of significance.**
In our case, it will help us to observe if there is association between different factors and average SGPA.

Regression and Prediction: In statistics, linear regression is a statistical model which estimates the linear relationship between a scalar response and one or more explanatory variables (also known as dependent and independent variables). The case of one explanatory variable is called **simple linear regression**; for more than one, the process is called **multiple linear regression**. In our case we will use multiple linear regression to predict the 5th semester SGPA of students. We have a dataset of categorical data, so first we need to construct dummy variables then we can proceed for regression.

- **Dummy Variables:** In regression analysis, a dummy variable (also known as indicator variable) is one that takes a binary value (0 or 1) to indicate the absence or presence of some categorical effect that may be expected to shift the outcome. Dummy variables are commonly used in regression analysis to represent categorical variables that have more than two levels. In our case, multiple dummy variables would be created to represent each level of the variable, and only one dummy variable would take on a value of 1 for each observation. Dummy variables are useful because they allow us to include categorical variables in our analysis, which would otherwise be difficult to include due to their non-numeric nature. They can also help us to control for confounding factors and improve the validity of our results.
- **Regression:** Given a data set $\{y_i, x_{i1}, \dots, x_{ip}\}, i = 1(1)n$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and the vector of regressors $\mathbf{x}$ is linear. This relationship is modeled through an error variable ε — an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i, i = 1(1)n$$

Now, using this model, we can predict y by least square methods. Suppose, $x_i = [x_1^i, x_2^i, \dots, x_m^i]$ and the model's parameters are $\beta = [\beta_0, \beta_1, \dots, \beta_m]$ then the model's prediction would be

$$y_i \approx \beta_0 + \sum_{i=1}^{m} \beta_j \times x_j^i, i = 1(1)n$$

Analysis of Variance (ANOVA): Analysis of variance (ANOVA) is a collection of statistical models and their associated estimation procedures (such as the "variation" among and between groups) used to analyze the differences among means. ANOVA was developed by the statistician Ronald Fisher. ANOVA is based on the law of total variance, where the observed variance in a particular variable is partitioned into components attributable to different sources of variation. In its simplest form, ANOVA provides a statistical test of whether two or more population means are equal, and therefore generalizes the *t*-test beyond two means. In other words, the ANOVA is used to test the difference between two or more means.

Suppose we have k populations (or groups or classes) P_1, P_2, … , P_k, which are assumed to follow normal distribution with the same variance σ^2, but different means μ_1, μ_2, … , μ_k, i.e. the populations are homoscedastic. Let y_{ij} be the j-th observation of the i-th sample of size n_i, which is drawn from the i-th population P_i. These k populations are the only available populations in which we are interested, and that's why a fixed effects model is appropriate here.

Our model will be,

$$\boldsymbol{y_{ij} = \mu + \alpha_i + e_{ij}}$$

where, $\boldsymbol{\mu}$ = additive constant
$\boldsymbol{\alpha_i}$ =fixed effect due to i-th population
$\boldsymbol{e_{ij}}$ = random error

Our assumption will be, $\boldsymbol{e_{ij}}$**'s are i.i.d.** $\boldsymbol{N(0, \sigma^2)}$ $\forall$ $\boldsymbol{(i, j)}$.

Our hypothesis will be,

$$\boldsymbol{H_0: \alpha_i = 0 \; \forall \; i \;\; vs \; H_1: \alpha_i \neq 0} \text{ for atleast one } \boldsymbol{i}$$

Our Test Statistic will be $\boldsymbol{F = \frac{MSB}{MSE} \sim F_{\alpha; k-1, n-k}}$

We will reject $\boldsymbol{H_0}$ if **observed F > F-value at 5% level of significance.**

Test for normality using Shapiro-Wilk test: Shapiro-Wilks test is a test for normality. The null-hypothesis of this test is that the population is normally distributed. Thus, if the p value is less than the chosen alpha level, then the null hypothesis is rejected and there is evidence that the data tested are not normally distributed. On the other hand, if the p value is greater than the chosen alpha level, then the null hypothesis (that the data came from a normally distributed population) cannot be rejected.

Confusion Matrix: In problem of statistical classification, a confusion matrix, also known as error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one; in unsupervised learning it is usually called a matching matrix. Each row of the matrix represents the instances in an actual class while each column represents the instances in a predicted class, or vice versa – both variants are found in the literature. The diagonal of the matrix therefore represents all instances that are correctly predicted. The name stems from the fact that it makes it easy to see whether the system is confusing two classes (i.e. commonly mislabelling one as another).It is a special kind of contingency table, with two dimensions ("actual" and "predicted"), and identical sets of "classes" in both dimensions (each combination of dimension and class is a variable in the contingency table).

We can find accuracy, precision, specificity and sensitivity through confusion matrix. These measures has some certain formulas, which are shown below

Accuracy: (TP + TN) / (TP + TN + FP + FN)

Sensitivity: TP / (TP + FN)

Precision: TP / (TP + FP)

Specificity: TN / (TN + FP)

Where, TP: True Positive
TN: True Negative
FP: False Positive
FN: False Negative

3 DATA VISUALISATION

Before starting analyzing data, we must visualize the data through some diagrams. It is very important, because- we cannot say anything about the data just by seeing it at a glance. We should choose our right direction of analysis by observing the nature of the diagram.

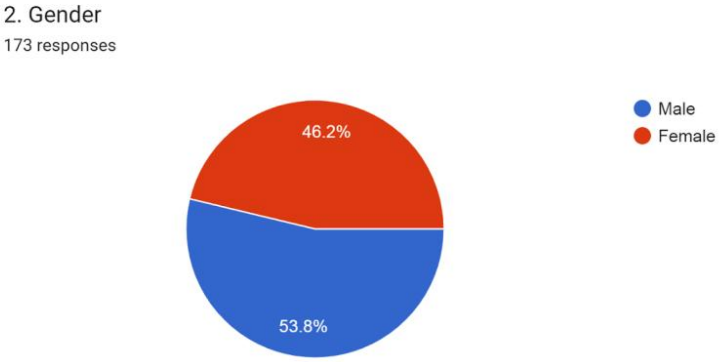


Fig-3.1

Fig-3.1 represents the percentage of male and females from our whole collected data. We can see number of males (53.8%) who have responded is slightly greater than that of the female (46.3%).

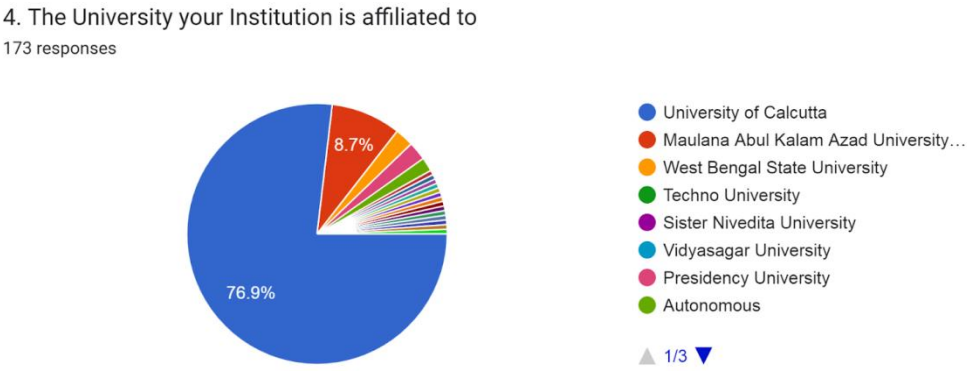


Fig-3.2

Fig-3.2 represents the percentage of respondents from different universities. And by our collected data, maximum of students (76.9%) are from University of Calcutta.

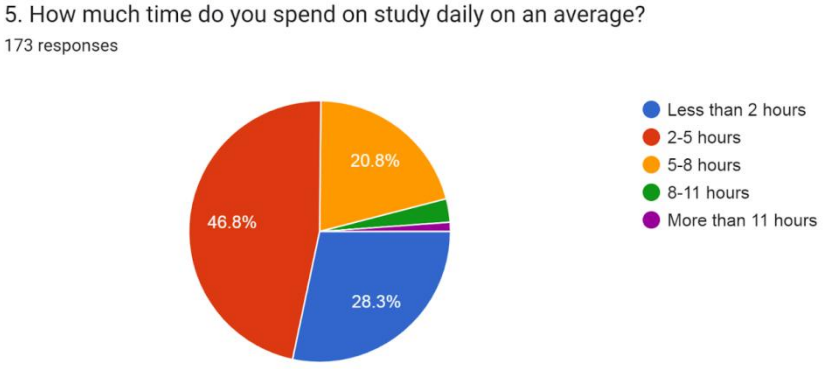


Fig-3.3

Fig-3.3 represents the amount of daily study time of students. And maximum student (46.8%) actually studies 2-5 hours daily. A big portion of students study less than 2 hours daily (28.3%).While 20.8% students study for 5-8 hours daily.

12. How much time do you spend on mobile weekly on an average? (Aggregated total of mobile phone usage)

173 responses

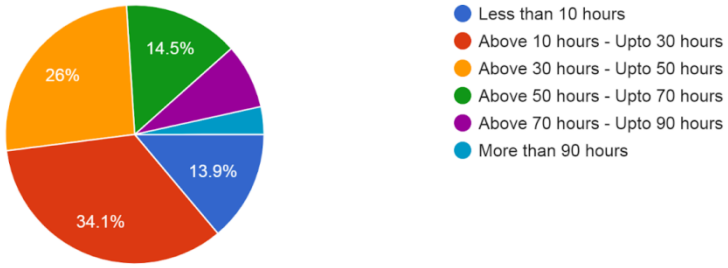


Fig-3.4

Fig-3.4 represents the amount of weekly mobile usage of students. It is observed that 34.1% students use their mobile phone for 10-30 hours weekly. While 26% students use their mobile phone for 30-50 hours weekly. There are a good amount of students who use mobile phone for more than 50 hours

13. How much time do you spend on mobile for study weekly on an average ?

173 responses

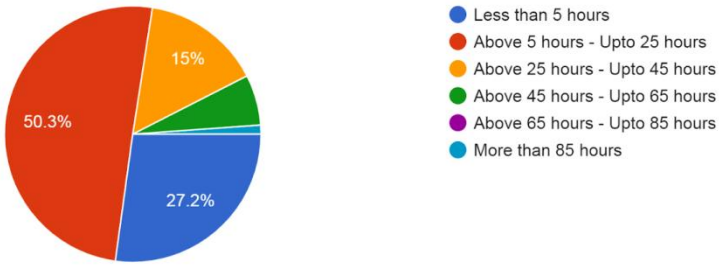


Fig-3.5

Fig-3.5 represents the amount of time spent on mobile for study purposes weekly by students. And half of the population (50.3%) uses their mobile phone for study for 5-25 hours weekly. So it may be said that students actually very much rely on their mobile phone for study purposes.

14. How much time do you spend on mobile for gaming weekly on an average ?

173 responses

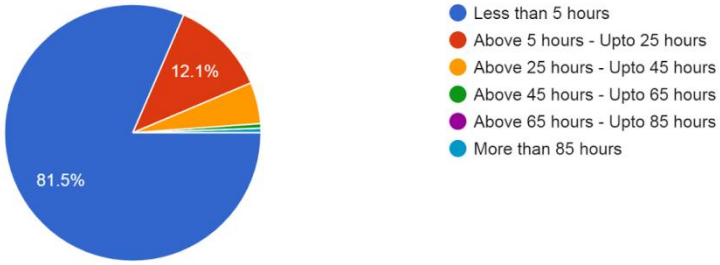


Fig-3.6

Fig-3.6 represents the amount of time students spend on mobile for gaming weekly. A whopping 81.3% of students, i.e. majority of the population spend less than 5 hours on mobile for gaming purposes.

