



PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA
KAVATHE MAHANKAL, Dist. Sangli (Maharashtra) Pin- 416 405
Principal Prof. (Dr.) M. K. Patil M.Sc., M. Phil., Ph.D. Mob.9421185277
Phone-02341-295220 Email: kmpvp@rediffmail.com Website: www.pvpkm.ac.in
Jr. College Index No. J 22.04.002

Criterion-II: Teaching Learning and Evaluation

Key Indicator 2.3 Teaching Learning Process

2.3.1 Student centric methods, such as experiential learning, participative learning and problem-solving methodologies are used for enhancing learning experiences and teachers use ICT- enabled tools including online resources for effective teaching and learning process

PROBLEM SOLVING METHODOLOGIES


RESEARCH PROJECTS

INDEX

Sr. No.	Name of Activity	Problem Solving Projects Conducted by
1.	Statistical Analysis of Academic Performance of Undergraduate Students	B.Sc.-III (Statistics Department)
2.	Statistical Analysis of Road Site Workers	B.Sc.-III (Statistics Department)
3.	Statistical Analysis of Fast-Food Consumption among Students	B.Sc.-III (Statistics Department)
4.	Statistical Analysis of Customer Perception towards OTT Platforms	B.Sc.-III (Statistics Department)
5.	Statistical Analysis Effect of Mobile and Internet on the Health of Children	B.Sc.-III (Statistics Department)
6.	Statistical Analysis Production and Marketing of Grapes	B.Sc.-III (Statistics Department)
7.	Statistical Perception towards Online Shopping	B.Sc.-III (Statistics Department)
8.	Statistical Analysis of Milk Production	B.Sc.-III (Statistics Department)
9.	Statistical Analysis of Demand of Bakery Product	B.Sc.-III (Statistics Department)
10.	Statistical Analysis of Bike Users	B.Sc.-III (Statistics Department)
11.	Consumer Perception towards Himalaya Products	B.Sc.-III (Statistics Department)
12.	Impact of Covid-19 on Social Welfare	B.Sc.-III (Statistics Department)
13.	Effect of Online Education System on Quality of Education	B.Sc.-III (Statistics Department)
14.	Statistical Analysis of Effect of Spending time on Computer/ Mobile Platforms and On-Screen Activities in Real Life	B.Sc.-III (Statistics Department)
15.	Role of Media on Human Development	B.Sc.-III (Statistics Department)

16.	To Study the Challenges of Online Education Over Traditional Education	B.Sc.-III (Statistics Department)
17.	To Study the Relation Between Age, Weight and Blood Presser	B.Sc.-III (Statistics Department)
18.	Roles of Fertilizers on Grapes Production	B.Sc.-III (Statistics Department)
19.	Statistical Analysis of Effect of Fertilizers and Soil Type on Sugarcane	B.Sc.-III (Statistics Department)
20.	To Study the time Spent by Student on social media	B.Sc.-III (Statistics Department)




 -PRINCIPAL
 Padmabhushan Vasantodada Patil
 Mahavidyalaya, K. Mahankal, Dist. Sangli

Project B.Sc. III In Year 2023-24

Sr. No.	Name of the Students	Title of the Project	Name of Mentor
1	Kute Pranav Kundlik	Statistical analysis of academic performance of undergraduate students	Mr. V. V. Koshti
	Kadam Aniket Mahadev		
	Kore Pratik Jayant		
	Bandagar Vishal Annaso		
2	Mane Laxmi Sahadev	Statistical analysis of Road-Site workers	Dr. M. K. Patil
	Kodag Dhanashri Dattatray		
	Sawant Swapnali Suryakant		
	Ghejji Bhagyashri Ramanna		
	Patil Shweta Basagonda		
3	Chavan Swati Deepak	Statistical analysis of fast-food consumption among students	Miss. S. A. Patil
	Mahajan Ashwini Annaso		
	Choudhary Geeta Dayaram		
	Patil Pratiksha Anandrav		
4	Latthe Rohini Mukesh	Statistical analysis of customer perception towards OTT platforms	Miss. S. M. Zarekar
	Bhosale Sushmita Abaso		
	Pawar Yogita Anil		
	Nikam Snehal Suresh		
	Chougule Yashoda Nathaji		
5	Mali Priyanka Balu	Statistical analysis of effect of mobile and internet on health of children	Miss. S. S. Patil
	Jadhav Aarti Tanaji		
	Shinde Shailaja Dilip		
	Shinde Ashwini Pandurang		
	Patil Akshata Balaso		
6	Devarshi Rahul Pargonda	Statistical Analysis of Production and Marketing of Grapes	Mr. G. D. Satpute
	Olekar Pratik Bandu		
	Suryawanshi Digvijay Sanjay		
	Nagane Akshay Pandurang		

A

Project Report

On

“STATISTICAL ANALYSIS ON CUSTOMER PREFERENCE
TOWARDS OTT PLATFORM”

Submitted To

(Department Of Statistics)

Shikshan Prasarak Sanstha's

PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA,
KAVATHE MAHANKAL

By

Miss.Latthe Rohini Mukesh

Miss. Pawar Yogita Anil

Miss.Bhosale Sushmita Abaso

Miss. Nikam Snehal Suresh

Miss.Chougule Yashodha Nathaji

Under the Guidance of

Asso.Koshti V.V.(Head of dept.) Mr.

G. D. Satpute

Mrs. Patil S.S.

Miss.Zarekar S.M.

Through

The principal

Prof.Dr.M.K.Patil.PADMABHUSHAN VASANTRAODADA PATIL
MAHAVIDYALAYA, KAVATHE MAHANKAL.

For the academic year 2023-24

ShikshanPrasarakSanstha's

**PADMABHUSHANVASANTRAODADAPATIL
MAHAVIDYALAYA, KAVATHE MAHANKAL**

(DepartmentOfStatistics)

CERTIFICATE

This is to certify that the following students are working under my Guidance.They have satisfactorily carried out and completed the project Work for the B.Sc Part III. They are presently submitted their project entitled.

This work is genuine, and the project or the part of it has not has submitted for the award of any other degree.

“Statistical Analysis on Customer Preference Towards OTT Platform”

Sr.No.	Name	PRN No.
1	Miss.LattheRohiniMukesh	2021019957
2	Miss.PawarYogitaAnil	2021020096
3	Miss.BhosaleSushmitaAbaso	2021020505
4	Miss.NikamSnehalSuresh	2021010322
5	Miss.ChouguleYashodhaNathaji	2021020076



Teacher in Charge



Examiner



Head Department of statistics

ACKNOWLEDGEMENT

We have great pleasure while submitting this project report on “Statistical Analysis on Customer Preference Towards OTT Platform” in practical fulfilment of B.Sc III. We thank Prof. Dr. M.K. Patil. for doing this project.

This report has been prepared under the guidance of Prof. Koshti V.V. (Head of department). We would like to express our profound gratitude towards him for his guidance and timely support throughout the completion of this project .

We would like to express our higher regards to our department of statistics, and also appreciate our respected teacher Asso. G.D. Satpute, Mrs. Patil S.S., Miss. Zarekar S.S. Also, we would like to thank all non-teaching staff for their help and co-operation. In the end, we express our respect to our friends who inspire and support us in this work.

INDEX

Sr.No.	NameOfTopic	PageNo.
1	Introduction.	04
2	Objectives	05
3	Methodofdata collection.	06
4	Statisticaltools.	07
5	Analysisofdata.	08-13
6	Overallconclusion.	14
7	References	15
8	Questionnaire.	16

INTRODUCTION



Over-the-top (OTT) in 2008 Reliance entertainment started the first dependent Indian OTT platform. OTT platform are where people can have access over the content via internet. OTT platform eliminate the factor of visiting theatres as everything is available at our comfort zone. There is a common talk that, because of OTT there is a smaller number of people have cable access. OTT platforms have influenced a lot of viewers because of easy accessibility through internet in our comfort zone.

In this paper a study on customer preference towards OTT. Analysis of the consumption includes determination of the OTT platforms used. Time spent on them and their experience. Consumers can watch OTT content on phones, smart TVs, set-top boxes, tablets, desktop and laptop computers.

We study how many respondents are using OTT platforms also we determine factor influencing; we check which app is more used by respondents. OTT platform is neither inherently bad nor good it all depends on how it is being used.

OBJECTIVES

1. To recognize popular OTT platforms and contents preference of viewers in OTT platform.
2. To study the comparison between cinemahall and OTT platform.
3. To make a study on the the impact of COVID-19 pandemic on the OTT platform.
4. To analyse age Group Wise classification of respondents to OTT platform.
5. To measure the pattern of use of OTT platform among customers.
6. To find the major OTT platform subscribed by the customers.
7. To check dependency between profession and times pendon OTT platform.
8. To analyselanguage preference of respondents to OTT platform.
9. To find the ratings of the OTT platform be ston experience.
10. To check dependency between gender and opinion about the OTT platform.

SAMPLESIZEDETERMINATION

The population size of Kavathe-Mahankal Tehsil is 152327 according to census of 2011. We used Yamane's formula to determine the sample size at approximately 6% margin of error and we got the sample size approximately 298. The Yamane's formula for the sample size determination is given by,

$$n = \frac{N}{Ne^2 + 1}$$

Where, n is sample size

N is the population size

E is the margin of error.

DATA COLLECTION

The main objective of the project is "To analysis customer preference towards OTT platforms." We have collected sample of various age group of people from different areas of Kavathe Mahankal tehsil. We prepare a questionnaire to fulfill the objectives behind the project accurately. The data collection procedure was done using google forms and direct interviewing with people.

STATISTICAL TOOLS

1) Graphical representation:

❖ Piechart

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size the data, and it is type of pictorial representation of data. A pie chart requires list of categorical variables and numerical variable. Here the term pie represents the whole. All the slices of the pie chart add up to make whole equalling 100% and 360 degree.

❖ Bar chart

A bar chart is a graphical representation used to display and compare discrete categories of data through rectangular bars, where the length or height of each bar is proportional to the frequency or value of the corresponding category or strips. It can be either horizontal or vertical.

❖ Donut chart

A donut chart is a graphical representation used to show proportions of categories that make up the whole, but the center can also be used to display data.

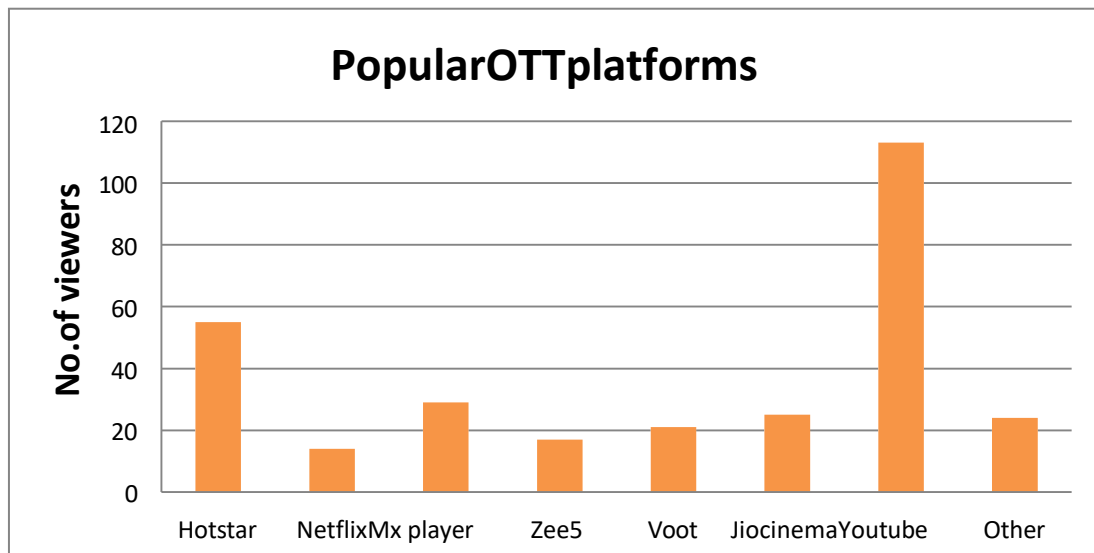
2) Hypothesis testing:

❖ Chi-square test for independence of attributes:

The null hypothesis of this test is that the attributes are independent on each other. The test is used to check the independency of the attributes. Observed frequencies are tabulated in the contingency table and the expected frequencies are calculated. The test checks whether the observed frequencies and expected frequencies are significantly different or not.

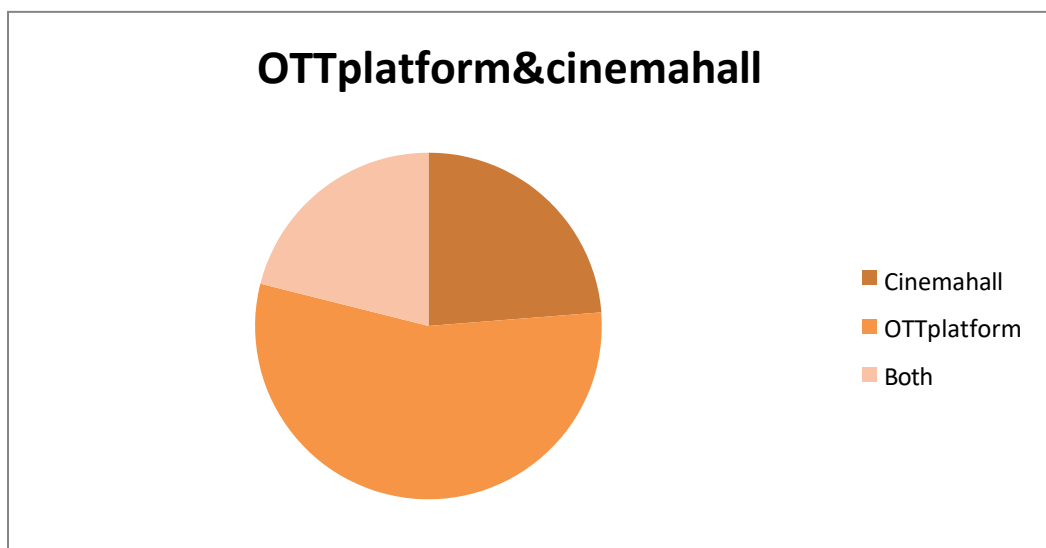
STATISTICAL ANALYSIS

1) To recognize popular OTT platforms and contents preference of viewers in OTT platform.



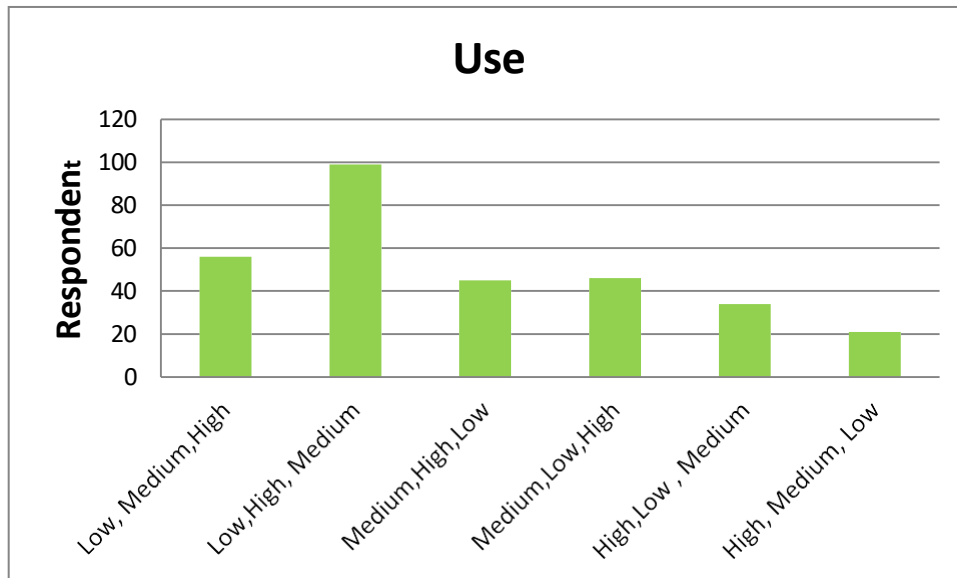
Interpretation: Most popular OTT platform is Youtube and people give less preference to Netflix.

2) To study the comparison between cinemahall and OTT platform.



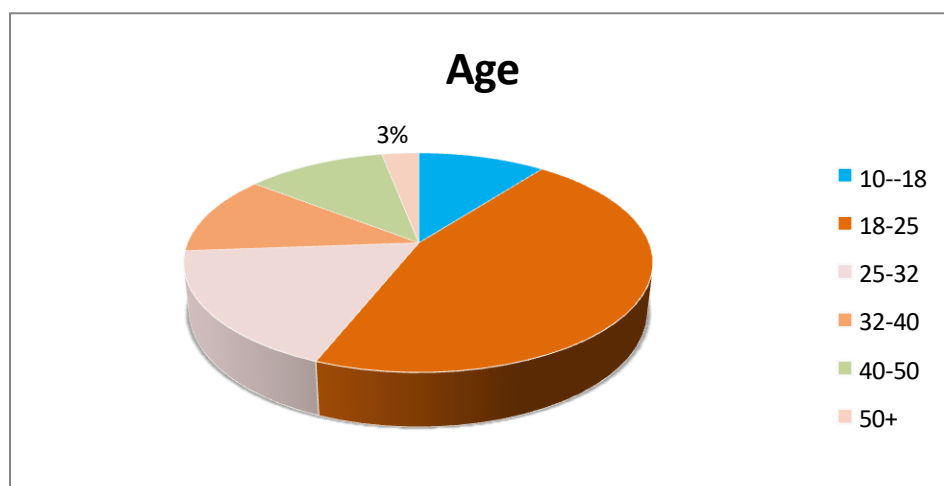
Interpretation: People prefer OTT platforms than cinemahall.

3) To make a study on the impact of COVID-19 pandemic on the OTT platform.



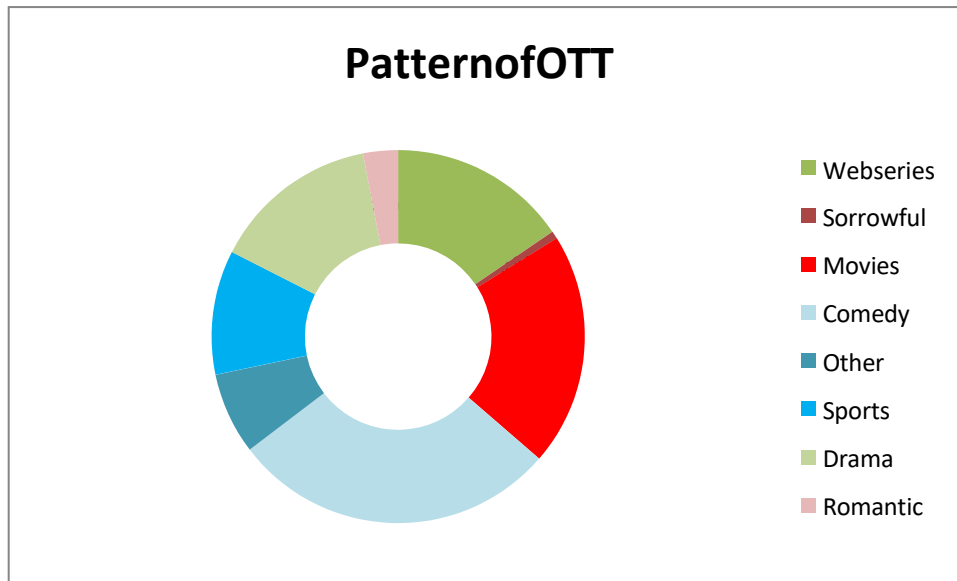
Interpretation: From graph we can observe that people uses low OTT platforms before COVID-19, high in COVID-19 and medium after COVID-19.

4) To analyse age Group Wise classification of respondents to OTT platform.



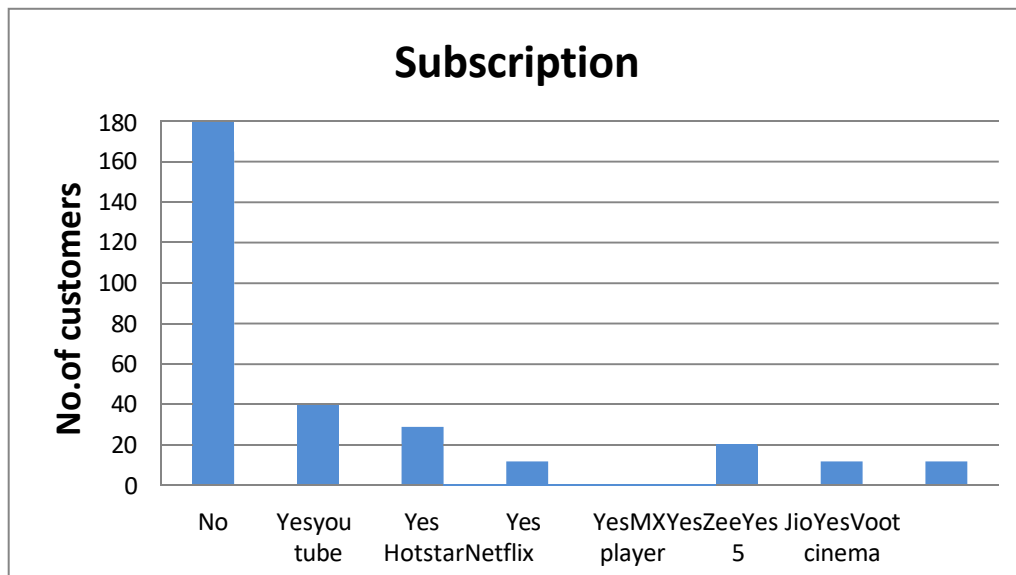
Interpretation: In 18-25 age group people more uses OTT platforms and 50+ age group less uses OTT platform.

5) To measure the pattern of use of OTT platform among customers.



Interpretation: From the given data, people use OTT platform to watch for comedy and then for movies.

6) To find the measures OTT platform subscribed by the customers.



Interpretation: Most of the people don't want to give any subscription.

7) To check dependency between profession and time spend on OTT platform.

H_0 : The relationship between profession and time spend on OTT platform is independent.

H_1 : The relationship between profession and time spend on OTT platform is dependent.

The contingency table is,

	Student	Employee	Homeworker	Others	Total
30min	49	16	10	5	80
1hr	47	26	15	7	95
2hr	26	29	9	11	75
More than 2	29	7	4	8	48
Total	151	78	38	31	298

The tabulated value is,

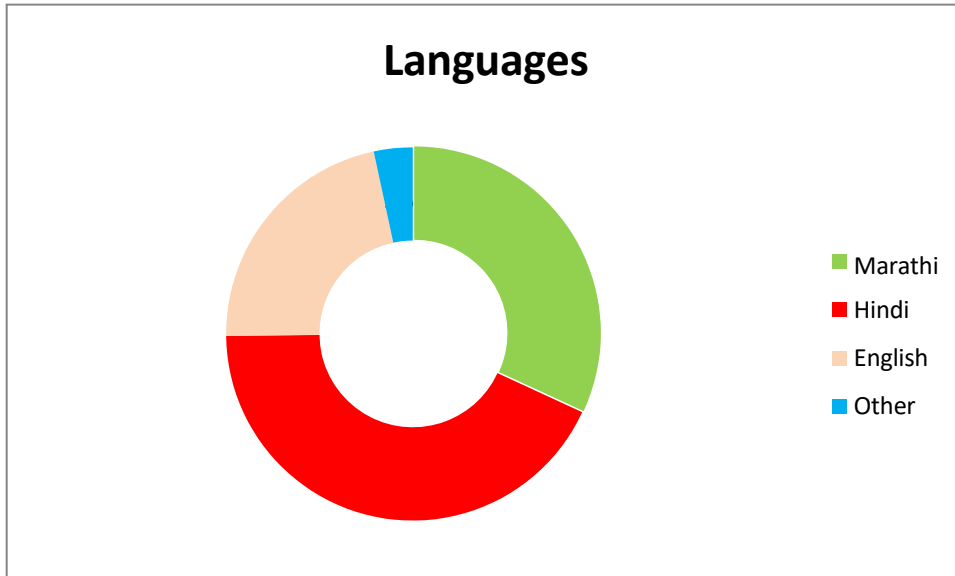
$$\chi^2 = 16.92$$

$$\begin{aligned} \text{Calculated } \chi^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= 18.1307 \end{aligned}$$

Therefore we reject the Null hypothesis at 5% level of significance.

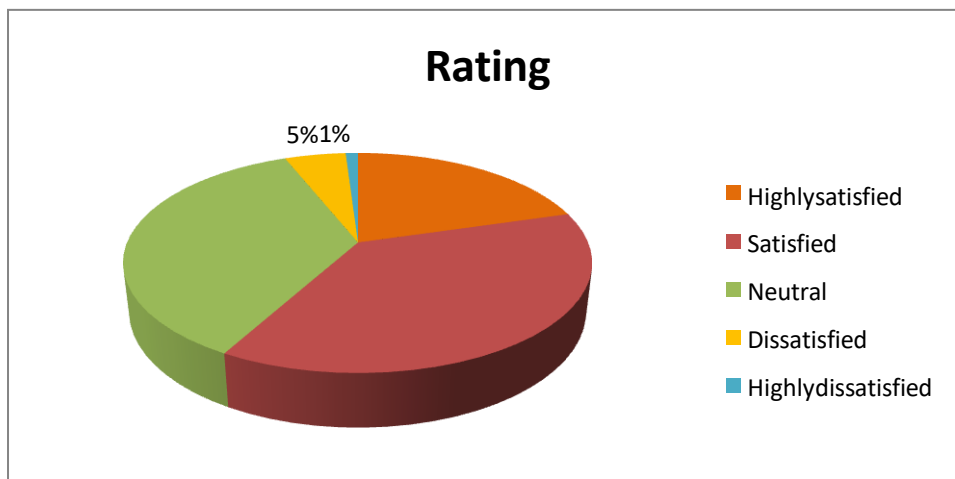
We conclude that, the data provides that time spend is dependent on profession.

8) To analyse language preference of respondentst ott platform.



Interpretation:Out of 298 peoples,128 people choose Hindi language. And also Marathi language are preferable among them.

9) To find theratings of the OTT platform beston experience.



Interpretation:There are 38% people are satisfied by using OTT platform.

10) To check dependency between gender and opinion about the OTT platform.

H_0 : The relationship between gender and opinion about OTT platform is independent.

H_1 : The relationship between gender and opinion about OTT platform is dependent.

The contingency table is,

	Yes	No	Total
Male	132	15	147
Female	140	11	151
Total	272	26	298

The tabulated value is,

$$\chi^2 = 3.841$$

$$\begin{aligned} \text{Calculated } \chi^2 &= \frac{N(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)} \\ &= \frac{298(1452-2100)^2}{147 \times 151 \times 26 \times 272} \\ &= 0.7971 \end{aligned}$$

Therefore we accept the Null hypothesis at 5% level of significance.