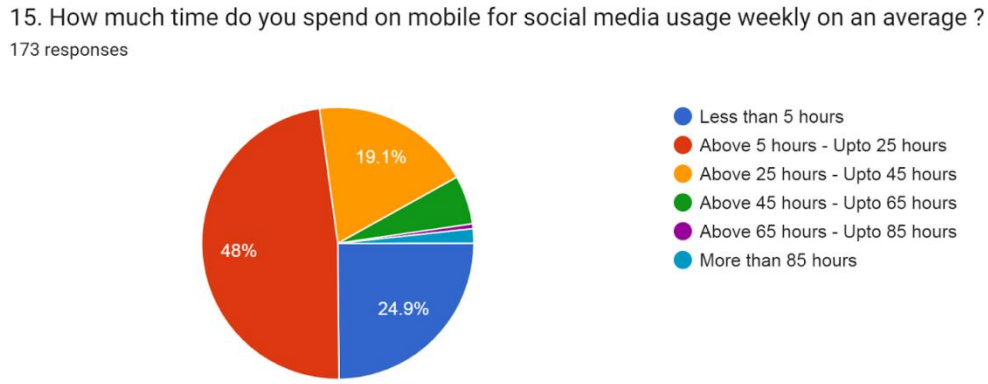


Fig-3.7

Fig-3.7 represents the amount of time students spend on mobile for social media usage weekly. Almost half of the respondents (48%) use it for 5-25 hours weekly. While 24.9% students use social media less than 5 hours weekly.

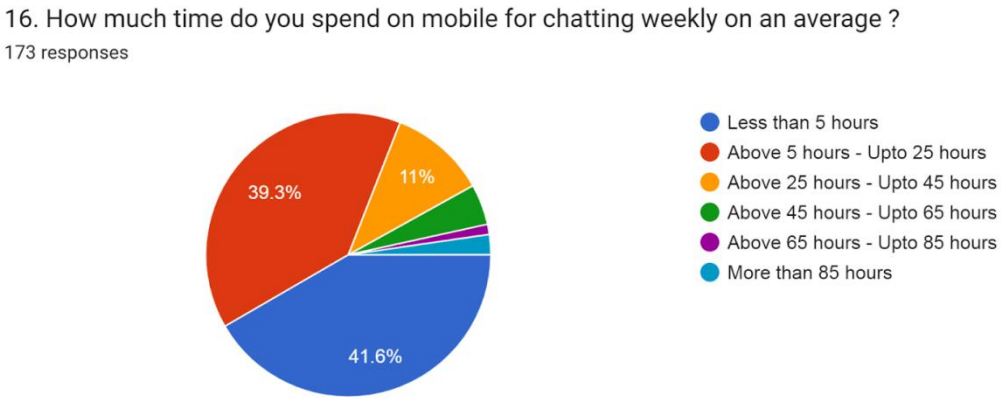


Fig-3.8

Fig-3.8 represents the amount of time students spend on mobile for chatting weekly. 41.6% students chat for less than 5 hours weekly while 39.3% students chats for 5-25 hours weekly.

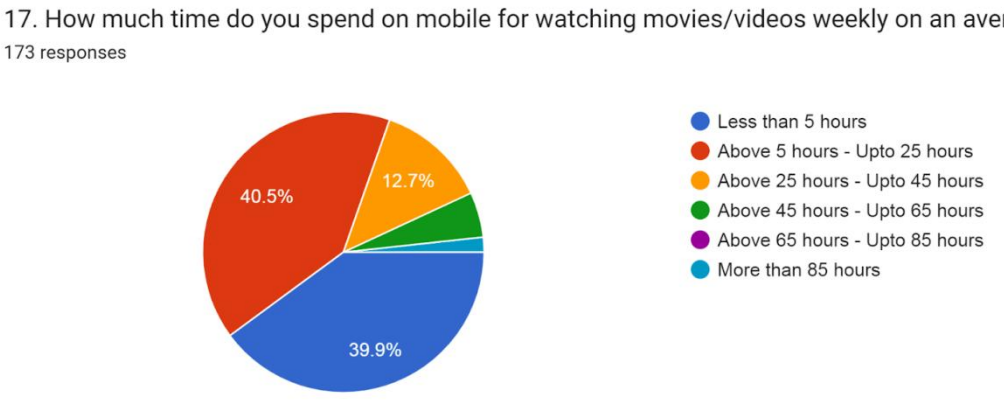


Fig-3.9

Fig-3.9 represents the amount of time students spend on mobile for watching videos weekly. 40.5% students spend 5-25 hours weekly for watching videos. 39.9% students spend less than 5 hours daily for watching videos.

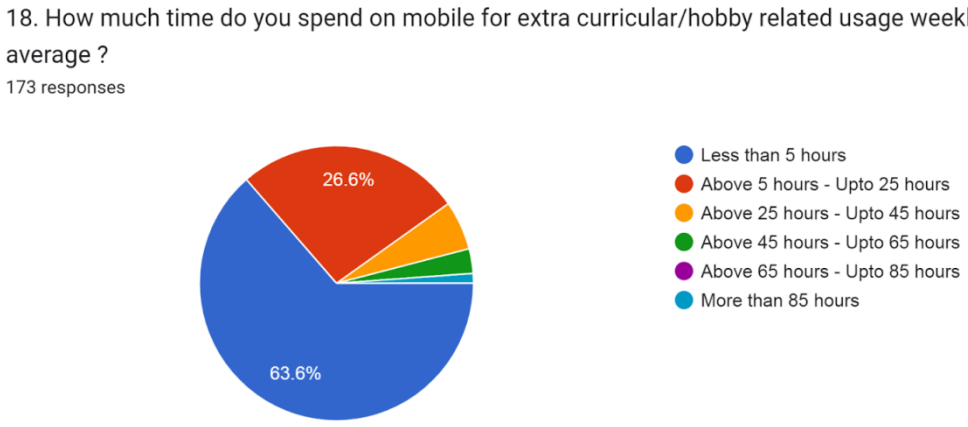


Fig-3.10

Fig-3.10 represents the amount of time students spend on mobile for hobby related usage weekly. Majority of the respondents (63.6%) uses mobile for hobby related usage less than 5 hours weekly. 26.6% students uses it for 5-25 hours weekly for that

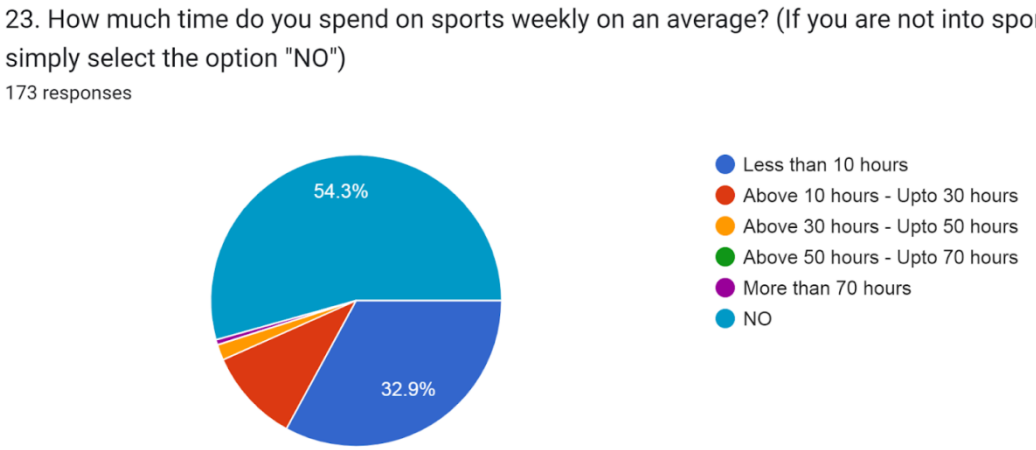


Fig-3.11

Fig-3.11 represents the amount of time students spend on sports weekly. More than half of the respondents (54.3%) are not even involved in sports. 32.9% students who are involved in sports spend less than 10 hours weekly on that.

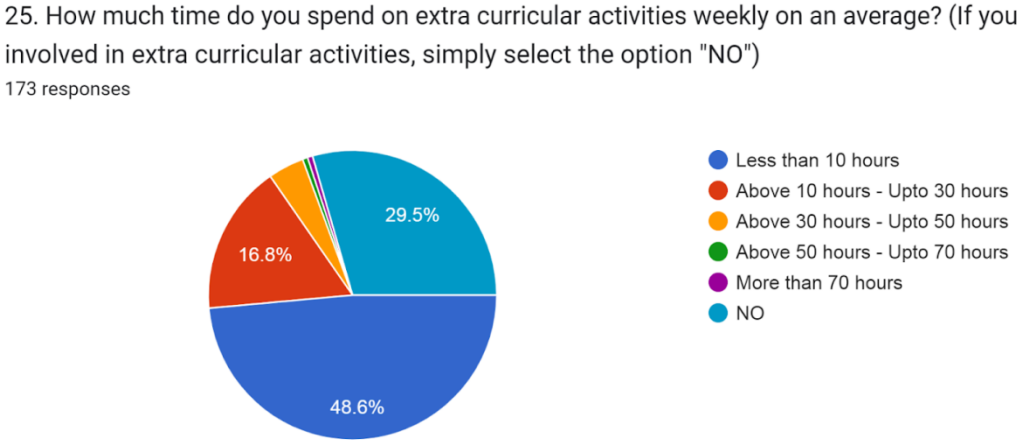


Fig-3.12

Fig-3.12 represents the amount of time students spend on extracurricular activity weekly. 48.6% students spend less than 10 hours weekly & 16.8% students spend 10-30 hours weekly for that purpose, while 29.5% students are not involved in extracurricular activities.

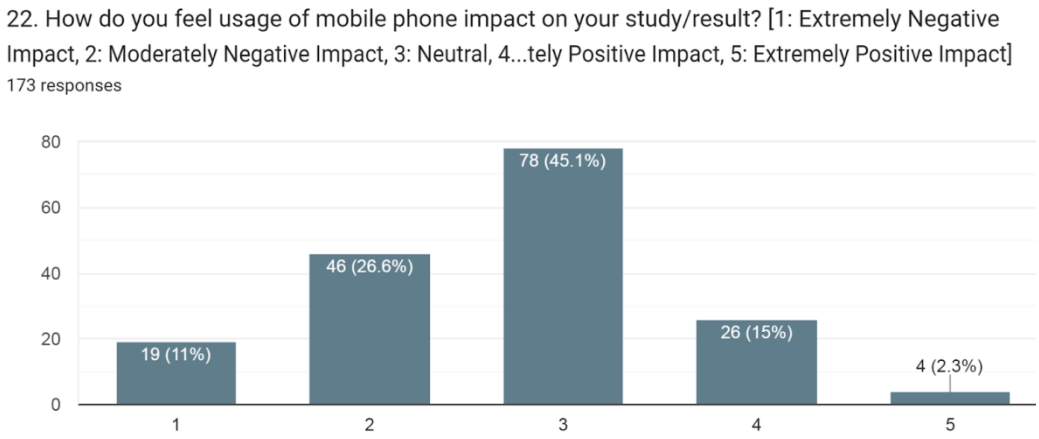


Fig-3.13

Fig-3.13 represents how the students feel about impact of mobile phone on study. 45.1% students are neutral in this matter. 26.6% students think it has moderately negative impact on study. 11% think it has extremely negative impact on study, While 15% and 2.3% student thinks it has moderately positive & extremely positive impact on study.