We conclude that, the data provides no evidence to claim that opinion is independent on the gender.

CONCLUSION

- The most popular OTT platform among people is YouTube.
- The most of the people still prefer to OTT platform than cinema hall.
- People use low OTT platforms before COVID-19, high in COVID-19 and medium after COVID-19.
- In 18-25 age group people more use OTT platforms.
- Mostly people use OTT platform to watch for comedy.
- Most of the people don't want to give any subscription.
- People's time spend is dependent on their profession.
- Out of 298 people, 128 people choose Hindi language. And also Marathi language is preferable among them.
- There are 38% people are satisfied by using OTT platform.
- People's opinion about OTT platform is independent on gender.

LIMITATIONS

- This study is limited to 298 respondents. So finding and suggestions given on the basis of the study cannot be assumed to the entire population.
- The study also restricts itself to the geographical area of the Kavathe Mahankal tehsil

SCOPE

- OTT Media Research and data on OTT players Regularly build on data with regards to the performance and reach of OTT players in India.
- Constantly review newly launched shows, movies,news channels interms of view, reachandaudiences engagement parameters.

REFERENCES

- Fundamentals of mathematical statistics by Gupta and Kapoor.
- 100 Statistical Tests by Gopalk. kanji.
- Study on customer preference towards OTT platform during a pandemic by PAVITHRA.D (April 2022)

QUESTIONNAIRE

- Age
 - A) 10-18 B)18-25 C)25-32
 - D) 32-40 E)40-50 F)Morethan 50
- Gender
 - A) Male B)Female
- WhichOTTplatform doyouprefer themost ?
 - A) Hotstar B)Netflix C) Mxplayer D)Zee5
 - E) Voot F)Jio cinemaG)Youtube H)Other
- WhichlanguagedoyouprefertowatchinOTT?
 - A) Marathi B) Hindi
 - C) English D)Other
- Do youthinkwatchinmoviesinOTTischeaperandconvenientthanothertraditional medium ?
 - A) Yes B)No
- Ina future,ifa moviereleasedincinema hallor onOTT together, whatwouldyou prefer mostly ?
 - A) Cinemahall B) OTT platform C) Both
- Do youatisfyonOTTplatform ?
 - A) Highlysatisfied B)Satisfied C) Neutral
 - D) Dissatisfied E)Highlydissatisfied
- Howmuchtimedoyou spendonOTTplatform ?
 - A) 30 minute B)1hour C)2 hour D)Morethan 2hour
- WhatdoyouwanttowatchonOTTplatform?
 - A) Drama B) Romantic C) Sorrowful D)Movies
 - E) Web series F) Sports G)Other
- Howwas theusageon OTTplatformbeforeCovid-19, in Covid-19 and howis it after Covid-19 ?
 - A) Low,Medium,High B)Low,High, Medium
 - C)Medium,High,Low D)Medium,Low,High
 - E)High,Low,Medium F)High,Medium, Low
- Didyoubuythesubscription?Ifyes,thenforwhichOTTplatform

A project report on,
**“A Study of Socio Economic Condition and
Health Hazard among
Road Side Workers in Jath City”**

Submitted to,
COLLEGE PROJECT

Organised by,
DEPARTMENT OF STATISTICS

By
**Miss. Dhanashri Dattatray Kodag.
Miss. Laxmi Sahadev Mane.
Miss. Swapnali Suryakant Sawant.
Miss. Shweta Basagonda Patil.
Miss. Bhagyashri Ramanna Ghejji.**

Under the Guidance of,
**Mr. V. V. Koshti
Mr. G. D. Satpute
Mrs. S. S. Patil
Miss. S. M. Zarekar**

Through
**Prof. Dr. M. K. Patil
The principal,
Padmabhushan Vasanttraodada Patil Mahavidyalaya, KavatheMahankal.**

ACKNOWLEDGEMENT

We have great pleasure while submitting this project report on “A Study of Socio Economic Condition and Health Hazard among Road Side Workers in Jath City” in practical fulfilment of B.Sc. III. We thank Prof.Mr.V.V. Koshti (HOD, Statistics Department) for giving permission for doing this project.This report has been prepared under the guidance of Mrs. S.S. Patil . We would like to express our profound gratitude towards him for his guidance and timely support throughout the completion of this project.We would like to express our higher regards to our department of statistics, and also appreciate our respected teachers Mr. G. D. Satpute , Mrs. S. M. Zarekar . Also, we would like to thank all non-teaching staff for their help and co-operation. In the end, we express our respect to our friends who inspire and support us in this work.

ABSTRACT

Construction works on high speed roads in mountainous landscape can be considered as exceptionally dangerous. During construction period, sometimes people have to work in the middle of the lanes with heavy and harmful equipment and materials so that workers should have to be very cautious while working in these areas. Since construction plays a vital role to the development of the nation, the rate of accident can be reduced if sufficient attention for the safety is provided at the construction site. This research was carried out to analyze the health and safety of casual workers and to verify whether the clauses mentioned in the contract regarding workplace health and safety has been followed by the contractors at project location in construction projects in Road Improvement Project (Phase-II). Safe drinking water, first-aid equipment, sanitary facilities, provision of personal protection equipment (PPE) and training of casual workers on safety procedures, were the measures needed for addressing the health and safety facilities confronting casual workers on construction sites. Inadequate instructions concerning work are ranked as the main cause of the accident on a construction site. Safe drinking water and safety signs are ranked as the most important health and safety facility that need to be provided on site to the casual workers.

INDEX

Sr. No.	Content	Page No.
1	Introduction	3
2	Objectives	4
3	Sample size Determination	5
4	Data Collection	6
5	Statistical tools	7- 8
6	Statistical Analysis	9- 19
7	Conclusion	20
8	Limitations	21
9	Scope	22
10	References	23
11	Quationnarie	24- 25

INTRODUCTION

Jath is a town and taluka headquarters in Miraj subdivisions of Sangli district in southern Maharashtra. Total area is 2258.28 km² (871.93 sq mi). The total population is 3,27,747 as per 2011 census. The population density is 150/km² (380/sq mi). According to the 2011 census, Jath taluka has a the male population is 1,68,493 and the female population is 1,59,254. The sex ratio of Jath taluka is 945 females per 1000 males. The Jath taluka has an average literacy rate of 71.63%. Road site are the most visible section of the informal economy. Large-scale spatial planning and urban development initiatives like road infrastructure have become increasingly popular in cities worldwide. These projects improve connectivity, increase cities competitiveness, and attract investment, When inhaled into the human body, silica can lead to lung cancer and other respiratory tract infections such as silicosis, cough, cold, etc. These diseases impact human health greatly and can have even worsening effects on people with underlying health conditions such as asthma, which may lead to the loss of life. On the other hand, road construction is linked to environmental pollution and risks, which impact people's health and well-being in urban and rural areas. Also, the varying implications of road construction activity on residents, commuters, traders, hawkers, and on-site workers on the road strongly correlate with the place's socioeconomic activity. Traders along the stretch of road construction activity are displaced to make way for the construction to go on as when the road is completed, it will improve living conditions and trading conditions; but the displacement of traders and hawkers during road construction greatly impacts their socio-economic lives. According to a recent report by the World Resources Institute (a Maharashtra government knowledge partner), building dust is responsible for 71% of Mumbai's air pollution. Air, water, and noise pollution result from these pollutants, causing irreversible damage to human health, with a concerning link between pollutants and cancer. The study aim is to explore the level of socio economic condition and health hazards experienced by the road site workers in the Jath Taluka. Total population of Road site workers in Sangli is 65000 and Total population of Road site Workers in Jath is 1358 we take response of 309 women and men Road site workers in Jath.

OBJECTIVES

- To study the working condition of Road site workers in Jath.
- To study Socio Economic condition among road site workers in Jath.
- To study health hazard of Road site workers in Jath.

SAMPLE SIZE DETERMINATION :-

The population size of Road site workers in Sangli is 65000 according to the given survey. According to this survey population size of Road site workers in Jath 1358. We used Yamane's formula to determine the sample size at approximately 5% margin of error and we got the sample size approximately 309. The Yamane's formula for the sample size determination is given by,

$$n = \frac{N}{1 + Ne^2}$$

Where, n is sample size

N is the population size

E is the margin of error.

DATA COLLECTION

The design adapted for this study is exploratory type. The population of Road site workers in Sangli city is 65000 and The population of Road site workers in Jath is 1358 . To study the working condition a Study of Socio EconomicCondition and Health Hazard among Road Side Workers in Jath City.

We collect the data of Road site Works by cluster sampling method to visiting the house of Road site Workers . The present study is based on primary data. The primary data is collected through field work. The interview of 309 Workers conducted and data is collected from Jath. The observations, discussion, interview and questionnaire technique is applied.

STATISTICAL TOOLS

- **Exploratory Data Analysis:**

Exploratory data analysis is an analysis approach that identifies general patterns in the data. These patterns include outlier and features of the data that might be unexpected. EDA is an important first step in any data analysis.

- **Graphical representation:**

- **Pie chart**

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size the data, and it is type of pictorial representation of data .A pie chart requires list of categorical variables and numerical variables. Here the term pie represents the whole. All the slices of the pie chart add up to make whole equaling 100% and 360 degrees.

- **Bar chart**

A bar chart is a graphical representation used to display and compare discrete categories of data through rectangular bars, where the length or height of each bar is proportional to the frequency or value of the corresponding category or strips. It can be either horizontal or vertical.

- **Line Chart**

Line chart is the graphical tool that is used to represent the trend in the data over the time. Line chart consists of the series of the data points connected by the line segments.

- **Hypothesis testing:**

- **Chi-square test for independence of attributes:**

The null hypothesis of this test is that the attributes are independent on each other. The test is used to check the independency of the attributes. Observed frequencies are tabulated in the contingency table and the

expected frequencies are calculated. The test checks whether the observed frequencies and expected frequencies are significantly different or not.

- **Fisher's Exact test for independence**

Fisher's exact test is a statistical significance test used in the analysis of contingency tables. The null hypothesis is that the relative proportions of one variable are independent of the second variable; in other words, the proportions at one variable are the same for different values of the second variable. The alternative the relative proportions of one variable are not independent of the second variable.

- **Proportion test:**

Let us consider the parameter p of the population proportion. For instance, we might want to know the proportion of males within a total population of adults when we conduct a survey. A test of proportion will assess whether or not a sample from a population represents the true proportion of the entire population.

$$H_0: P=P_0$$

$$H_1: P \neq P_0$$

- MS-Excel
- R- Software

STATISTICAL ANALYSIS:

❖ Exploratory Data Analysis:

Variables	Mean	SD	Median	Minimum	Maximum
Age	39.307	11.5427	38	17	71
Work start	20.6537	3.0937	20	14	32
Hrs. per day	6.3527	1.1685	6	4	9
Workers	2	0.8710	2	1	4
Income	5737.86	1334.783	5700	500	9500
Savings	1707.3	856.1	1610	100	6000

From the survey of road site workers, we have observed that, age of road site workers in between 17 to 71 years and average age is 39.307 years. Age of most of the road site workers is 38 years.

They start work from minimum 6 years and maximum 32 years and average of work starts of road site workers is 20.6537 years and most of road site workers work start from 20 years.

The hours per day working of road site workers are in between 4 hrs to 9 hrs and average working hours per day is 8.362 hrs and most of road site workers work 6 hrs per day.

In the family of road site workers there are minimum 1 and maximum 7 workers are worked in this same occupation. Average number of workers is $1.7799 \approx 2$ per family. Most of families there is single worker working in this occupation.

The income of road site workers is in between 500 Rs to 9500 Rs and average income of road site workers is 5737.86 Rs. Most of road site workers earn 5700 Rs per month.

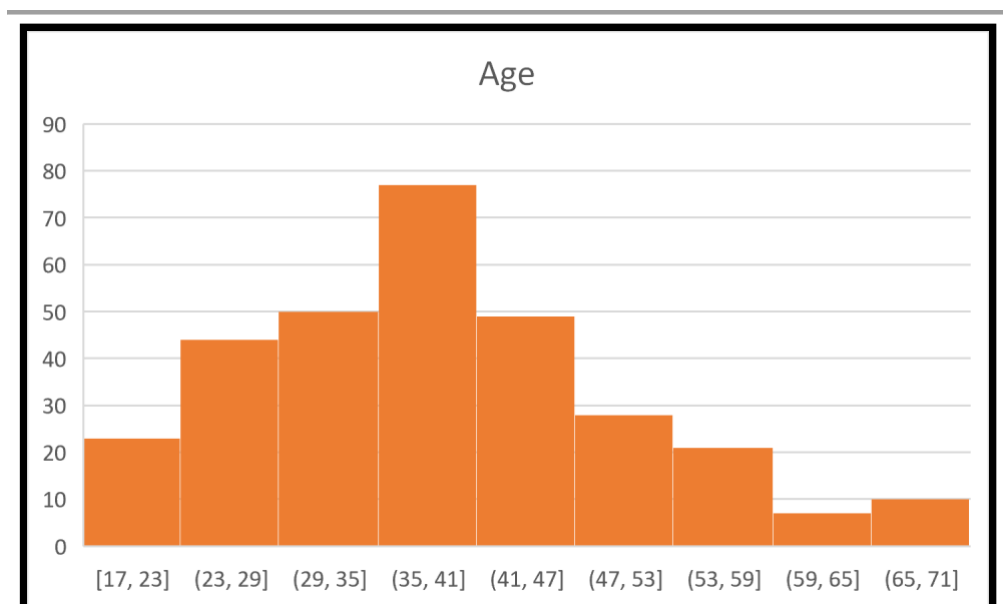
The 50% Road site workers save money for future and save 100 Rs to 6000 Rs per month and average saving of bidi rollers is 1707.3 Rs. The most of bidi rollers save 1610 Rs per month.