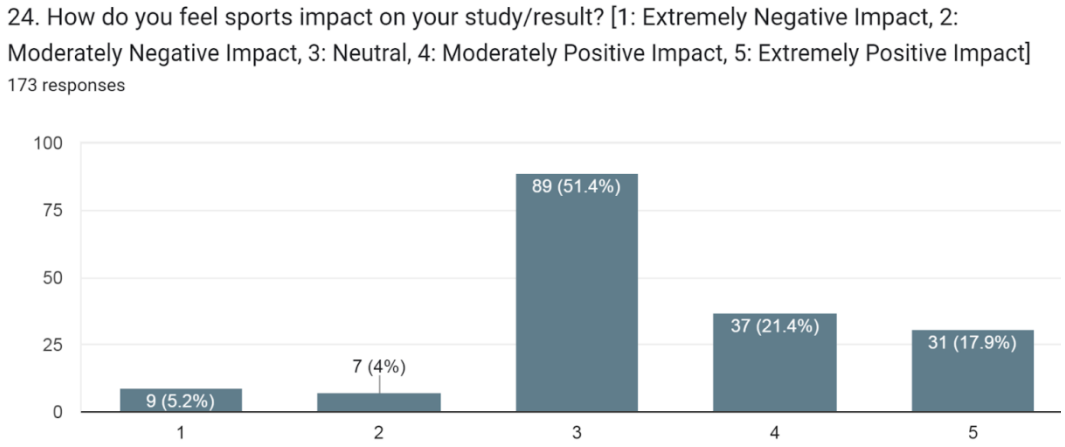


Fig-3.14

Fig-3.14 represents how the students feel about impact of sports on study. 51.4% students are neutral in this matter. 4% & 5.2% students think it has moderately negative & extremely negative impact on study. While 21.4% and 17.9% student thinks it has moderately positive & extremely

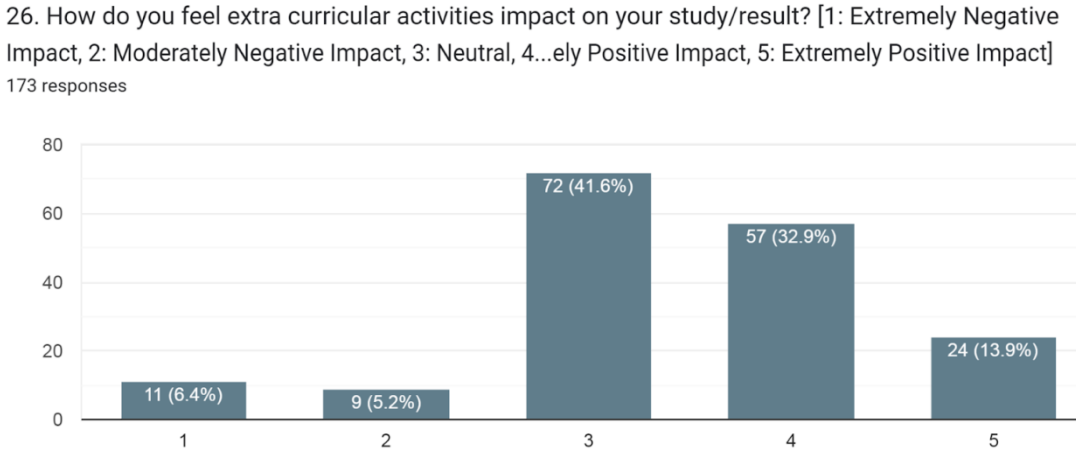


Fig-3.15

Fig-3.15 represents how the students feel about impact of extracurricular activities on study. 41.6% students are neutral in this matter. 5.2% & 6.4% students think it has moderately negative & extremely negative impact on study. While 32.9% and 13.9% student thinks it has moderately positive & extremely positive impact on study.

.

4 ANALYSIS

Chapter 4.1: Checking association between different factors using Goodman-Kruskal's gamma

As mentioned before in **"Methodology"**, our first task is to check for association between different factors (e.g. Time spend on study, time spend on mobile for gaming, time spend on sports etc.) and average of SGPA's of four semesters. In order to find Goodman-Kruskal's gamma & test for association, we need to construct contingency tables between different factors and average SGPA. Let us define,

- N_c, the number of pairs of cases ranked in the same order on both variables (number of concordant pairs),
- N_d, the number of pairs of cases ranked in reversed order on both variables (number of reversed pairs),
- Goodman-Kruskal's Gamma: $G = \frac{N_c - N_d}{N_c + N_d}$, $T = G\sqrt{\frac{(N_c + N_d)}{n(1-G^2)}}$

Factor wise contingency tables, value of Goodman-Kruskal's gamma and associations between them are shown below:

- **Association between total study hour and average SGPA:**

Table 4.1.1: Contingency table for total study hour and average SGPA

Average SGPA \ Total Study Hour	Less than 2 hr	2-5 hr	5-8 hr	8-11 hr	More than 11 hr	TOTAL
5 to 6	2	1	0	0	0	3
6 to 7	5	11	2	0	0	18
7 to 8	11	27	8	3	1	50
8 to 9	4	12	15	1	0	32
More than 9	5	7	2	1	1	16
TOTAL	27	58	27	5	2	119

Here,

$$N_c = 2091, \quad N_d = 1280, \quad G = 0.24058143, \quad T = 1.242857226, \quad p\,value = 0.216384448$$

Therefore, there is **no association** between **total study hour** and **average SGPA**.

- **Association between total time spend on mobile and average SGPA:**

Table 4.1.2: Contingency table for total time spend on mobile and average SGPA

Average SGPA \ Total Hour Spend On Mobile	Less than 10 hr	10-30 hr	30-50 hr	50-70 hr	70-90 hr	More Than 90 hr	TOTAL
5 to 6	1	1	1	0	0	0	3
6 to 7	2	9	5	1	1	0	18
7 to 8	6	13	17	6	5	3	50
8 to 9	7	12	6	4	3	0	32
More than 9	1	4	4	5	1	1	16
TOTAL	17	39	33	16	10	4	119

Here,

$$N_a = 2127, \quad N_i = 1758, \quad G = 0.094980695, \quad T = 0.540243471, \quad p\ value = 0.590047333$$

Therefore, there is **no association** between **total time spend on mobile** and **average SGPA**.

- **Association between total time spend on mobile for study and average SGPA:**

Table 4.1.3: Contingency table for total time spend on mobile for study and average SGPA

Average SGPA \ Total Hour Spend On Mobile for Study	Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
5 to 6	1	2	0	0	0	0	3
6 to 7	2	13	3	0	0	0	18
7 to 8	13	26	5	6	0	0	50
8 to 9	9	12	8	3	0	0	32
More than 9	4	9	2	0	0	1	16
TOTAL	29	62	18	9	0	1	119

Here,

$$N_a = 1643, \quad N_i = 1582, \quad G = 0.018914729, \quad T = 0.098449663, \quad p\ value = 0.921742363$$

Therefore, there is **no association** between **total time spend on mobile for study** and **average SGPA**.

- **Association between total time spend on mobile for gaming and average SGPA:**

Table 4.1.4: Contingency table for total time spend on mobile for gaming and average SGPA

		Total Hour Spend On Mobile for Gaming						
		Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
Average SGPA	5 to 6	2	1	0	0	0	0	3
	6 to 7	14	4	0	0	0	0	18
	7 to 8	44	4	2	0	0	0	50
	8 to 9	26	2	4	0	0	0	32
	More than 9	10	4	1	0	0	1	16
	TOTAL	96	15	7	0	0	1	119

Here,

$$N_a = 1054, \quad N_i = 718, \quad G = 0.189616253, \quad T = 0.718427162, \quad p\ value = 0.473914069$$

Therefore, there is **no association** between **total time spend on mobile for gaming** and **average SGPA**.

- **Association between total time spend on mobile for social media and average SGPA:**

Table 4.1.5: Contingency table for total time spend on mobile for social media and average SGPA

		Total Hour Spend On Mobile for Social Media						
		Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
Average SGPA	5 to 6	1	0	2	0	0	0	3
	6 to 7	6	10	2	0	0	0	18
	7 to 8	8	23	12	6	1	0	50
	8 to 9	11	14	5	2	0	0	32
	More than 9	4	8	3	0	0	1	16
	TOTAL	30	55	24	8	1	1	119

Here,

$$N_a = 1632, \quad N_i = 1786, \quad G = -0.045055588, \quad T = -0.241223688, \quad p\ value = 0.809800412$$

Therefore, there is **a no association** between **total time spend on mobile for social media** and **average SGPA**.

- **Association between total time spend on mobile for chatting and average SGPA:**

Table 4.1.6: Contingency table for total time spend on mobile for chatting and average SGPA

Average SGPA	Total Hour Spend On Mobile for Chatting: Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
5 to 6	1	2	0	0	0	0	3
6 to 7	9	9	0	0	0	0	18
7 to 8	19	17	6	5	2	1	50
8 to 9	16	12	3	1	0	0	32
More than 9	4	7	3	1	0	1	16
TOTAL	49	47	12	7	2	2	119

Here,

$$N_a = 1821, \quad N_i = 1482, \quad G = 0.102633969, \quad T = 0.537863938, \quad p\ value = 0.591683986$$

Therefore, there is **no association** between **total time spend on mobile for chatting** and **average SGPA**.

- **Association between total time spend on mobile for watching videos and average SGPA:**

Table 4.1.7: Contingency table for total time spend on mobile for watching videos and average SGPA

Average SGPA	Total Hour Spend On Mobile for Watching Videos: Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
5 to 6	1	1	1	0	0	0	3
6 to 7	10	6	2	0	0	0	18
7 to 8	15	25	6	4	0	0	50
8 to 9	12	15	4	1	0	0	32
More than 9	3	8	3	1	0	1	16
TOTAL	41	55	16	6	0	1	119

Here,

$$N_a = 1926, \quad N_i = 1356, \quad G = 0.173674589, \quad T = 0.898217458, \quad p\ value = 0.370898749$$

Therefore, there is **no association** between **total time spend on mobile for watching videos** and **average SGPA**.

- **Association between total time spend on mobile for extracurricular activities and average SGPA:**

Table 4.1.8: Contingency table for total time spend on mobile for extracurricular activities and average SGPA

Average SGPA \ Total Hour Spend On Mobile for Extra Curricular Activities	Less than 5 hr	5-25 hr	25-45 hr	45-65 hr	65-85 hr	More Than 85 hr	TOTAL
5 to 6	2	1	0	0	0	0	3
6 to 7	9	6	3	0	0	0	18
7 to 8	30	14	4	2	0	0	50
8 to 9	20	11	1	0	0	0	32
More than 9	11	5	0	0	0	0	16
TOTAL	72	37	8	2	0	0	119

Here,

$N_a = 1107, \quad N_i = 1564, \quad G = -0.171096967, \quad T = -0.798645535, \quad p\ value = 0.426100484$

Therefore, there is a **no association** between **total time spend on mobile for extracurricular activities** and **average SGPA**.

- **Association between total time spend on sports and average SGPA:**

Table 4.1.9: Contingency table for total time spend on sports and average SGPA

Average SGPA \ Total Hour Spend On Sports	Less than 10 hr	10-30 hr	30-50 hr	50-70 hr	More than 70 hr	NO	TOTAL
5 to 6	1	0	0	0	0	2	3
6 to 7	3	2	0	0	0	13	18
7 to 8	14	5	0	0	0	31	50
8 to 9	14	3	0	0	0	15	32
More than 9	7	1	2	0	0	6	16
TOTAL	39	11	2	0	0	67	119

Here,

$N_a = 1018, \quad N_i = 1883, \quad G = -0.298173044, \quad T = -1.405238624, \quad p\ value = 0.162577617$

Therefore, there is a **no association** between **total time spend on sports** and **average SGPA**.

- **Association between total time spend on extracurricular activities and average SGPA:**

Table 4.1.10: Contingency table for total time spend on extracurricular activities and average SGPA

Average SGPA \ Total Hour Spend On Extra Curricular Activities	Less than 10 hr	10-30 hr	30-50 hr	50-70 hr	More than 70 hr	NO	TOTAL
5 to 6	1	1	0	0	0	1	3
6 to 7	9	5	1	0	0	3	18
7 to 8	27	6	2	0	0	15	50
8 to 9	14	7	0	0	0	11	32
More than 9	8	2	1	0	0	5	16
TOTAL	59	21	4	0	0	35	119

Here,

$$N_a = 1728, \quad N_i = 1501, \quad G = 0.070300403, \quad T = 0.365293984, \quad p\ value = 0.715545921$$

Therefore, there is **no association** between **total time spend on extracurricular activities** and **average SGPA**.

- **Association between total time spend on sleep and average SGPA:**

Table 4.1.11: Contingency table for total time spend on sleep and average SGPA

Average SGPA \ Total Hour Spend On Sleep	Less than 2 hr	2-4 hr	4-6 hr	6-8 hr	8-10 hr	More than 10 hr	TOTAL
5 to 6	1	0	0	1	0	1	3
6 to 7	0	0	6	7	5	0	18
7 to 8	0	2	14	26	7	1	50
8 to 9	0	2	3	24	2	1	32
More than 9	0	0	6	8	1	1	16
TOTAL	1	4	29	66	15	4	119

Here,

$$N_a = 1532, \quad N_i = 1632, \quad G = -0.031605563, \quad T = -0.162888815, \quad p\ value = 0.870884607$$

Therefore, there is a **no association** between **total time spend on sleep** and **average SGPA**.

Summary: A summary of the above performed tests for associations are tabulated in a table for ease of understanding.

Table 4.1.12: Summarized table of associations

Factors	G	T	p value
Total Study Hour	0.24058	1.242857226	0.216384
Total Hour Spend On Mobile	0.09498	0.540243471	0.590047
Total Hour Spend On Mobile for Study	0.01891	0.098449563	0.921742
Total Hour Spend On Mobile for Gaming	0.18962	0.718427162	0.473914
Total Hour Spend On Mobile for Social Media	-0.04506	-0.241223688	0.8098
Total Hour Spend On Mobile for Chatting	0.10263	0.537863938	0.591684
Total Hour Spend On Mobile for Watching Videos	0.17367	0.898217458	0.370899
Total Hour Spend On Mobile for Extra Curricular Activities	-0.1711	-0.798645535	0.4261
Total Hour Spend On Sports	-0.29817	-1.348382657	0.180118
Total Hour Spend On Extra Curricular Activities	0.0703	0.365293984	0.715546
Total Hour Spend On Sleep	-0.03161	-0.162888815	0.870885

From the table shown above it can be observed that no factors has a significant association with average SGPA (i.e. students' results) as p-values of all factors are greater than 0.05. And as mentioned in **"Methodology"** earlier, if p-value > 0.05, we should accept the null hypothesis; i.e. **there is no association between factors and average SGPA**. However, we can also observe that Goodman-Kruskal gamma's value for **"Total Study Hour"** and **"Total Hour Spend On Sports"** are **0.24058** and **-0.29817** respectively. Which are the most and least values of gamma. More positive value indicates a positive association of factors with average SGPA and more negative value indicates a negative association for the same. One aim of our project was analysis of different factors to students' results, and another was to predict next semester results of students'. To achieve our second goal, we should construct a regression model for prediction. Since the values of G for the two factors viz. **"Total Study Hour"** and **"Total Hour Spend On Sports"** seem to have some sort of association, we will use them as explanatory variables. Further discussion follows in the next chapter.

Chapter 4.2: ANOVA for different factors

ANOVA provides a statistical test of whether two or more population means are equal, and therefore generalizes the *t*-test beyond two means. In other words, the ANOVA is used to test the difference between two or more means. Here in our case we are performing ANOVA between average SGPA & four different factors; i.e. total time spend on study, total time spend on mobile for study, total time spend on sports & total time spend on extracurricular activities. We want to see if there is significant difference between the levels of each factor. The ANOVA tables are shown below. We will use ANOVA one way fixed effect model, $\boldsymbol{y_{ij} = \mu + \alpha_i + e_{ij}}$, where $\boldsymbol{\alpha_i}$ = effect due to i-th level of that factor.

- **ANOVA between total mobile usage and average SGPA:**

Here our hypothesis of interest is;

$$H_{01}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A1}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.1: ANOVA Table between total mobile usage and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4.231913	5	0.846383	1.013991	0.412818	2.293911
Within Groups	95.1563	114	0.834704			
Total	99.38822	119				

Here, $F(Obs.) = 1.013991 < F_{0.95,5,114} = 2.293911$
Therefore, H_{01} is accepted, we can conclude that there is no significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are sampled from normal distribution, as p-value=0.41>0.05.

- **ANOVA between total time spend on mobile for study and average SGPA:**

Here our hypothesis of interest is;

$$H_{02}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A2}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.2: ANOVA Table between total time spend on mobile for study and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.223694	4	0.805923	0.963777	0.430258	2.450571
Within Groups	96.16452	115	0.836213			
Total	99.38822	119				

Here, $F(Obs.) = 0.96378 < F_{0.95,4,115} = 2.450571$
Therefore, H_{02} is accepted, we can conclude that there is no significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are sampled from normal distribution, as p-value=0.43>0.05.

- **ANOVA between total time spend on mobile for Extracurricular activities and average SGPA:**

Here our hypothesis of interest is;

$$H_{03}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A3}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.3: ANOVA Table between total time spend on mobile for Extracurricular activities and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	2.607691	3	0.86923	1.041849	0.376851	2.682809
Within Groups	96.78053	116	0.834315			
Total	99.38822	119				

Here, $F(Obs.) = 1.041849 < F_{0.95,3,116} = 2.682809$
Therefore, H_{03} is accepted, we can conclude that there is no significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are sampled from normal distribution, as p-value=0.38>0.05.

- **ANOVA between total time spend on sports and average SGPA:**

Here our hypothesis of interest is;

$$H_{04}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A4}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.4: ANOVA Table between total time spend on sports and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	9.970095	3	3.323365	4.311322	0.00638	2.682809
Within Groups	89.41812	116	0.770846			
Total	99.38822	119				

Here, $F(Obs.) = 4.311322 > F_{0.95,3,116} = 2.682809$
Therefore, H_{04} is rejected, we can conclude that there is significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are not sampled from normal distribution, as p-value=0.00638<0.05.

- **ANOVA between total time spend on study and average SGPA:**

Here our hypothesis of interest is;

$$H_{05}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A5}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.5: ANOVA Table between total time spend on study and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	4.248988	4	1.062247	1.283996	0.280397	2.450571
Within Groups	95.13923	115	0.827298			
Total	99.38822	119				

Here, $F(Obs.) = 1.283996 < F_{0.95,4,115} = 2.450671$
Therefore, H_{05} is accepted, we can conclude that there is no significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are sampled from normal distribution, as p-value=0.28>0.05.

- **ANOVA between total time spend on Extracurricular activities and average SGPA:**

Here our hypothesis of interest is;

$$H_{06}: \alpha_i = 0 \; \forall \; i \;\; vs. \;\; H_{A6}: \alpha_i \neq 0 \text{ for at least one } i$$

Table 4.2.6: ANOVA Table between total time spend on Extracurricular activities and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	0.622884	3	0.207628	0.243859	0.865555	2.682809
Within Groups	98.76533	116	0.851425			
Total	99.38822	119				

Here, $F(Obs.) = 0.243859 > \; F_{0.95,3,116} = 2.682809$
Therefore, H_{06} is accepted, we can conclude that there is no significant difference among the means of different groups.

Test for normality using Shapiro-Wilk test: To check whether the value is sampled from a normal distribution or not, we will test for normality. We can check that by observing the p-value of the given ANOVA table. If the p-value is greater than the chosen significance level (in our case 5% level) then the values are sampled from a normal distribution. Again, if the p-value is lesser than the chosen significance level (in our case 5% level) then the values are not sampled from a normal distribution.
Therefore, in our case, the values are sampled from normal distribution, as p-value=0.86>0.05.

Table 4.2.7: A summarized table showing the results of ANOVA

Factors	Significant or not
Total Study Time	No
Total Time Spend on Sports	Yes
Total Time Spend on Extracurricular Activities	No
Total Time Spend on Mobile	No
Total Time Spend on Mobile for Study	No
Total Time Spend on Mobile for Extracurricular Activities	No

From the above table, it can be seen that only one factor's levels has significant difference viz. **"Total Time Spend on Sports".** Intuitively it seems sceptical. We can observe from our collected data that a huge number of students are actually not involved to sports. That is very normal, as sports is not a common domain for everybody. So we performed another ANOVA on **"Total Time Spend on Sports"** but this time excluding the students who are not involved in sports. The result is shown below

Table 4.2.8: ANOVA Table between total time spend on sports (excluding those who are not involved) and average SGPA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	3.909074	2	1.954537	2.360223	0.104836	3.18261
Within Groups	41.40578	50	0.828116			
Total	45.31485	52				

So from the above ANOVA table, it can be seen that

$$F(Obs.) = 2.360223 < F_{0.95,2,50} = 3.18261$$

This suggests that there is no significant difference between the levels of "Total Time Spend on Sports". Also in this case p-value = 0.10 > 0.05, i.e. the normality assumption also holds by Shapiro-Wilk test.

So, we can conclude that there is no significant difference between the levels of all factors.