Opinion to timely payment	Frequency	Percentage
Yes	247	79.93528
No	62	20.06472
Total	309	1

Around 79.93528 % of workers said that they get timely payment to their wages. While as 20.06472 % said that they don't get it on time but gets delayed by 2 to 3 days.

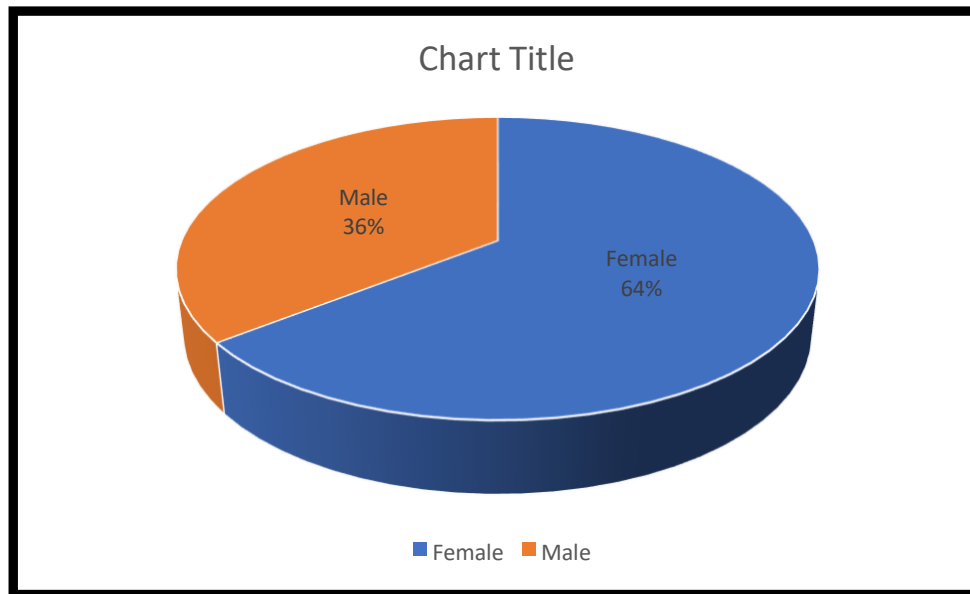
❖ Graphical Representation

1. Bar chart for age of Road site workers



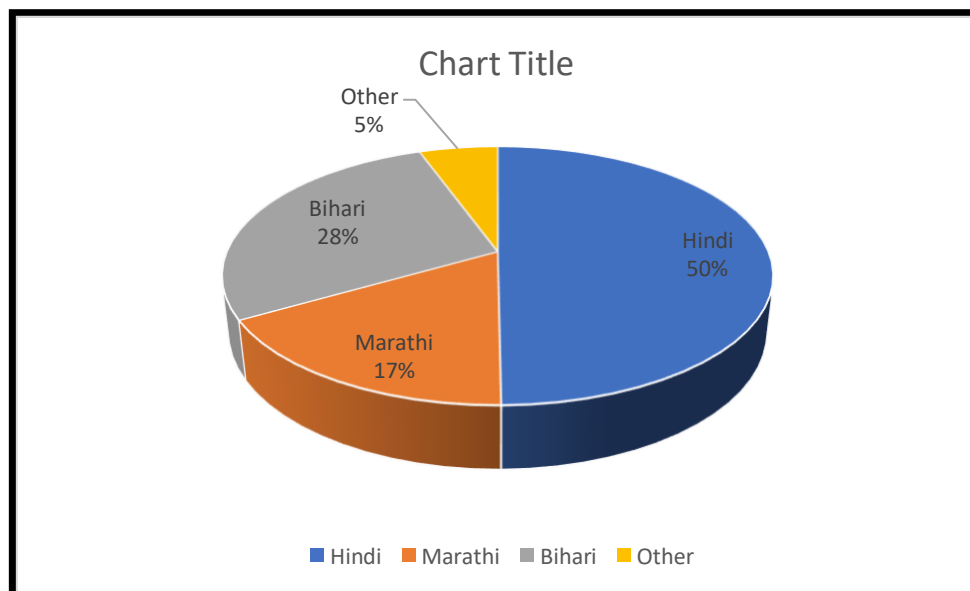
Interpretation: From above bar chart, we interpret that, Age of most of the Road site workers are in between 35 to 41 years.

2. Pie chart for Gender of road site workers



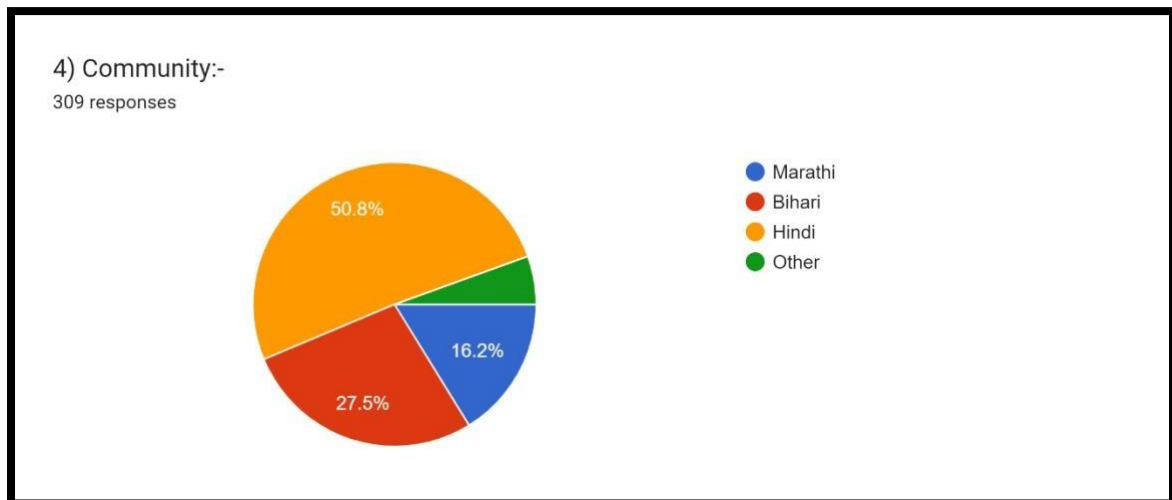
Interpretation: From above pie chart, we interpret that, number of female workers are greater than male workers. The 35.6% male workers and 64.4% female.

3. Pie chart for Language of road site workers



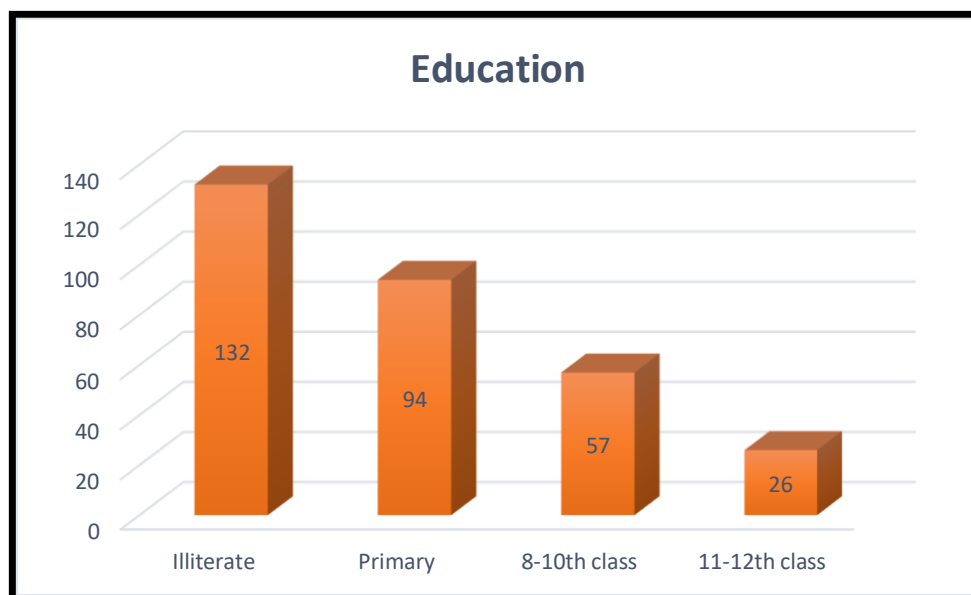
Interpretation: From above pie chart , we interpret that, most of road site workers speak Hindi language.

4. Pie chart for Community of Road site workers



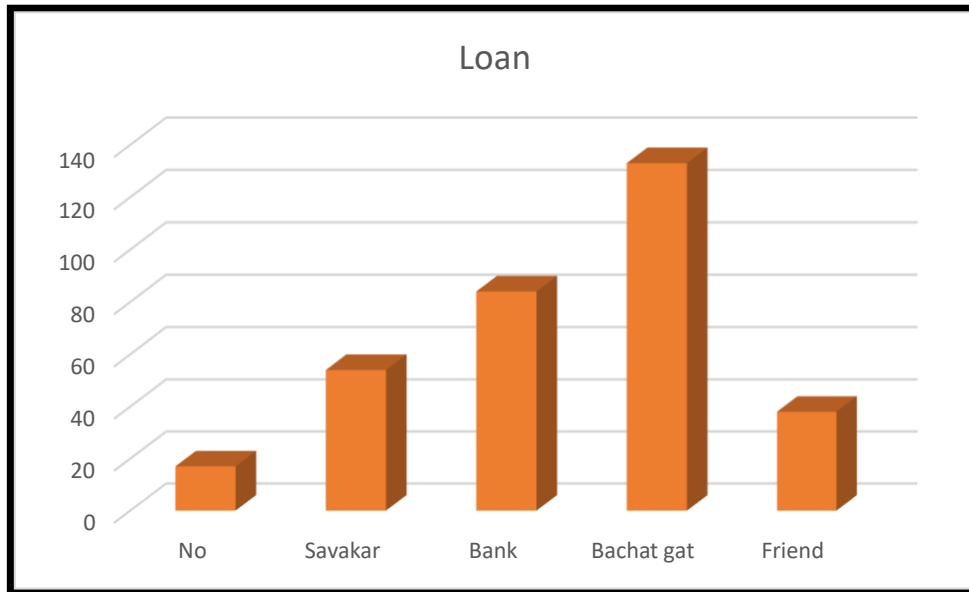
Interpretation: From above pie chart, we interpret that, 50.8% road site workers are of Hindi community.

5. Bar chart for Education of road site workers



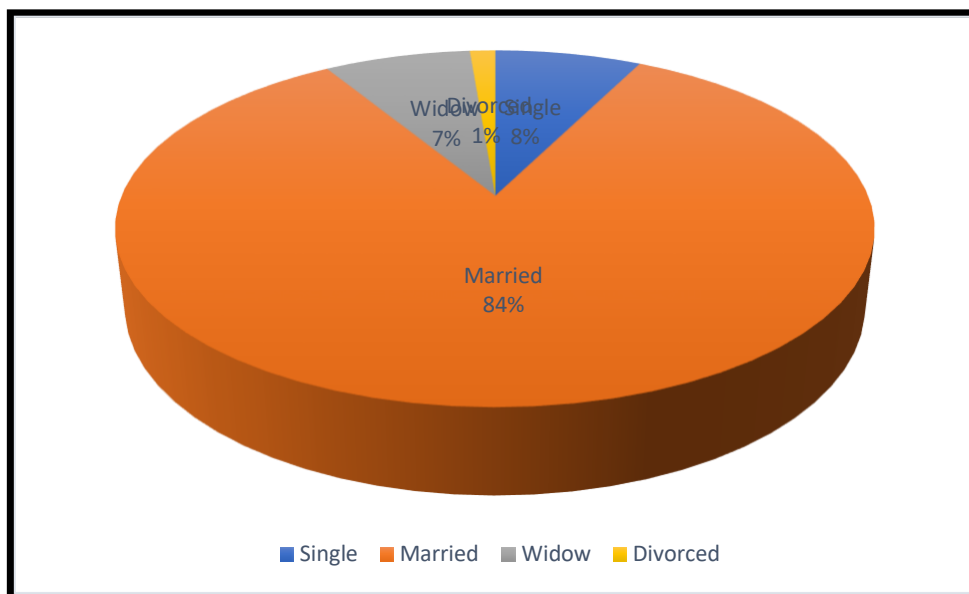
Interpretation: From above for pie chart, we interpret that, illiterate workers are more than the educational workers.

6. Bar chart for loan:



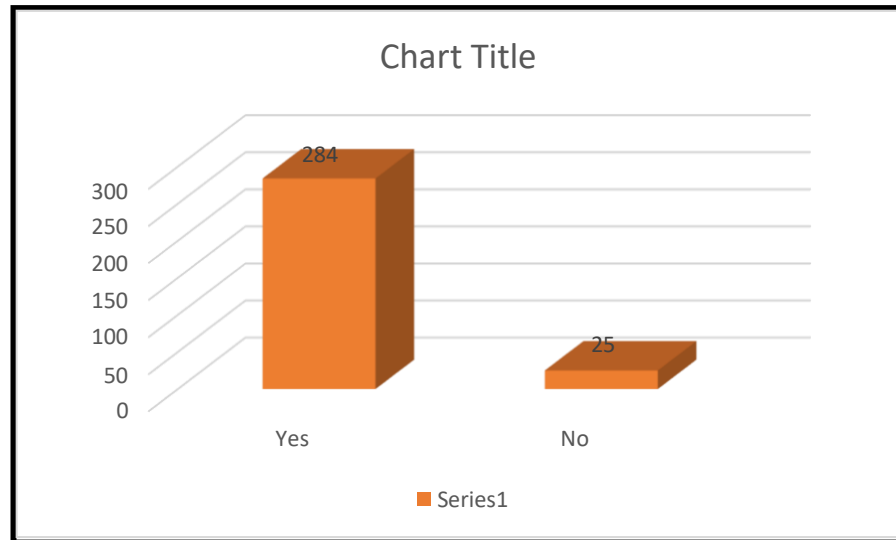
Interpretation: From above chart, we interpret that, greater 50% workers are takes loan and most of worker takes loan from bachat gat.