Chapter 4.3: Regression and Prediction

As mentioned in the previous chapter we will use **"Total Study Hour"** (Say,x_1) and **"Total Hour Spend On Sports"** (Say,x_2) as the explanatory variables for our regression. Average SGPA (Say, y) will be our response variable. We have to use concept of **dummy variables** to perform regression as our data is categorical. In order to construct dummy variables, we should define some new variables which will represent different levels of the explanatory variables. **"Total Study Hour"** has 5 levels and and **"Total Hour Spend On Sports"** has 6 levels. So we have to define 4 new variables for total study hour and 5 new variables for total hour spend on study.
Let us define, z_1, z_2, z_3, z_4 corresponding to the levels of "Total Study Hour", w_1, w_2, w_3, w_4, w_5 corresponding to the levels of "Total Hour Spend On Sports", and y_1, y_2 corresponding to the SGPA as

$$y_1 = \text{average SGPA upto } 4^{\text{th}} \text{ semester}, \; y_2 = 5^{\text{th}} \text{ semester SGPA}$$

$$z_1 = \begin{cases} 1, & if \;\; x_1 \; lies \; in \text{ "Less than 2 hours"} \\ 0, & otherwise \end{cases}; \quad z_2 = \begin{cases} 1, & if \;\; x_1 \; lies \; in \text{ "2 - 5 hours"} \\ 0, & otherwise \end{cases}$$

$$z_3 = \begin{cases} 1, & if \;\; x_1 \; lies \; in \text{ "5 - 8 hours"} \\ 0, & otherwise \end{cases}; \qquad z_4 = \begin{cases} 1, & if \;\; x_1 \; lies \; in \text{ "8 - 11 hours"} \\ 0, & otherwise \end{cases}$$

And,

$$w_1 = \begin{cases} 1, & if \;\; x_2 \; lies \; in \text{ "Less than 10 hours"} \\ 0, & otherwise \end{cases}; \quad w_2 = \begin{cases} 1, & if \;\; x_2 \; lies \; in \text{ "Above 10 hours – Upto 30 hours"} \\ 0, & otherwise \end{cases}$$

$$w_3 = \begin{cases} 1, & if \;\; x_2 \; lies \; in \text{ "Above 30 hours – Upto 50 hours"} \\ 0, & otherwise \end{cases};$$

$$w_4 = \begin{cases} 1, & if \;\; x_2 \; lies \; in \text{ "Above 50 hours – Upto 70 hours"} \\ 0, & otherwise \end{cases};$$

$$w_5 = \begin{cases} 1, & if \;\; x_2 \; lies \; in \text{ "More than 70 hours"} \\ 0, & otherwise \end{cases}$$

Now, using these variables we have constructed dummy variables and performed regression using **MS EXCEL.** Our regression line is given by,

$$\hat{Y}_1 = 8.33604 + (-0.89481)\, z_1 + (-0.8511)\, z_2 + (-0.4849)\, z_3 + (-0.30412) z_4 + 0.4429\, w_1 + 0.18872\, w_2 + 1.9094\, w_3$$

Now, we can use this regression line to predict students' average SGPAs upto 4^{th} semester, which is shown by a table in the next page and also represented graphically in page-28 (Fig. 4.3.1).

Table 4.3.1: Predicted Vs Actual Average SGPA Upto 4th Semester

Name	Predicted Average SGPA	Actual Average SGPA	Name	Predicted Average SGPA	Actual Average SGPA
Diptajit Chakraborty	7.485	7.400	Prat	7.441	9.125
Saheb Tirtha Dey	8.475	9.518	Rupsa Chakraborty	7.851	8.188
Pritam Saha	7.485	7.603	Rathantari Mahapatra	7.851	8.125
Shagnick Roy	7.441	7.795	Trisha Chatterjee	7.485	6.275
Sayan Dewanjee	7.928	8.575	Tiyasa Chakraborty	7.851	7.125
Trambak Banerjee	7.928	8.125	Debarati Mukherjee	7.851	6.548
Indranil Ghosh	7.884	8.170	GENIA PAUL	8.221	7.775
Sneha Paul	7.485	8.050	A	8.294	8.025
SAPTAJIT DUTTA	8.032	8.275	Priyanka Das	7.485	7.275
Bristi Maity	7.884	9.075	Arkajyoti Banerjee	7.928	7.700
Deep Singha	7.485	9.133	Arunima Saha	7.485	5.841
Sirsha Ghosh	8.779	7.775	Pritam Saha	7.441	9.813
Subhayan Sahoo	7.674	8.450	Santanu Paul	7.674	6.575
Sarfaraz Molla	7.441	5.970	Upasana Das	8.294	8.850
Srija Acharjee	8.294	7.054	Akash Sen	8.294	8.653
Niladri Banerjee	7.928	6.775	Sudipa Das	7.485	7.790
Atreyee Bose	7.884	8.000	Rohan Mitra	7.928	7.413
Daipayan Bhattacharyya	7.441	7.368	Sandipta Malakar	7.485	7.545
Aniruddha Mukherjee	8.294	7.775	Snehasis Koley	8.040	9.565
Shibaji Das	7.485	7.247	Arijit Roy	7.630	7.225
Ahan Mazumdar	7.851	8.100	Sabina Yasmin	7.441	6.468
Swarnali Dutta	7.485	7.531	Jamirul Chowdhury	7.928	7.025
Soham Choudhury	7.485	7.913	Arkaadeb Kapat	7.485	6.885
Tathastu Ghosh	8.040	8.582	Shreya Maiti	7.928	7.200
Ayush	7.928	9.300	Titli Biswas	8.032	7.625
Subhanjan Debnath	7.851	7.789	Tamojit Ghosh	7.928	7.225
Shikta	7.928	8.083	Soumyadeep Basak	7.928	6.725
Avik Bhuin	7.928	7.200	Annesha Khamrui	7.485	7.388
Rohit Sen	8.294	6.850	Krittika Chatterjee	7.928	7.925
Asmita De	7.485	7.221	Srijeeta Bhattacharjee	7.485	8.500

Name	Predicted Average SGPA	Actual Average SGPA	Name	Predicted Average SGPA	Actual Average SGPA
Chitra Thakur	7.485	7.275	**Romit**	7.630	6.650
Surjyadeep Bose	7.674	7.887	**Sara Ahamed**	7.485	6.855
Siboham Pattanayak	7.928	9.883	**Sulagna Das**	7.485	7.670
Soumadip Das	7.928	8.075	**Sankhya Sen**	9.394	9.450
Mayukh Mukherjee	7.485	8.040	**Rajdeep Mondal**	7.441	9.148
Safalya Pan	7.884	8.998	**Aniket Agarwal**	7.851	7.135
Lalchand Gorai	8.294	8.525	**Parna Dutta**	7.485	6.800
Tanupriyo Das	7.928	8.713	**Priya Mondal**	7.851	8.095
Poushomi Chatterjee	7.485	8.098	**Soumyajyoti Chakraborty**	7.928	9.375
Antara Naskar	7.441	7.770	**A S Praisie Jemimah**	7.928	9.500
Sagnik Chowdhury	7.485	8.850	**Adityabarna Bose**	7.851	7.475
Aishani Chowdhury	7.485	6.438	**Parthib Saha**	7.851	9.188
Sagnik kundu	7.884	5.575	**Adrisha Dutta**	7.485	8.758
Souvik Nath	7.928	9.120	**Rohit Mondal**	7.884	8.240
ANIKET SAHA	7.485	6.825	**Somnath Banerjee**	9.351	9.295
Ritam Bhattacharya	8.040	8.225	**Soumita Datta**	7.630	7.750
Ankana Ghosh	7.851	8.075	**Snigdha Saha**	7.851	8.354
Joy Ganguly	7.851	8.928	**Souvik Ghosh**	8.294	8.110
Tanush Mondal	7.851	7.125	**Shreenidhi Roy**	7.441	7.050
Sharanya Das	7.928	7.048	**Debayan Sen**	7.485	7.406
Soumya Deep saha	8.336	9.340	**Sreeja Maity**	7.485	7.361
Soma Sikder	7.441	7.563	**Anushka Dutta**	7.441	6.800
Puja Pandey	7.441	6.575	**Priyanka Dasgupta**	7.441	7.475
Anisha Dutta	7.851	7.325	**Ananya Ghosh**	7.441	7.128
Debaditya Chakraborty	7.884	7.735	**Adrita Sen**	7.928	7.158
SK	7.851	8.305	**Soumili Hati**	7.485	6.825
Kousik Pramanik	7.884	8.340	**Rajdeep mukherjee**	7.485	7.433
Ananya Das	7.485	7.408	**Shifa Badgujar**	7.441	6.225
Bappaditya Debnath	8.221	7.788	**Rizwan Rizwi**	7.485	6.525
Abhijnan Sarkar	7.928	7.685	**Anuska Ghosh**	7.441	7.050

Now, we will again perform regression using **MS EXCEL** to predict 5th semester SGPA**.** Our regression line is given by,

$$\hat{Y}_2 = -1.3695 + 1.0995\, y_1 + 0.5364\, z_1 + 0.4121\, z_2 + 0.348\, z_3 + 0.7339\, z_4 + (-0.1051) w_1 + (-0.3035)\, w_2$$

Now, we can use this regression line to predict 5th semester SGPA, which is shown by a table in the next page and also represented graphically in page-28 (Fig. 4.3.2).

Table 4.3.2: Predicted Vs Actual 5th Semester SGPA

Name	Predicted SGPA	Actual SGPA
Diptajit Chakraborty	7.179	7.284
Saheb Tirtha Dey	9.724	10.000
Pritam Saha	7.402	7.600
Shagnick Roy	7.738	9.000
Sayan Dewanjee	8.366	7.939
Trambak Banerjee	7.871	8.100
Indranil Ghosh	8.045	9.110
Sneha Paul	7.894	7.460
SAPTAJIT DUTTA	8.463	8.187
Bristi Maity	9.040	8.810
Deep Singha	9.084	9.000
Niladri Banerjee	6.387	6.199
Aniruddha Mukherjee	7.422	7.568
Ahan Mazumdar	7.885	8.101
Soham Choudhury	7.743	8.050
Tathastu Ghosh	8.111	8.736
Subhanjan Debnath	7.542	8.700
Shikta	7.824	8.030
Rohit Sen	6.405	5.841
Asmita De	6.983	7.391
Rupsa Chakraborty	7.981	7.200
Debarati Mukherjee	6.178	6.554
Arkajyoti Banerjee	7.404	7.837
Akash Sen	8.387	8.189
Sudipa Das	7.608	7.500
Rohan Mitra	7.088	6.600
Sandipta Malakar	7.339	7.118
Arijit Roy	6.807	7.000
Arkaadeb Kapat	6.613	6.680
Soumyadeep Basak	6.332	7.450
Krittika Chatterjee	7.651	7.725
Sagnik Chowdhury	8.773	8.674
Aishani Chowdhury	6.121	5.610
Ritam Bhattacharya	7.719	6.901
Soumya Deep saha	8.900	8.900
Soma Sikder	7.482	7.056
Puja Pandey	6.396	5.984
Debaditya Chakraborty	7.567	5.968
Kousik Pramanik	8.232	9.000
Ananya Das	7.187	6.600
Parthib Saha	9.080	8.920
Rohit Mondal	8.122	7.500

Fig.4.3.1: Predicted vs Actual Average SGPA upto 4th semester (Table 4.3.1)

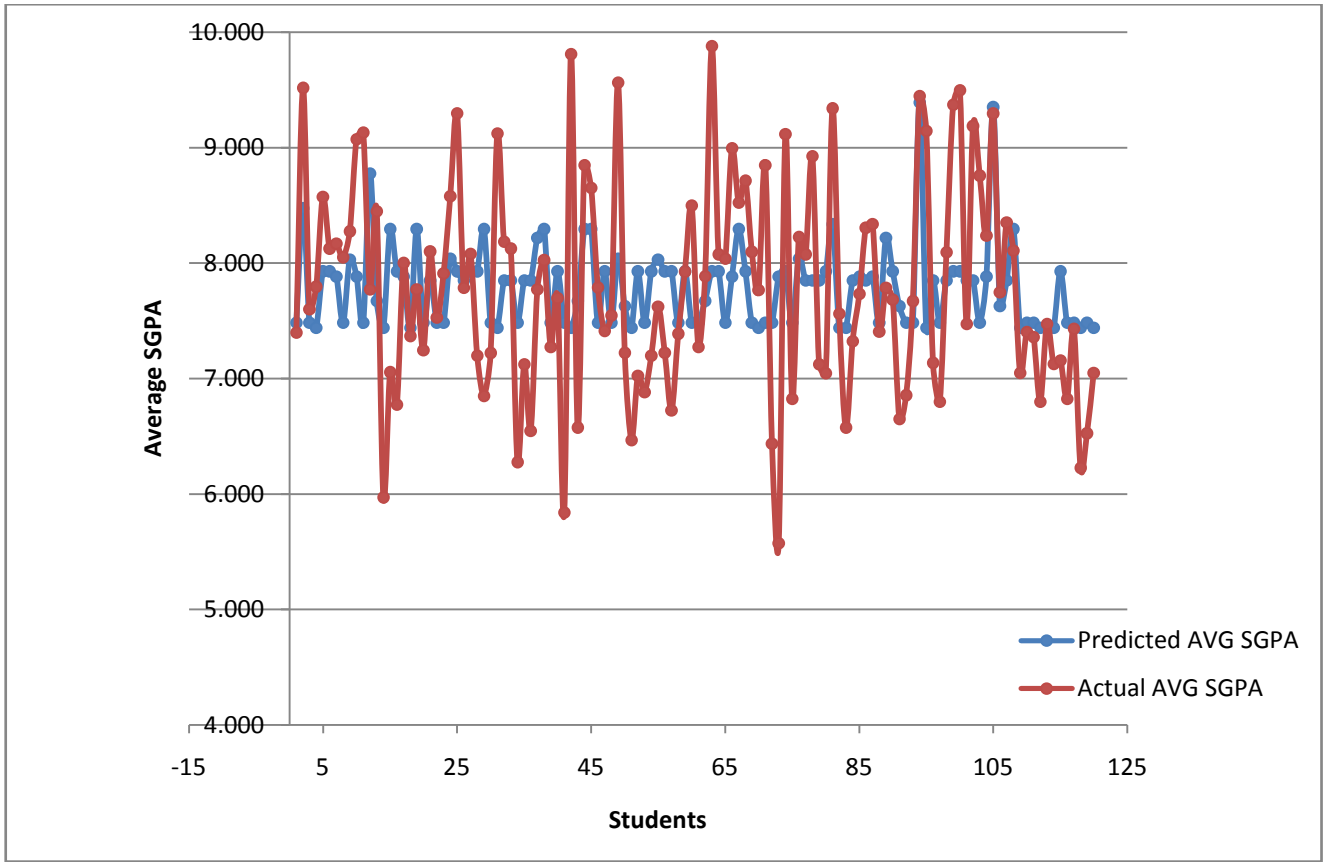
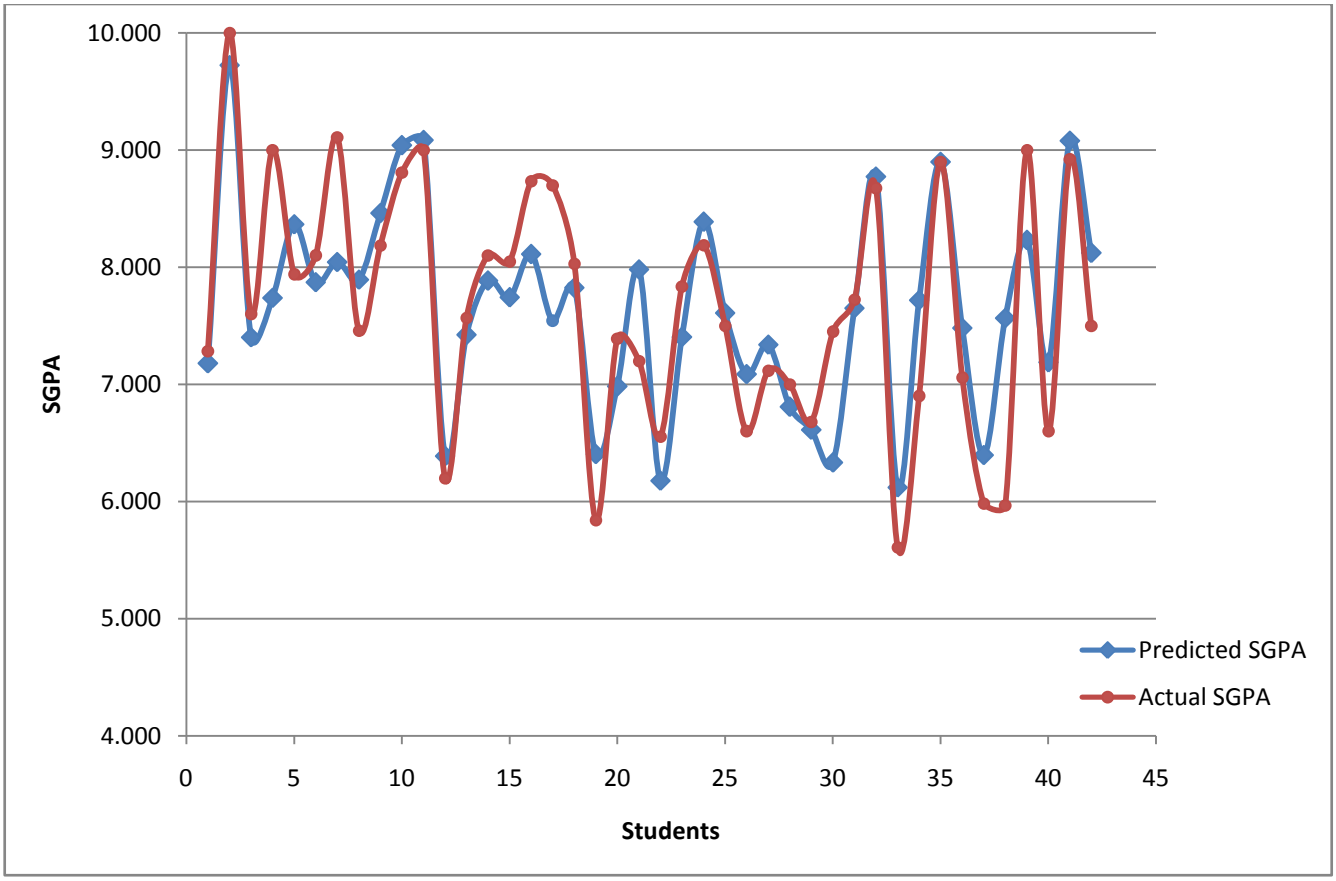


Fig.4.3.2: Predicted vs Actual 5th semester SGPA (Table 4.3.2)



Chapter 4.4: Confusion Matrix

Now based on the comparison of predicted average SGPA and actual average SGPA upto 4th semester shown in Table: 4.2.1, we will construct a confusion matrix. That will help us to find the accuracy, precision, sensitivity and specificity of our fit.

Table 4.4.1: Confusion Matrix based on predicted and actual average SGPA upto 4th semester

Actual Value / Predicted Value	False	True
Negative	21	0
Positive	0	99

Here,

True Positive **:** The number of samples correctly predicted as positive is 99

True Negative : The number of samples correctly predicted as negative is 0

False Positive : The number of samples incorrectly predicted as positive is 0

False Negative : The number of samples incorrectly predicted as negative is 21

Now,

Accuracy : 0.825

Precision **:** 1

Sensitivity **:** 0.825

Specificity : Undefined

Table 4.4.2: Confusion Matrix based on predicted and actual 5th semester SGPA

Actual Value / Predicted Value	False	True
Negative	4	6
Positive	3	29

Here,

True Positive **:** The number of samples correctly predicted as positive is 29

True Negative : The number of samples correctly predicted as negative is 6

False Positive : The number of samples incorrectly predicted as positive is 3

False Negative : The number of samples incorrectly predicted as negative is 4

Now,

Accuracy : 0.834

Precision **:** 0.90625

Sensitivity **:** 0.878

Specificity : 0.67

5 CONCLUSION

After all these works, now it's time to make a conclusion based on the results we have got. Our project had two aims:

i) To analyse whether or not different factors affect students' results, and
ii) To predict 5th semester results of students.

To fulfill our first goal, we used Goodman-Kruskal gamma and ANOVA. By Goodman-Kruskal gamma we have checked for associations between different factors and average SGPA. As shown in chapter 4.1, it can be clearly seen that there is no significant association between different factors and average SGPA, i.e. actually all these factors do not really have any impact on students' results, at least in case of our study. However, "Total Study Hour" and "Total Time Spend on Sports" seem to have some sort of association. Then we performed ANOVA to check if there is any significant difference among the levels of the factors. From chapter 4.2, it can be observed that there doesn't have any significant difference among the levels of different factors.

Now, to fulfill our second goal, we have done prediction using regression. From chapter 4.3, we can predict the 5th semester SGPA of all students. Then we have done a comparison between the observed and predicted 5th semester SGPA for a sample of 42 students. It can be seen that predictions are more or less good. To check if our model fits well or not, we have constructed a confusion matrix. From the confusion matrix we have found accuracy, precision, specificity and sensitivity of our fitted model to the sample. In case of our sample, our precision is 0.906 and accuracy is 83%, i.e. pretty good.

I wonder how good the prediction would be to our 6th semester SGPA!

REFERENCES

- Giri, Prasanta Kumar and Banerjee, Jiban – Introduction to Statistics
- Gun, A.M., Gupta, M.K. and Dasgupta, B. – An Outline of Statistical Theory, Vol. 2
- Gun, A.M., Gupta, M.K. and Dasgupta, B. – Fundamentals of Statistics, Vol. 2
- Chatterjee, Samprit and Hadi, Alis – Regression Analysis by Example
- WWW.Wikipedia.Com
- WWW.Youtube.Com
- WWW.Medium.Com

A C K N O W L E D G E M E N T

I would like to express my heartfelt gratitude and extend my sincerest appreciation to the individuals who have played a significant role in the successful completion of my project. Their guidance, support, and expertise have been invaluable throughout my journey.

First and foremost, I am deeply indebted to my **supervisor**, **Prof. Anup Kumar Giri**, for his unwavering dedication, insightful feedback, and continuous encouragement. His profound knowledge in the field of statistics and his willingness to share his expertise have been instrumental in shaping the direction of my project. I am truly grateful for his mentorship and guidance, which have been invaluable assets in this endeavour.

I would also like to extend my sincere thanks to the **Head of the Statistics, Department**, **Maulana Azad College**, **Prof. Partha Pal**, for his constant support and encouragement. His vision and leadership have provided me with a conducive environment to pursue my research interests. I am grateful for his valuable insights and the opportunities he has provided for my intellectual growth.

Additionally, I would like to express my gratitude to the other esteemed professor of the Statistics Department, **Prof. Tuhinsubhra Bhattacharya.** His expertise and willingness to share his knowledge have been of immense help in refining my research ideas and broadening my understanding of statistical concepts. Their constructive criticism and valuable suggestions have played a pivotal role in shaping my project.

I would also like to acknowledge my fellow friends who have supported me throughout this journey. Their camaraderie and encouragement have been a constant source of motivation. In particular, my friends **Tathastu Ghosh, Arkajyoti Banerjee, Pritam Saha, Shagnick Roy and Eshan De** have provided me immense support and guidance towards completion of my project, in every respect.

Lastly, I would like to thank **Maulana Azad College** for providing me with a nurturing academic environment and the necessary resources to pursue my project. I am grateful for the opportunities and platform that the college has offered me to explore my research interests and enhance my statistical knowledge.