7. Pie chart for Marital status for road site workers:



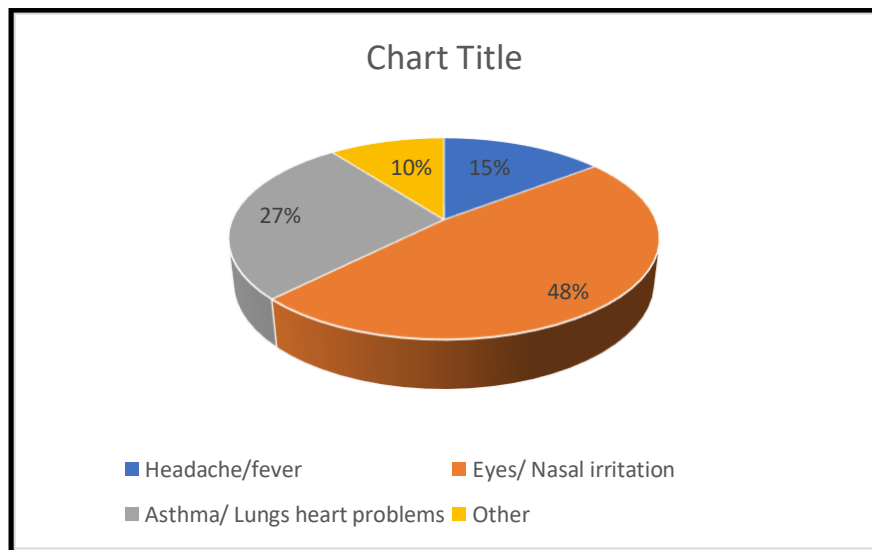
Interpretation: From above pie chart, we interpret that, 84% married, 8% Single, 7% widow, 1% divorced workers .

8. Bar-chart for health problem:



Interpretation: From above chart, we interpret that, the most of the road site workers are suffered from the health problem.

9. Pie-chart for type of different diseases:



Interpretation: From above chart, we interpret that, the most of the road site workers are suffered from disease like Eyes/ nasal irritation.

❖ Fishers's exact test for independence:

Here we want to test,

H_0 : House and type of house of road site workers are independent.

H_1 : House and type of house of road site workers are dependent.

Observation table:

		Type of House		
		Tin/mud/glass	Slab/tile	Total
House	Owned	17	15	32
	Rented	243	4	247
	Total	260	19	279

Procedure:

We apply fisher's exact test using R software we get,

Fisher's Exact Test for Count Data

p-value= 6.777e-13

Alternative hypothesis: true odds ratio is not equal to 1

Sample estimates:

Odds ratio :0.01939014

Here P value 6.777e-13 < 0.05,

Hence we reject H_0

Conclusion :

We conclude that, House and type of house of road site workers are dependent.

❖ **Pearson's Chi-square test for independence:**

Here, we want to test,

H_0 : Loan is independent on marital status of road site workers.

H_1 : Loan is dependent on marital status of road site workers.

Observation table:

		Marital Status		
		Married	Unmarried	Total
Loan	Yes	237	19	256
	No	17	6	23
	Total	254	25	279

Procedure:

By using R software we get,

Pearson's Chi-squared test with Yate's continuity correction

X-squared = 6.8699, df = 1, p-value = 0.008766

Here, p-value is $0.008766 < 0.05$

Hence we reject H_0 .

Conclusion:

We conclude that, Loan is independent on marital status of road site workers.

❖ Proportion Test:

Here we claim that, 85% Road Site Workers are save money for future.

We have to test,

$H_0: P=0.85$ (i.e.85% road site workers are save money for future)

$H_1: P \neq 0.85$

Saving money for future	Save	264
	Do not save	45
	Total	309

Procedure:

By using R software we get,

1-sample proportions test with continuity correction

data: 264 out of 309, null probability 0.85

X-squared = 0.018339, df = 1, p-value = 0.8923

alternative hypothesis: true p is not equal to 0.85

95 percent confidence interval:

0.8089145 0.8907669

sample estimates: p = 0.8543689

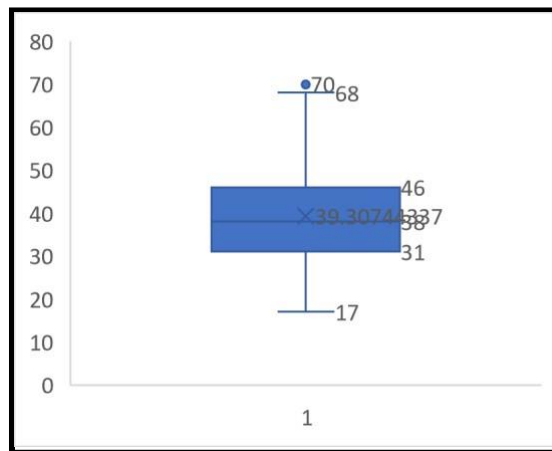
Here, p value is $0.8923 > 0.85$, hence it fails to reject H_0

Conclusion:

85% road site workers are save money for future.

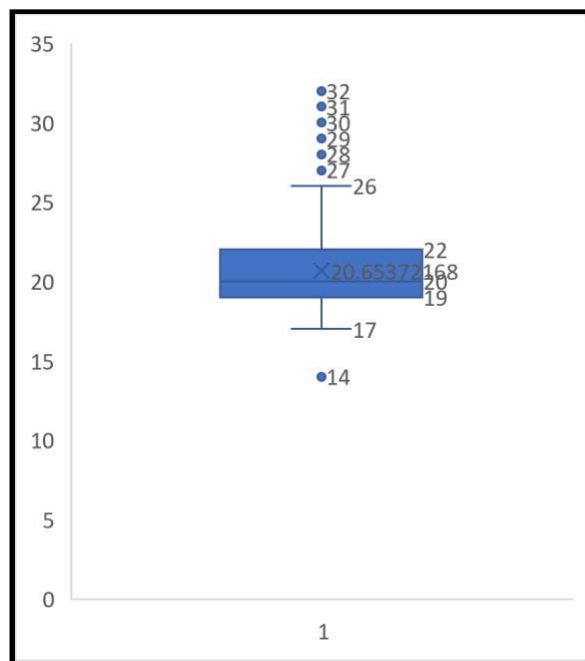
❖ Age and income

Age



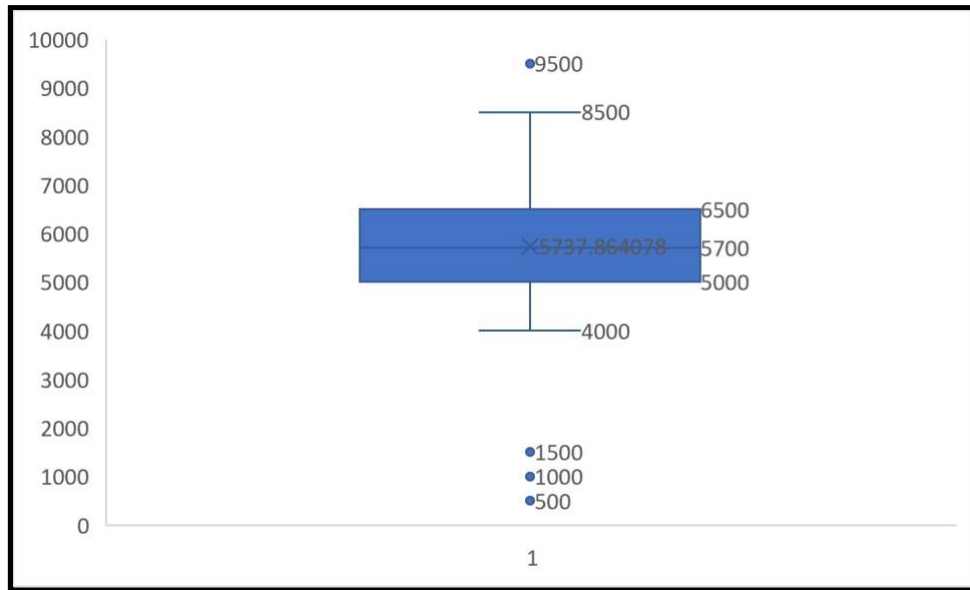
On an average, the age of Road site workers is 39 year old

Work start from age



On an average the road site workers start this work from year 20.

Income status



On an average, the income of the road site workers is 5737.

CONCLUSION

- Age of most of the Road site workers are in between 35 to 41 years.
- Around 79.93528 % of workers said that they get timely payment to their wages. While as 20.06472 % said that they don't get it on time but gets delayed by 2 to 3 days.
- The number of female workers are greater than male workers.
- The most of road site workers speak Hindi language.
- The 50.8% road site workers are of Hindi community.
- The illiterate workers are more than the educational workers.
- The greater than 50% workers are takes loan and most of worker takes loan from bachat gat.
- The 84% workers are married.
- We conclude that, House and type of house of road site workers are dependent.
- The most of the road site workers are suffered from the health problem.
- We conclude that, Loan is independent on marital status of road site workers.
- 85% road site workers are save money for future.

LIMITATIONS

- Sample size is limited.
- Income and Working hours take as rounded figure.
- Thus survey is limited to Jath city.

SCOPE

1. Study can be used to improve the life style of Road site workers.
2. The study can be used to improve the awareness about education.
3. The study can be used to provide the hospital facilities for road site workers.
4. The study can be used to improve the some important facilities related to the health, education, economical, etc.

REFERENCES

1. Fundamentals of Mathematical Statistics by S.G. Gupta & V.K. Kapoor.
2. Internet .
3. Research papers.
4. Categorical data analysis (second edition) by Alan Agresti.
5. Some report on Road site workers.

QUESTIONNAIRE

1. Age:
2. Gender: 1) Male 2) Female
3. Language: 1) Hindi 2) Marathi 3) Bihari 4) Other
4. Community: 1) Hindi 2) Marathi 3) Bihari 4) Other
5. Marital Status: 1) Married 2) Single 3) Widow 4) Divorced
6. No. of Childrens:
7. No. of persons in family: 1) Single 2) 2-5 3) 6-10 4) > 10
8. Educational Qualification: 1) Illiterate 2) Primary 3) 8th-10th 4) 11-12th
9. House: 1) Owned 2) Rented
10. Type of house: 1) Tin 2) Mud/glass 3) Slab 4) Tile 5) Other
11. From what age you have started this work?
12. How many hours do you work per day ?
13. No. of family members working in this occupation ?
1) 1 2) 2 3) 3 4) 4
14. Do you get payment on time? 1) Yes 2) No
15. Do you save money for future? 1) Yes 2) No
16. Monthly income of individual?
17. Do you face a health Problem? 1) Yes 2) No
18. Type of health problem: 1) Headache/fever 2) Eyes/ Nasal irritation
3) Asthma/ Lungs heart problems 4) Other

19. Did you get a loan?

1)Yes

2)No

20. Did you get loan from?

1)Bank

2)Savakar

3) Friend

4) Bachat gat

A
Project Report
On
**“STATISTICAL ANALYSIS EFFECT OF USE
OF
MOBILE AND INTERNET ON CHILDREN’S HEALTH”**
Submitted to

(Department Of Statistics)
Shikshan Prasark Sanstha’s
**PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA
KAVATHE MAHANKAL**

BY
Miss. Shinde Shailaja Dilip.
Miss. Jadhav Aarti Tanaji.
Miss. Mali Priyanaka Balu.
Miss. Patil Akshata Balaso.
Miss. Shinde Ashwini Pandurang.

Under the guidance of
Prof. Koshti V.V (Head of Department)
Miss. Patil S. S.
Miss. Zarekar S. M.
Mr. Satpute G. D.

Through
The Principal

Dr. Patil M. K.

Padmabhushan Vasanthaodada Patil Mahavidyalaya, Kavathe Mahankal.

FOR THE ACADEMIC YEAR,

2023-2024

Shikshan prasarak sanstha's
PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA
KAVATHE MAHANKAL.
(Department of Statistics)

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and complete the project work for the B. Sc. Part III. They are presently submitted their entitled .

This work is genuine, original, and the project or the part of it has not has Submitted for the award of any other degree .

“Statistical analysis effect of use of mobile and internet on children's health”

Miss. Shinde Shailaja Dilip.
Miss. Jadhav Aarti Tanaji.
Miss. Mali Priyanaka Balu.
Miss. Patil Akshata Balaso.
Miss. Shinde Ashwini Pandurang .



Teacher in Charge



Examiner



Head Department

DECLARATION

We hereby declare that the project entitled “**Statistical analysis effect of use of mobile and internet on children’s health**”.

It is the original work carried out by us at the department of statistics. This project or part of it has not been submitted previously for any other degree or any other purpose.

Sr. No.	Roll No.	Name	PRN No.
1.		Shinde S. D.	2021020547
2.		Jadhav A. T.	2021020533
3.		Mali P. B.	2021020015
4.		Patil A. B.	2021020022
5.		Shinde A. P.	2021020583

Place: Kavathe Mahankal

Date:

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude to Prof. Dr. Patil M. K. who gave backing and his valuable guidance for this project work. We further thank to Asso. Prof. V.V. Koshti, Asst. Prof. S. S. Patil, Asst. Prof. S. M. Zarekar and Asst. Prof. G. D. Satapute. Who gave us their encouragement, kind support and constant inspiration during our project, work further we are thankful to all of us who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our sprit up and their immerse support during every stage of project work.