A P P E N D I X

R Code for data cleaning:

```
>data=read.csv("C\\Whole Data.csv", header=T,sep= ",")
>dim(data)
>data=data[,-1]
>data=na.omit(data)
>rownames(data)=NULL
>SGPA=as.matrix(data[,6:9])
>class(SGPA);typeof(SGPA)
>sem.mean=rowMeans(SGPA)
>screen.time=data[,12]
>screen.time=factor(screen.time)
>str(screen.time)
>on.screen.study=data[,13]
>on.screen.study=factor(on.screen.study)
>str(on.screen.study)
>hobbies=data[,18]
>hobbies=factor(hobbies)
>sports=data[,23]
>sports=factor(sports)
```

R Code for ANOVA:

```
>model1=aov(sem.mean~screen.time)
>summary(model1)
```

This is the R code for ANOVA between average SGPA and total mobile usage. We used similar coding for ANOVA between average SGPA and other factors.

How different factors affect students' results ?

Hello, I am Santarpan Pal, a 3rd year BSc. Statistics(Hons) student from Maulana Azad College, Kolkata. I am preparing my final semester project work on the topic "**How different factors affect students' results**"- only for students who are currently studying in 3rd year.

I would be thankful if you give some of your valuable time to fill this form in. It will be really helpful for me.

Your response will only be used for academic research purposes.

Thanks in advance.

* Indicates required question

1. Email *

2. 1.Your Name *

3. 2. Gender *

   Mark only one oval.

   - ◯ Male
   - ◯ Female

4. 3. Name of your institution *

5. 4. The University your Institution is affiliated to *

Mark only one oval.

- ◯ University of Calcutta
- ◯ Maulana Abul Kalam Azad University of Technology
- ◯ West Bengal State University
- ◯ Techno University
- ◯ Sister Nivedita University
- ◯ Vidyasagar University
- ◯ Other: _____

6. 5. How much time do you spend on study daily on an average? *

Mark only one oval.

- ◯ Less than 2 hours
- ◯ 2-5 hours
- ◯ 5-8 hours
- ◯ 8-11 hours
- ◯ More than 11 hours

7. 6. Your 1st semester SGPA *
(If you have backlog and can't provide SGPA, simply write "NA")

_____

8. 7. Your 2nd semester SGPA *
(If you have backlog and can't provide SGPA, simply write "NA")

_____

9. 8. Your 3rd semester SGPA *
(If you have backlog and can't provide SGPA, simply write "NA")

_____

10. 9. Your 4th semester SGPA *
(If you have backlog and can't provide SGPA, simply write "NA")

_____

11. 10. Which range of SGPA you expected yourself to get in the previous semester (4th semester)? *

Mark only one oval.

- ◯ Above 4 - Upto 5
- ◯ Above 5 - Upto 6
- ◯ Above 6 - Upto 7
- ◯ Above 7 - Upto 8
- ◯ Above 8 - Upto 9
- ◯ Above 9

12. 11. Which range of SGPA you expect yourself to get in the next semester (5th semester)? *

Mark only one oval.

- ◯ Above 4 - Upto 5
- ◯ Above 5 - Upto 6
- ◯ Above 6 - Upto 7
- ◯ Above 7 - Upto 8
- ◯ Above 8 - Upto 9
- ◯ More than 9

13. 12. How much time do you spend on mobile weekly on an average? * (Aggregated total of mobile phone usage)

Mark only one oval.

- Less than 10 hours
- Above 10 hours - Upto 30 hours
- Above 30 hours - Upto 50 hours
- Above 50 hours - Upto 70 hours
- Above 70 hours - Upto 90 hours
- More than 90 hours

14. 13. How much time do you spend on mobile for study weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

15. 14. How much time do you spend on mobile for gaming weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

16. 15. How much time do you spend on mobile for social media usage weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

17. 16. How much time do you spend on mobile for chatting weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

18. 17. How much time do you spend on mobile for watching movies/videos weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

19. 18. How much time do you spend on mobile for extra curricular/hobby related usage weekly on an average ? *

Mark only one oval.

- Less than 5 hours
- Above 5 hours - Upto 25 hours
- Above 25 hours - Upto 45 hours
- Above 45 hours - Upto 65 hours
- Above 65 hours - Upto 85 hours
- More than 85 hours

20. 19. Which kind of network service/s do you use for internet at your residence? *

Mark only one oval.

- Wi-fi through Broadband / Optical Fibre
- Mobile Data
- Both

21. 20. How much money do you spend on mobile recharge monthly on an average? *

Mark only one oval.

- Less than 50 Rs.
- Above 50 Rs.- Below 100 Rs.
- Above 100 Rs.- Below 200 Rs.
- Above 200 Rs.- Below 300 Rs.
- More than 300 Rs.

22. 21. Does your monthly recharge satisfy your requirements? *

Mark only one oval.

- ◯ Yes
- ◯ No

23. 22. How do you feel usage of mobile phone impact on your study/result? * [1: Extremely Negative Impact, 2: Moderately Negative Impact, 3: Neutral, 4: Moderately Positive Impact, 5: Extremely Positive Impact]

Mark only one oval.

- 1 ◯
- 2 ◯
- 3 ◯
- 4 ◯
- 5 ◯

24. 23. How much time do you spend on sports weekly on an average? *
(If you are not into sports, simply select the option "NO")

Mark only one oval.

- ◯ Less than 10 hours
- ◯ Above 10 hours - Upto 30 hours
- ◯ Above 30 hours - Upto 50 hours
- ◯ Above 50 hours - Upto 70 hours
- ◯ More than 70 hours
- ◯ NO

25. 24. How do you feel sports impact on your study/result? *
[1: Extremely Negative Impact, 2: Moderately Negative Impact, 3: Neutral, 4: Moderately Positive Impact, 5: Extremely Positive Impact]

Mark only one oval.

- 1 ◯
- 2 ◯
- 3 ◯
- 4 ◯
- 5 ◯

26. 25. How much time do you spend on extra curricular activities weekly on an average? * (If you are not involved in extra curricular activities, simply select the option "NO")

Mark only one oval.

- Less than 10 hours
- Above 10 hours - Upto 30 hours
- Above 30 hours - Upto 50 hours
- Above 50 hours - Upto 70 hours
- More than 70 hours
- NO

27. 26. How do you feel extra curricular activities impact on your study/result? * [1: Extremely Negative Impact, 2: Moderately Negative Impact, 3: Neutral, 4: Moderately Positive Impact, 5: Extremely Positive Impact]

Mark only one oval.

1	
2	
3	
4	
5	

28. 27. How much time do you sleep daily on an average? *

Mark only one oval.

- Less than 2 hours
- Above 2 hours - Upto 4 hours
- Above 4 hours - Upto 6 hours
- Above 6 hours - Upto 8 hours
- Above 8 hours - Upto 10 hours
- More than 10 hours

29. 28. Remarks (If any)

Sl. No.	1. Your Name	2. Gender	3. Name of your institution	4. The University your Institution is affiliated to	5. How much time do you spend on study daily on an average?
1	Diptajit Chakraborty	Male	Maulana Azad College	University of Calcutta	2-5 hours
2	Saheb Tirtha Dey	Male	Adamas University	Adamas University	8-11 hours
3	Pritam Saha	Male	St. Xavier's College kolkata	University of Calcutta	2-5 hours
4	Shagnick Roy	Male	Presidency University	Presidency University	Less than 2 hours
5	Sayan Dewanjee	Male	Surendranath College	University of Calcutta	2-5 hours
6	Trambak Banerjee	Male	Institute of Technical Education and Research	Siksha 'O' Anusandhan University, Bhubaneswar	2-5 hours
7	Indranil Ghosh	Male	Heritage Institute of Technology, Kolkata	Maulana Abul Kalam Azad University of Technology	Less than 2 hours
8	Sneha Paul	Female	Bengal Institute of Technology	Maulana Abul Kalam Azad University of Technology	2-5 hours
9	SAPTAJIT DUTTA	Male	MAULANA AZAD COLLEGE	University of Calcutta	8-11 hours
10	Bristi Maity	Female	Techno main salt lake	Maulana Abul Kalam Azad University of Technology	Less than 2 hours
11	Priti Prakash Mohanta	Male	Maulana Azad College	University of Calcutta	Less than 2 hours
12	দীপ সিংহ	Male	দমদম মতিঝিল কলেজ (DDMC)	West Bengal State University	2-5 hours
13	Saikat Maiti	Male	Surendranath college	University of Calcutta	5-8 hours
14	Sirsha Ghosh	Female	Lady Brabourne college	University of Calcutta	More than 11 hours
15	Subhayan Sahoo	Male	UEM	IEM	2-5 hours
16	Sarfaraz Molla	Male	Presidency University	Presidency University	Less than 2 hours
17	Srija Acharjee	Female	Vidyasagar College	University of Calcutta	5-8 hours
18	Subhadip panja	Male	Maulana Azad College	University of Calcutta	Less than 2 hours
19	Niladri Banerjee	Male	Maulana Azad College	University of Calcutta	2-5 hours
20	Atreyee Bose	Female	Maulana azad college	University of Calcutta	Less than 2 hours
21	Arshal Baskey	Male	Maulana Azad College	University of Calcutta	Less than 2 hours
22	Daipayan Bhattacharyya	Male	Maulana Azad College	University of Calcutta	Less than 2 hours
23	Aniruddha Mukherjee	Male	Surendranath college	University of Calcutta	5-8 hours
24	Shibaji Das	Male	Maulana Azad College	University of Calcutta	2-5 hours
25	Ahan Mazumdar	Male	Maulana Azad College	University of Calcutta	5-8 hours