INDEX

Sr. No.	Name	Page No.
1	Introduction	6
2	Objective	7
3	Data Collection	8
4	Statistical tools	9
5	Statistical analysis	11
6	Conclusion	16
7	Limitations, scope	17
8	References	18
9	Questionnaire	19

INTRODUCTION

The integration of mobile technology into children's daily lives has become increasingly prevalent in recent years, raising questions about its impact on various aspects of their development. In this statistics project, we aim to investigate the effects of mobile technology on children's behaviour, academic performance and overall well-being.

By analysing data on the usage patterns of mobile devices among children of different age groups, we seek to understand how screen time correlates with factors such as social interactions, cognitive abilities and emotional health.

The pervasive use of mobile devices among children has sparked debates about the long-term consequences of excessive screen time and digital media exposure. According to recent studies, the average child spends a significant amount of time engaged with screen ranging from watching videos and playing games to browsing social media and texting friends. This trend has raised questions about how prolonged screen time may influence various aspects of children's lives.

Through statistical analysis, we hope to uncover valuable insights that can inform parents, educators and policymakers about the potential benefits and risks associated with children's exposure to mobile technology.

OBJECTIVES

- To find the amount of screen time by children.
- To study which social media platform more used by children.
- To study how much time parents use mobile or laptop.
- To analyze the different diseases due to mobile.
- To check why children use mobile phones.
- To check whether social networking sites are popular in the girls and boys.
- To analyze use of mobile affects children's language.
- To check whether the children's screen time and parents spend time with children's are independence or not.

METHOD OF DATA COLLECTION

We surveyed Kavathe Mahankal area and collect the data. We select the Convenient sample by the random sampling method. We collect the data through the online questionnaire. 250 samples are the optimum sample size. We collect data from the online method. The data of our project is primary because we collect the data by taking online methods from each Parents .

SAMPLE SIZE DETERMINATION

The population size of Kavathe Mahankal Tehsil is 152327 according to census of 2011. We used Yamane's formula to determine the sample size at approximately 5% margin of error and we got the sample size approximately 250. The Yamane's formula for the sample size determination is given by,

$$n = \frac{N}{1+Ne^2}$$

Where,

n is the sample size

N is population size

E is the margin of error

STATISTICAL TOOLS

The following are some statistical tools leads us towards the appropriate results and finding.

A) GRAPHICAL REPRESENTATION :

To represent our data graphically, we use the multiple bar diagram, pie chart method, etc. To construct the appropriable divided bar diagram of given data we used Microsoft Excel.

1.Bar Chart: A bar chart or bar graph is a chart that presents grouped data with rectangular bars with a lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

2.Pie Chart: A pie chart displays data, information, and statistics in an easy-to-read pie-slice format with varying slice sizes telling you how much of one data element exists. The bigger the slice the more of that particular data was gathered.

3.Line Chart: A line chart or line plot or line graph or curve chart is a type of chart that displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart common in many fields. It is similar to a scatter plot except that the measurement points are ordered typically by their x-axis value) and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time a time series thus the line is often drawn chronologically. In these cases, they are known as run charts.

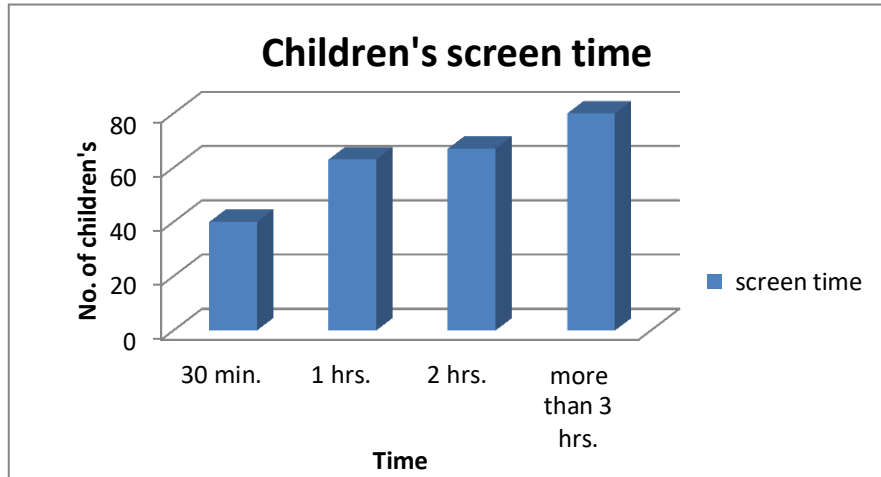
4. Pareto chart : A Pareto chart is a graph that indicates the frequency of defects, as well as their cumulative impact. Pareto charts are useful to find the defects to prioritize on order to observe the greatest overall improvement.

B) CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES:

The null hypothesis of this test is that the attributes are independent on each other. The test is used to check the independency of the attributes. Observed frequencies are tabulated in the contingency table and the expected frequencies are calculated. The test checks whether the observed frequencies and expected frequencies are significantly different or not.

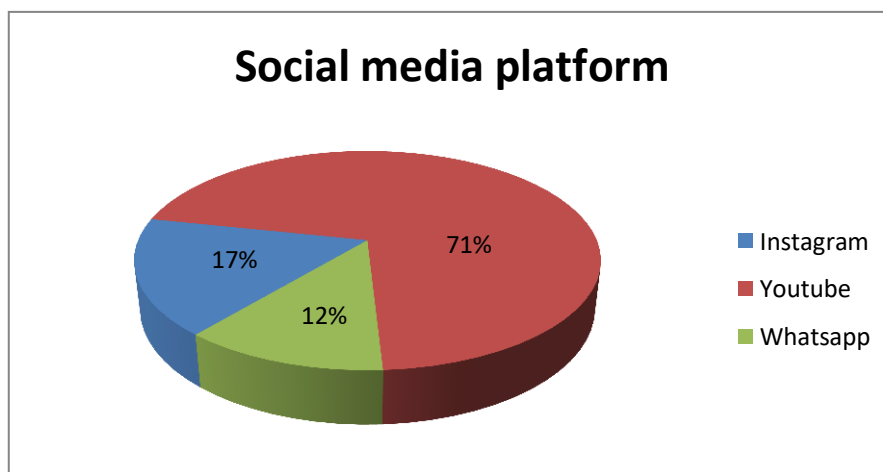
STATISTICAL ANALYSIS

1. To find the amount of time spend by children's on screen.



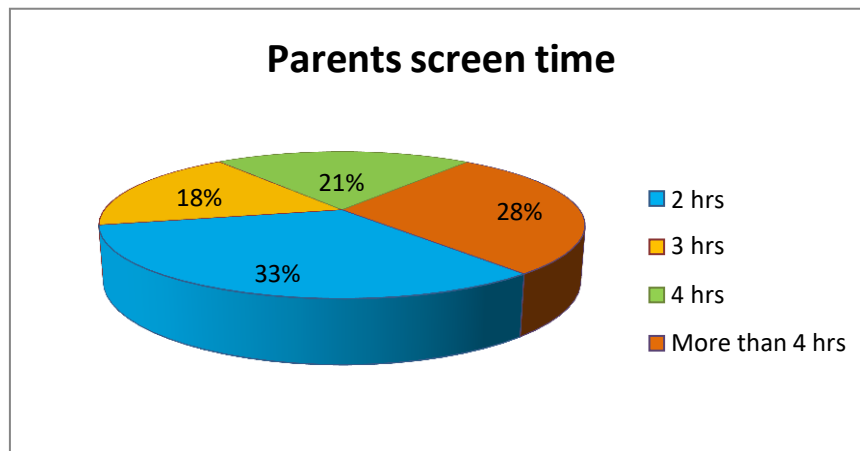
Interpretation: From the figure most of the children's use mobile more than 3 hours.

2. To study which social media platform more used by children's .



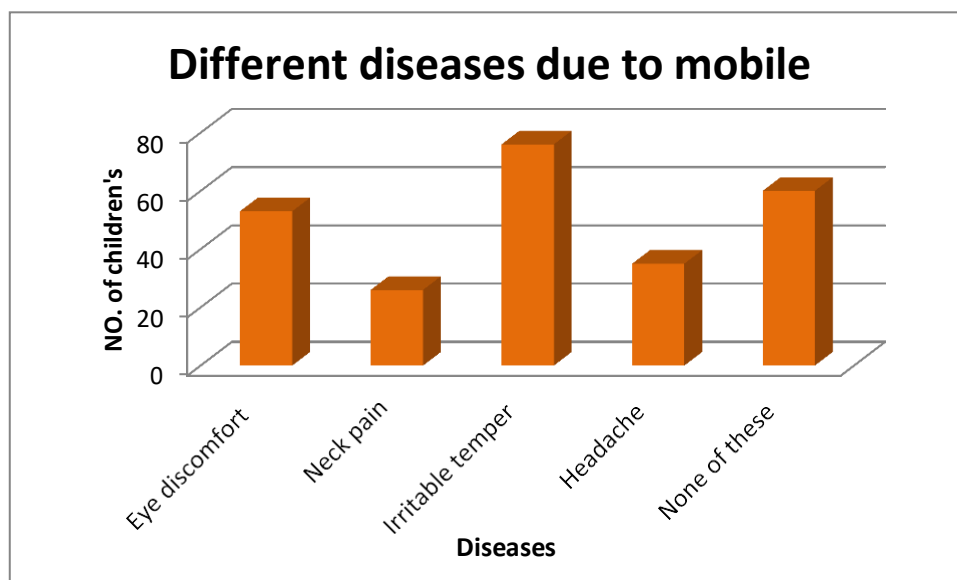
Interpretation: Most of the children's use YouTube .

3. To study how much time parents use mobile or laptop.



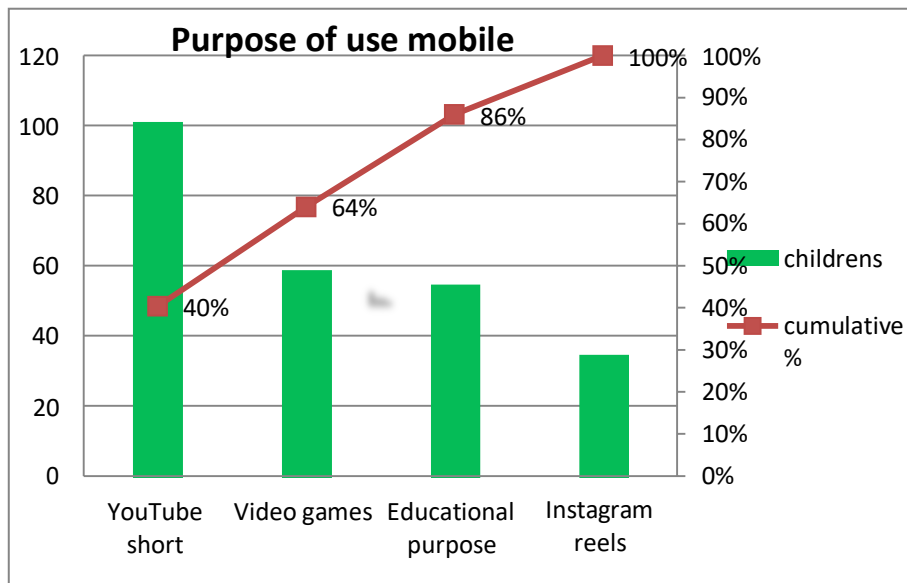
Interpretation: Most of the parents use mobile or laptop for 2 hours.

4. To check the different diseases due to mobile.



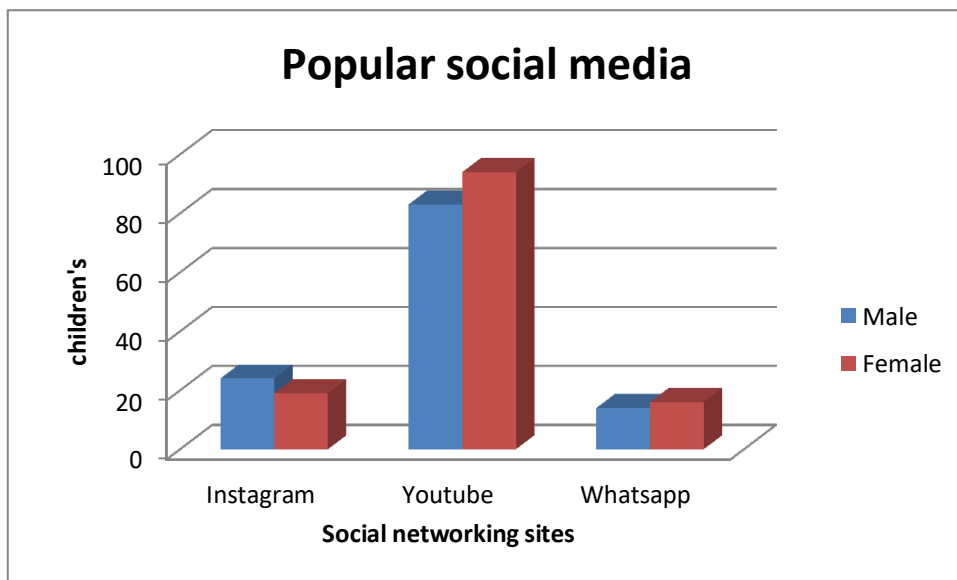
Interpretation: Out of 250 childrens 76 childrens affected by irritable temper issues.

5. To check why children's use mobile phones.



Interpretation: From the above graph we conclude that the 64% childrens use mobile for you tube shorts and video games.

6. To check whether social networking sites are popular in the male and female.



Interpretation: From the above graph, we observed that the most popular social networking site in male and female is you tube.

7. To analyse mobile improve children's language.

Test of independent of attributes:

H₀: The use of mobile and improvement of language are independent.

H₁: The use of mobile and improvement of language are dependent.

The contingency table is,

	Half hrs	1 hrs	2 hrs	More than 3 hrs	Total
Yes	16	36	38	53	143
No	24	27	29	27	107
Total	40	63	67	80	250

The test statistics is,

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2_{(cal)} = 7.5165 \quad \chi^2_{(tab)} = 7.8147279$$

$$\chi^2_{(cal)} = 7.5165 < \chi^2_{(tab)} = 7.8147279$$

Interpretation: We conclude that, the data provides enough evidence to claim that improvement of language is independent upon the use of mobile.

8. To find out the relation between time spend by children's on screen and parents time spend with children's.

Test of independent of attributes:

H₀: The children screen time and parents time spend with children are independent.

H₁: The children screen time and parents time spend with children are dependent.

The contingency table is,

	Half hrs	1 hrs	2 hrs	More than 3 hrs	Total
2 hrs	20	16	12	34	82
3 hrs	3	16	14	16	49
4 hrs	2	7	16	14	39
More than 4 hrs	15	24	25	16	80
Total	40	63	67	80	250

The test statistics is,

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$\chi^2_{(cal)} = 27.74145 \quad \chi^2_{(tab)} = 16.91898$$

$$\chi^2_{(cal)} = 27.74145 > \chi^2_{(tab)} = 16.91898$$

Interpretation: We conclude that, the data provides enough evidence to claim that children's screen time is depend on parents time spend with children.

OVERALL CONCLUSIONS

2. Most of the children's use mobile more than 3 hours.
3. Most of the children's use you tube.
4. Most of the parents use mobile or laptop 2 hours.
5. Out of 250 children's 76 children's affected by irritable temper issues.
6. Most preferred purpose use of mobile is, you tube shorts.
7. Most popular social networking site in male and female is you tube.
8. Improvement of language is independent upon the use of mobile.
9. Children's screen time is depend on parents time spend with children.

LIMITATIONS

- 1) Sample size (250) is very small to represent the whole area.
- 2) The analysis and finding can be made more accurate if the data on large scale is available.

SCOPE

- 1) This project can be extended to the state and national level.

REFERENCE

- fundamental of mathematical statistics by Gupta and Kapoor.
- <https://innohealthmagazine.com/2019/well-being/effects-mobile-phones/>
- https://www.researchgate.net/publication/316974592_Effects_of_Mobile_Phones_on_Children's_and_Adolescents'_Health_A_Commentary



QUESTIONNAIRE

1) Children's age?

- 0-2
- 2-5
- 5-7
- 7-15

2) Gender?

- Male
- Female

3) What is the education of parents?

- Primary
- Secondary
- Graduation
- Post-Graduation

4) How much time parents use mobile/laptop?

- 2 hours
- 3 hours
- 4 hours
- More than 4hours

5) How much time are you out of the house?

- 2 hours
- 3 hour
- 4 hours
- More than 4 hours

6) How much time parents spend with children's?

- 2 hours
- 3 hours
- 4 hours
- More than 4 hours

7) Children's daily screen time?

- 2 hours
- 3hours
- 4 hours
- More than 4 hours

8) Most used social media platform by children's?

- You Tube
- Instagram
- What's App

9) Reason for mobile used?

- Video games
- You Tube short
- Educational purposes
- Instagram reels

10) Do they have their own mobile?

- Yes
- No

11) Do they have their own account own social media ?

- Yes
- No

12) Does mobile make a difference in their language?

- Yes
- No

13) Different diseases due to mobile?

- Eye discomfort
- Neck pain
- Headache
- Irritable temper
- None of these

14) Do children's remember mobile phone while sleeping at night?

- Yes
- No

15) What parents can do to reduce children's screen time?

- Free response

A

Project Report

On

“STATISTICAL ANALYSIS ON ACADEMIC
PERFORMANCE OF THE UNDERGRADUATE STUDENTS”

Submitted To

ShikshanPrasarakSanstha’s

**PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA,
KAVATHE MAHANKAL**

(Department Of Statistics)

By

Mr. Kute Pranav Kundlik

Mr. Kore Pratik Jayant

Mr. Kadam Aniket Mahadev

Mr. Bandgar Vishal Annaso

Under the Guidance of

Mr. Koshti V.V.(Head of dept.)

Mr. G. D. Satpute

Miss. Patil S.S.

Miss. Zarekar S.M.

Through

Hon. Prin. Prof (Dr) M.K. Patil

**PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA,
KAVATHE MAHANKAL.**

For the academic year 2023-2024

ShikshanPrasarakSanstha's

**PADMABHUSHAN VASANTRAODADA PATIL
MAHAVIDYALAYA, KAVATHE MAHANKAL**

(Department Of Statistics)

CERTIFICATE

This is to certify that the following students are working under my Guidance. They have satisfactorily carried out and completed the project Work for the B.Sc Part III. They are presently submitted their project entitled.

This work is genuine, and the project or the part of it has not has submitted for the award of any other degree.

“Statistical Analysis on Academic Performance Of The Undergraduate Student”

Mr. Kute Pranav Kundlik

Mr. Kore Pratik Jayant

Mr. Kadam Aniket Mahadev

Mr. Bandgar Vishal Annaso



Teacher in Charge



Examiner



Head Department of statistic

INDEX

Sr.No.	Name Of Topic	Page No.
1	Introduction.	04
2	Method of data collection.	05
3	Objectives.	06
4	Statistical tools.	07
5	Analysis of data.	08
6	Overall conclusion.	20
7	Limitations and scope.	21
8	Bibliography.	22
9	Questionnaire.	23

INTRODUCTION

The final year of college is a pivotal juncture, where students solidify academic foundations and prepare for their future endeavors. This study explores various aspects of undergraduate education, focusing on B.A, B.Com, and B.Sc streams. Our objectives include determining the most preferred subject in the final year (TY) across these streams, examining the dependency between gender and pursuing higher studies, comparing attendance percentages, identifying reasons for irregular attendance, and exploring competitive exam preparation methods.

We aim to understand students' academic preferences, gender-based differences in educational pursuits, and attendance patterns. By addressing these objectives, we seek to provide insights that can enhance educational policies and practices, ultimately improving the academic experience and outcomes of undergraduate students.

SAMPLE SIZE DETERMINATION

The population size of PVP college is almost 350. We use Yamane's formula to determine the sample size at approximately 6% margin of error. The Yamane's formula for the sample size determination is given by,

$$n = \frac{N}{1+Ne^2}$$

Where, n is sample size
N is the population size
E is the margin of error.

DATA COLLECTION

The main objective of the project is "To analysis academic performance of the undergraduate students." We have collected sample of TY students of BA, BCom, BSc stream from PVP college. We prepare a questionnaire to fulfill the objectives behind the project accurately. The data collection procedure was done using google forms and direct interviewing with students.

OBJECTIVE

1. To determine the most preferred subject in TY across B.A, B.Com and B.Sc streams.
2. To check dependency between the gender and pursuing higher study across different streams.
3. To compare the attendance percentage across different streams.
4. To identify the reasons why students are irregular.
5. To identify which competitive exam are most of the student preparing across different streams and also to compare the approaches to competitive exam preparation employed by male and female students
6. To check dependency between the gender and attendance percentage across different streams.
7. To identify the preparation method to prepare for competitive exams

STATISTICAL TOOLS

1) Graphical representation:

❖ Pie chart

A pie chart is a type of graph that represents the data in the circular graph. The slices of pie show the relative size the data, and it is type of pictorial representation of data. A pie chart requires list of categorical variables and numerical variable. Here the term pie represents the whole. All the slices of the pie chart add up to make whole equalling 100% and 360 degree.

❖ Bar chart

A bar chart is a graphical representation used to display and compare discrete categories of data through rectangular bars, where the length or height of each bar is proportional to the frequency or value of the corresponding category or strips. It can be either horizontal or vertical.

❖ Donut chart

A donut chart is a graphical representation used to show proportions of categories that make up the whole, but the center can also be used to display data.

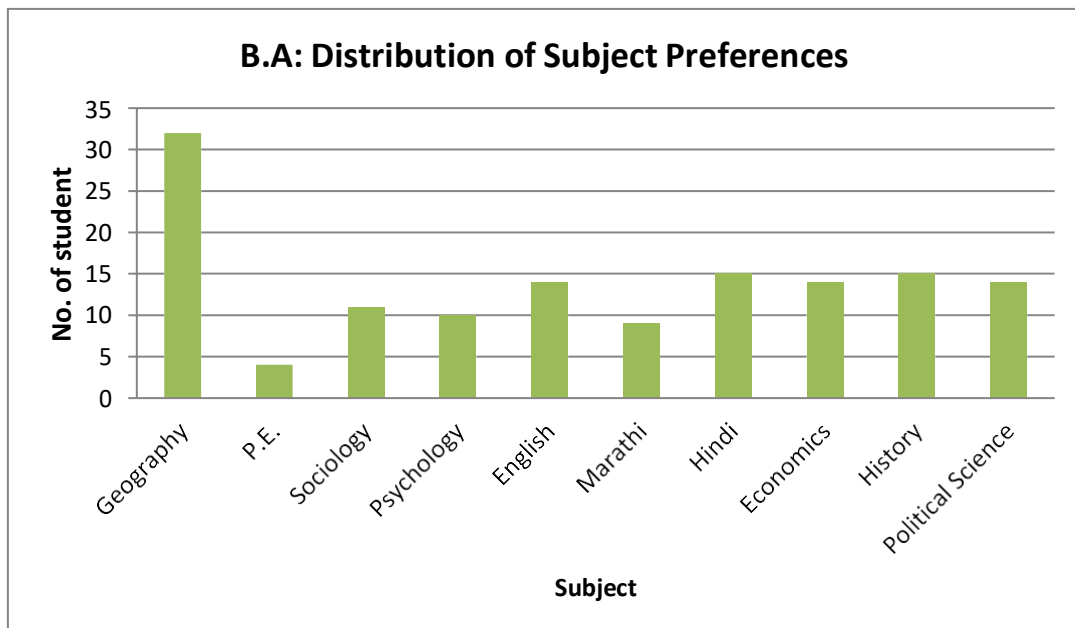
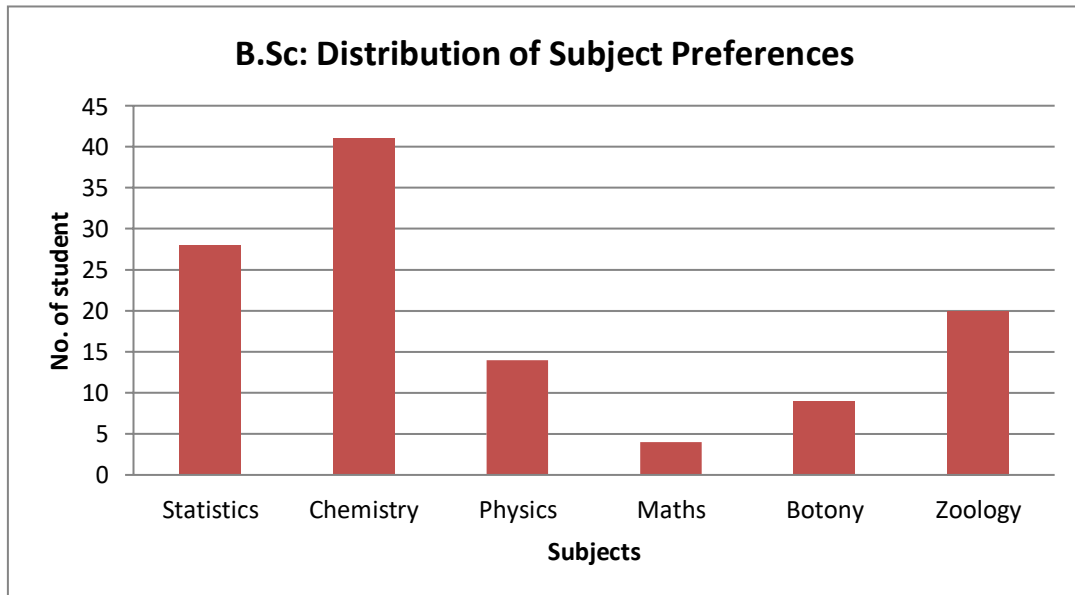
2) Hypothesis testing:

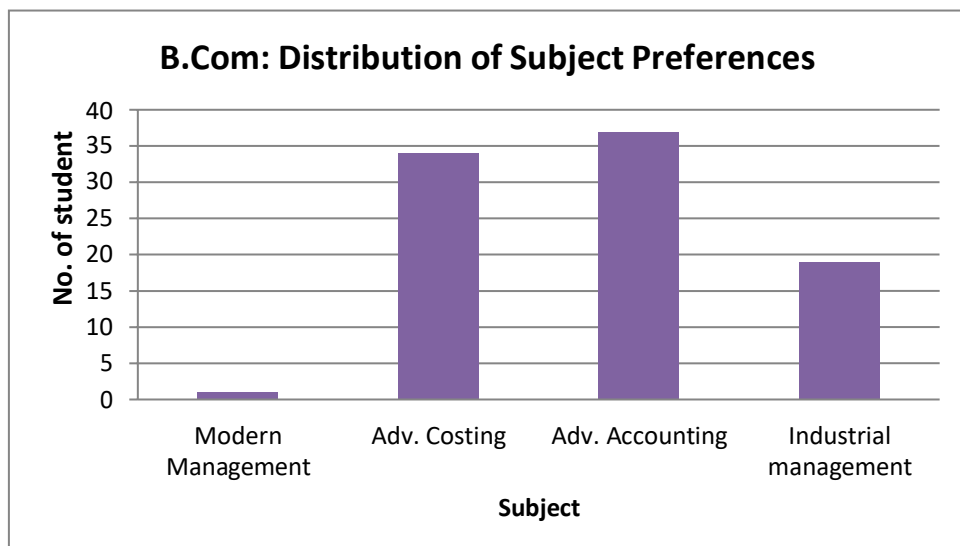
❖ Chi-square test for independence of attributes:

The null hypothesis of this test is that the attributes are independent on each other. The test is used to check the independency of the attributes. Observed frequencies are tabulated in the contingency table and the expected frequencies are calculated. The test checks whether the observed frequencies and expected frequencies are significantly different or not.