Sl. No.	6. Your 1st semester SGPA (If you have backlog and can't provide SGPA, simply write "NA")	7. Your 2nd semester SGPA (If you have backlog and can't provide SGPA, simply write "NA")	8. Your 3rd semester SGPA (If you have backlog and can't provide SGPA, simply write "NA")	9. Your 4th semester SGPA (If you have backlog and can't provide SGPA, simply write "NA")	10. Which range of SGPA you expected yourself to get in the previous semester (4th semester)?	11. Which range of SGPA you expect yourself to get in the next semester (5th semester)?	12. How much time do you spend on mobile weekly on an average? (Aggregated total of mobile phone usage)
1	9.1	6.4	6.8	7.3	Above 7 - Upto 8	Above 7 - Upto 8	Above 30 hours - Upto 50 hours
2	9.33	9.54	9.43	9.77	Above 9	More than 9	Less than 10 hours
3	8.6	8.12	6.19	7.5	Above 7 - Upto 8	Above 7 - Upto 8	Above 30 hours - Upto 50 hours
4	7.82	7.91	6.59	8.86	Above 8 - Upto 9	Above 7 - Upto 8	Above 70 hours - Upto 90 hours
5	9.3	8.6	8.3	8.1	Above 8 - Upto 9	Above 8 - Upto 9	Above 70 hours - Upto 90 hours
6	7.96	8.34	8.24	7.96	Above 8 - Upto 9	Above 8 - Upto 9	Less than 10 hours
7	8.77	7.5	8.01	8.4	Above 8 - Upto 9	More than 9	Above 70 hours - Upto 90 hours
8	9.7	9.4	6.7	6.4	Above 7 - Upto 8	Above 7 - Upto 8	Less than 10 hours
9	9.399	7.608	8.38	7.713	Above 7 - Upto 8	Above 7 - Upto 8	Above 70 hours - Upto 90 hours
10	9.46	9.44	8.59	8.81	Above 8 - Upto 9	Above 8 - Upto 9	Above 30 hours - Upto 50 hours
11	9.02	NA	6.23	5.84	Above 5 - Upto 6	Above 6 - Upto 7	Above 30 hours - Upto 50 hours
12	9.1	9.03	9.1	9.3	Above 7 - Upto 8	Above 8 - Upto 9	Above 50 hours - Upto 70 hours
13	9.03	6.05	6.13	NA	Above 6 - Upto 7	Above 6 - Upto 7	Above 10 hours - Upto 30 hours
14	8.6	7.5	7.5	7.5	Above 8 - Upto 9	Above 8 - Upto 9	Less than 10 hours
15	8.7	8.5	8.2	8.4	Above 8 - Upto 9	Above 8 - Upto 9	Less than 10 hours
16	7	6.09	4.29	6.5	Above 5 - Upto 6	Above 6 - Upto 7	Above 30 hours - Upto 50 hours
17	7.913	6.981	6.672	6.65	Above 7 - Upto 8	Above 6 - Upto 7	Above 30 hours - Upto 50 hours
18	9.02	NA	Na	Na	Above 6 - Upto 7	Above 6 - Upto 7	Above 10 hours - Upto 30 hours
19	9.3	5.6	6.2	6	Above 7 - Upto 8	Above 7 - Upto 8	Above 30 hours - Upto 50 hours
20	9.6	7.4	7.2	7.8	Above 7 - Upto 8	Above 7 - Upto 8	Above 70 hours - Upto 90 hours
21	7.2	NA	5.9	5.4	Above 5 - Upto 6	Above 4 - Upto 5	Less than 10 hours
22	8.992	7.124	6.182	7.175	Above 5 - Upto 6	Above 5 - Upto 6	More than 90 hours
23	9.2	7.1	7.3	7.5	Above 7 - Upto 8	Above 7 - Upto 8	Above 70 hours - Upto 90 hours
24	8.125	6.719	6.952	7.19	Above 6 - Upto 7	Above 6 - Upto 7	More than 90 hours
25	9.6	7.6	7.2	8	Above 8 - Upto 9	Above 7 - Upto 8	Less than 10 hours

Sl. No.	13. How much time do you spend on mobile for study weekly on an average ?	14. How much time do you spend on mobile for gaming weekly on an average ?	15. How much time do you spend on mobile for social media usage weekly on an average ?	16. How much time do you spend on mobile for chatting weekly on an average ?
1	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Above 45 hours - Upto 65 hours	Above 5 hours - Upto 25 hours
2	Less than 5 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours
3	Above 5 hours - Upto 25 hours	Less than 5 hours	Above 5 hours - Upto 25 hours	Above 45 hours - Upto 65 hours
4	Less than 5 hours	Less than 5 hours	Above 45 hours - Upto 65 hours	Above 45 hours - Upto 65 hours
5	Less than 5 hours	Above 25 hours - Upto 45 hours	Above 5 hours - Upto 25 hours	Above 25 hours - Upto 45 hours
6	Less than 5 hours	Less than 5 hours	Above 5 hours - Upto 25 hours	Less than 5 hours
7	Above 45 hours - Upto 65 hours	Less than 5 hours	Above 25 hours - Upto 45 hours	Above 25 hours - Upto 45 hours
8	Above 25 hours - Upto 45 hours	Less than 5 hours	Above 25 hours - Upto 45 hours	Above 5 hours - Upto 25 hours
9	Above 25 hours - Upto 45 hours	Above 25 hours - Upto 45 hours	Above 25 hours - Upto 45 hours	Above 5 hours - Upto 25 hours
10	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours
11	Above 5 hours - Upto 25 hours	Less than 5 hours	Less than 5 hours	Above 5 hours - Upto 25 hours
12	Less than 5 hours	Above 5 hours - Upto 25 hours	Above 25 hours - Upto 45 hours	Above 45 hours - Upto 65 hours
13	Above 25 hours - Upto 45 hours	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Less than 5 hours
14	Above 5 hours - Upto 25 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours
15	Less than 5 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours
16	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Above 25 hours - Upto 45 hours	Above 5 hours - Upto 25 hours
17	Above 5 hours - Upto 25 hours	Less than 5 hours	Above 45 hours - Upto 65 hours	Above 5 hours - Upto 25 hours
18	Above 5 hours - Upto 25 hours	Less than 5 hours	Above 5 hours - Upto 25 hours	Above 25 hours - Upto 45 hours
19	Above 5 hours - Upto 25 hours	Less than 5 hours	Less than 5 hours	Above 5 hours - Upto 25 hours
20	Less than 5 hours	Less than 5 hours	Above 25 hours - Upto 45 hours	Less than 5 hours
21	Less than 5 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours
22	Above 5 hours - Upto 25 hours	Less than 5 hours	Above 65 hours - Upto 85 hours	Above 45 hours - Upto 65 hours
23	Less than 5 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours
24	Less than 5 hours	Less than 5 hours	Above 5 hours - Upto 25 hours	More than 85 hours
25	Above 5 hours - Upto 25 hours	Less than 5 hours	Less than 5 hours	Less than 5 hours

Sl. No.	17. How much time do you spend on mobile for watching movies/videos weekly on an average ?	18. How much time do you spend on mobile for extra curricular/hobby related usage weekly on an average ?	19. Which kind of network service/s do you use for internet at your residence?	20. How much money do you spend on mobile recharge monthly on an average?
1	Less than 5 hours	Less than 5 hours	Mobile Data	Above 200 Rs.- Below 300 Rs.
2	Less than 5 hours	Less than 5 hours	Wi-fi through Broadband / Optical Fibre	Less than 50 Rs.
3	Above 5 hours - Upto 25 hours	Above 45 hours - Upto 65 hours	Both	More than 300 Rs.
4	Less than 5 hours	Less than 5 hours	Mobile Data	Above 200 Rs.- Below 300 Rs.
5	Less than 5 hours	Less than 5 hours	Wi-fi through Broadband / Optical Fibre	More than 300 Rs.
6	Less than 5 hours	Less than 5 hours	Mobile Data	Above 100 Rs.- Below 200 Rs.
7	Above 5 hours - Upto 25 hours	Less than 5 hours	Both	Above 200 Rs.- Below 300 Rs.
8	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Both	Above 200 Rs.- Below 300 Rs.
9	Above 5 hours - Upto 25 hours	Less than 5 hours	Both	Above 200 Rs.- Below 300 Rs.
10	Above 25 hours - Upto 45 hours	Less than 5 hours	Both	Above 200 Rs.- Below 300 Rs.
11	Above 5 hours - Upto 25 hours	Less than 5 hours	Wi-fi through Broadband / Optical Fibre	Above 100 Rs.- Below 200 Rs.
12	Above 45 hours - Upto 65 hours	Less than 5 hours	Both	Above 200 Rs.- Below 300 Rs.
13	Less than 5 hours	Above 5 hours - Upto 25 hours	Both	Above 200 Rs.- Below 300 Rs.
14	Less than 5 hours	Less than 5 hours	Both	Above 200 Rs.- Below 300 Rs.
15	Less than 5 hours	Less than 5 hours	Both	Less than 50 Rs.
16	Above 25 hours - Upto 45 hours	Less than 5 hours	Mobile Data	Above 100 Rs.- Below 200 Rs.
17	Above 5 hours - Upto 25 hours	Less than 5 hours	Both	More than 300 Rs.
18	Less than 5 hours	Less than 5 hours	Mobile Data	Above 100 Rs.- Below 200 Rs.
19	Less than 5 hours	Less than 5 hours	Mobile Data	Above 200 Rs.- Below 300 Rs.
20	Above 45 hours - Upto 65 hours	Above 45 hours - Upto 65 hours	Both	Above 100 Rs.- Below 200 Rs.
21	Less than 5 hours	Less than 5 hours	Mobile Data	Above 200 Rs.- Below 300 Rs.
22	Above 45 hours - Upto 65 hours	Above 5 hours - Upto 25 hours	Mobile Data	Above 200 Rs.- Below 300 Rs.
23	Above 5 hours - Upto 25 hours	Less than 5 hours	Wi-fi through Broadband / Optical Fibre	Above 100 Rs.- Below 200 Rs.
24	Above 5 hours - Upto 25 hours	Above 5 hours - Upto 25 hours	Both	More than 300 Rs.
25	Less than 5 hours	Less than 5 hours	Both	Above 100 Rs.- Below 200 Rs.

Sl. No.	21. Does your monthly recharge satisfy your requirements?	22. How do you feel usage of mobile phone impact on your study/result?	23. How much time do you spend on sports weekly on an average? (If you are not into sports, simply select the option "NO")	24. How do you feel sports impact on your study/result?	25. How much time do you spend on extra curricular activities weekly on an average? (If you are not involved in extra curricular activities, simply select the option "NO")	26. How do you feel extra curricular activities impact on your study/result?	27. How much time do you sleep daily on an average?
1	No	4	NO	5	NO	4	Above 8 hours - Upto 10 hours
2	Yes	4	Less than 10 hours	5	NO	4	Above 8 hours - Upto 10 hours
3	Yes	3	NO	3	Above 10 hours - Upto 30 hours	5	Above 6 hours - Upto 8 hours
4	No	3	NO	1	Less than 10 hours	3	Above 6 hours - Upto 8 hours
5	No	1	Less than 10 hours	4	Less than 10 hours	3	Above 2 hours - Upto 4 hours
6	Yes	2	Less than 10 hours	5	Less than 10 hours	3	Above 6 hours - Upto 8 hours
7	Yes	3	Less than 10 hours	5	Above 10 hours - Upto 30 hours	5	Above 6 hours - Upto 8 hours
8	Yes	1	NO	4	Less than 10 hours	5	Above 6 hours - Upto 8 hours
9	Yes	3	NO	3	NO	3	More than 10 hours
10	Yes	3	Less than 10 hours	3	Less than 10 hours	3	Above 6 hours - Upto 8 hours
11	Yes	2	NO	1	Above 10 hours - Upto 30 hours	1	Above 6 hours - Upto 8 hours
12	Yes	1	NO	5	NO	1	Above 4 hours - Upto 6 hours
13	Yes	3	Less than 10 hours	3	Less than 10 hours	4	Above 6 hours - Upto 8 hours
14	Yes	4	Less than 10 hours	3	Less than 10 hours	4	Above 2 hours - Upto 4 hours
15	No	1	Above 10 hours - Upto 30 hours	1	Above 10 hours - Upto 30 hours	2	Above 6 hours - Upto 8 hours
16	Yes	5	NO	4	NO	2	Above 6 hours - Upto 8 hours
17	Yes	3	Less than 10 hours	4	Less than 10 hours	3	Above 8 hours - Upto 10 hours
18	Yes	2	NO	3	Less than 10 hours	3	Above 6 hours - Upto 8 hours
19	Yes	2	Less than 10 hours	4	Less than 10 hours	4	Above 6 hours - Upto 8 hours
20	Yes	3	Less than 10 hours	4	Less than 10 hours	3	Above 6 hours - Upto 8 hours
21	Yes	3	Less than 10 hours	4	NO	3	Above 4 hours - Upto 6 hours
22	Yes	2	NO	3	NO	3	Above 6 hours - Upto 8 hours
23	Yes	3	Less than 10 hours	5	NO	3	Above 6 hours - Upto 8 hours
24	No	3	NO	3	Less than 10 hours	3	Above 6 hours - Upto 8 hours
25	Yes	1	NO	3	Less than 10 hours	3	Above 6 hours - Upto 8 hours