STATISTICAL ANALYSIS

1) To determine the most preferred subject in TY across B.A, B.Com and B.Sc streams.





Interpretation : Most preferred stream in TY across B.A, B.Com and B.Sc streams are ‘Geography’, ‘Advanced Accounting’ and ‘Chemistry’ respectively.

2) To check dependency between the gender and pursuing higher study across different streams

H_0 : Pursuing higher study is independent on gender

H_1 : Pursuing higher study is dependent on gender

B.Sc:

	Yes	No	Total
Male	23	24	47
Female	20	46	66
Total	43	70	113

Tabulated χ^2 is,

$$\chi^2 = 3.8414$$

$$\begin{aligned}\text{Calculated } \chi^2 &= \frac{N(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)} \\ &= \frac{113(23*46-24*20)^2}{(23+24)*(20+46)*(24+46)*(23+20)} \\ &= \mathbf{4.0432}\end{aligned}$$

Therefore we reject the Null hypothesis at 5% level of significance. We conclude that, the data provides enough evidence to claim that Pursuing higher study is dependent on gender.

B.Com:

	Yes	No	Total
Male	30	25	55
Female	10	27	37
Total	40	52	92

Tabulated χ^2 is,

$$\chi^2 = 3.8414$$

$$\begin{aligned}\text{Calculated } \chi^2 &= \frac{N(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)} \\ &= \frac{92(30*27-25*10)^2}{(30+25)*(10+27)*(25+27)*(30+10)} \\ &= \mathbf{6.8161}\end{aligned}$$

Therefore we reject the Null hypothesis at 5% level of significance.

We conclude that, the data provides enough evidence to claim that pursuing higher study is dependent on the gender.

B.A:

	Yes	No	Total
Male	16	57	73
Female	9	54	63
Total	25	111	136

Tabulated χ^2 is,

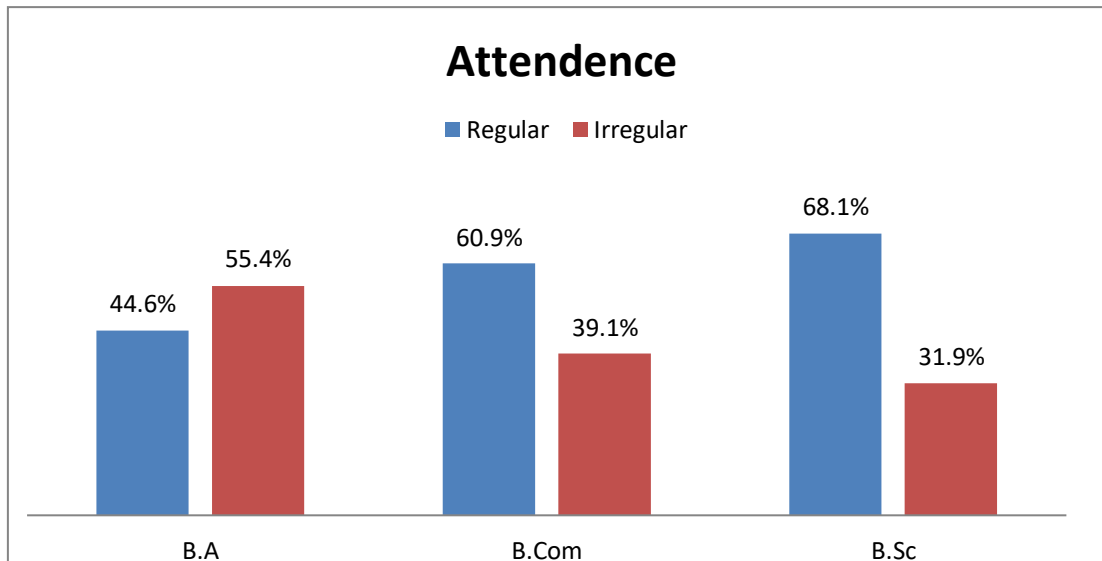
$$\chi^2 = 3.8414$$

$$\begin{aligned}\text{Calculated } \chi^2 &= \frac{N(ad-bc)^2}{(a+b)(c+d)(b+d)(a+c)} \\ &= \frac{136(16*54-57*9)^2}{(16+57)*(9+54)*(54+57)*(16+9)} \\ &= \mathbf{1.3128}\end{aligned}$$

Therefore we accept the Null hypothesis at 5% level of significance.

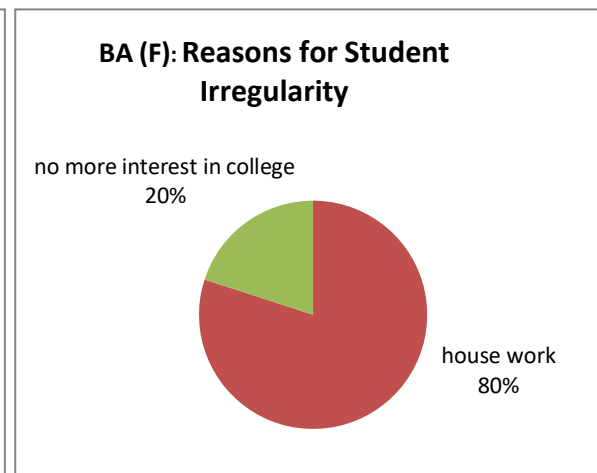
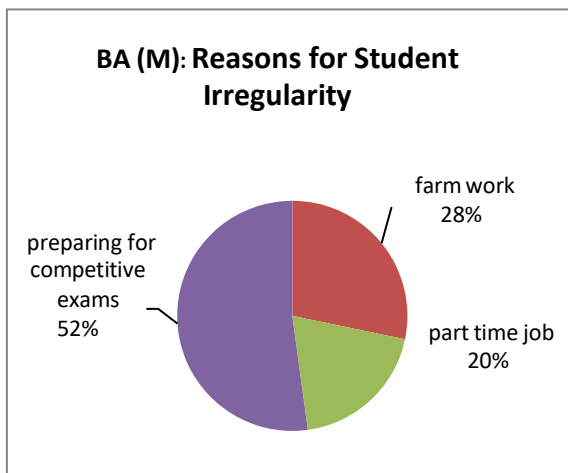
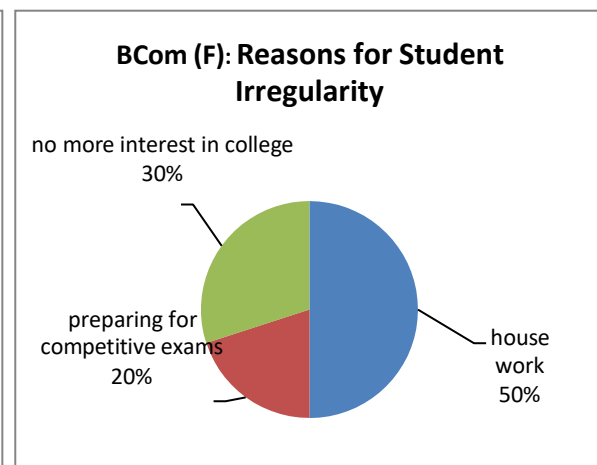
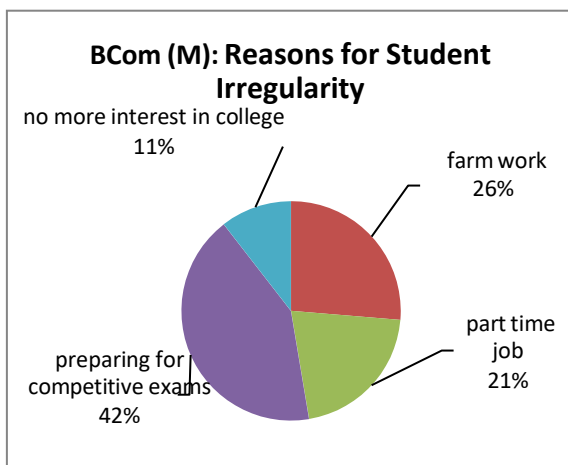
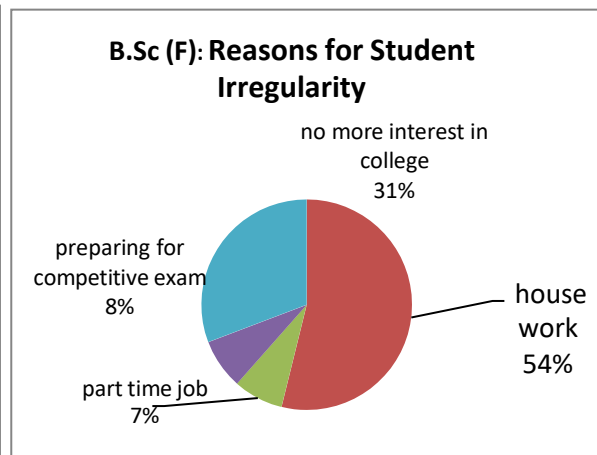
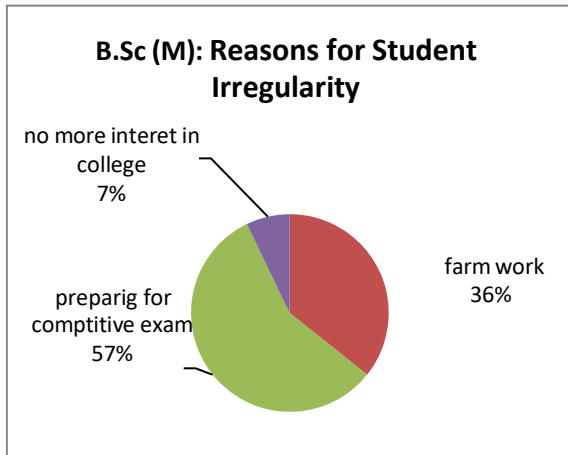
We conclude that, the data provides enough evidence to claim that opinion is independent on the gender.

3) To compare the attendance percentage across different streams.



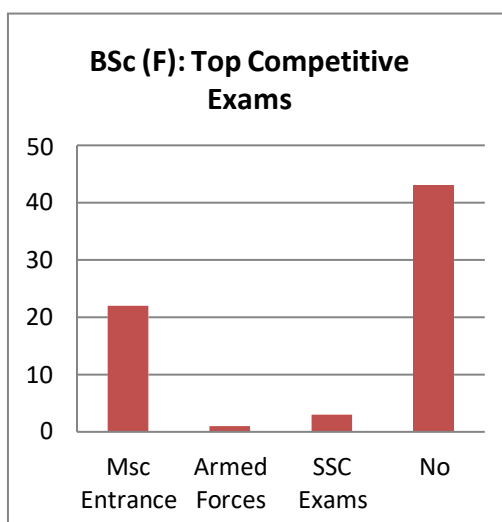
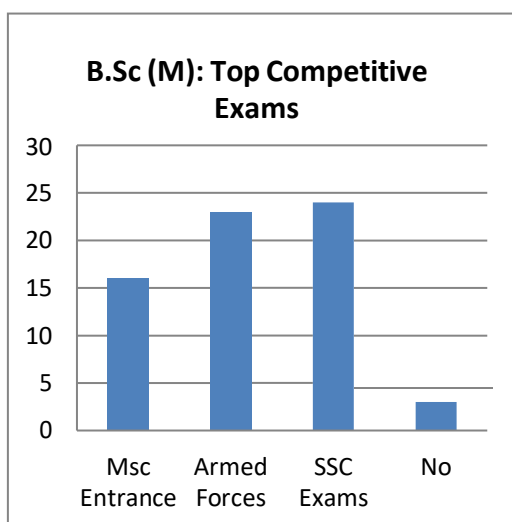
Interpretation : From graph we can observe that B.Sc students are more regular than B.A and B.Com students and B.A students are more irregular than B.Com and B.SC students

4) To identify the reasons why students are irregular.

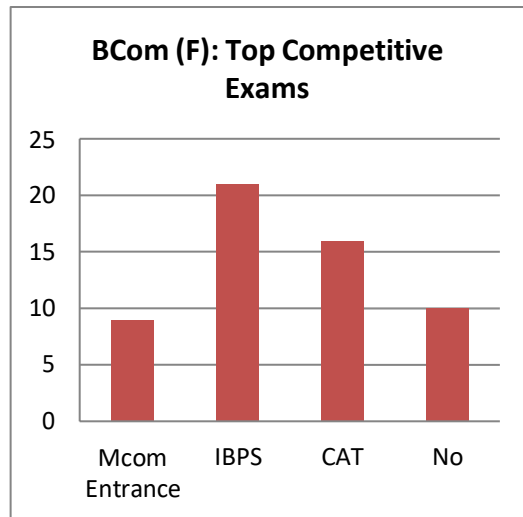
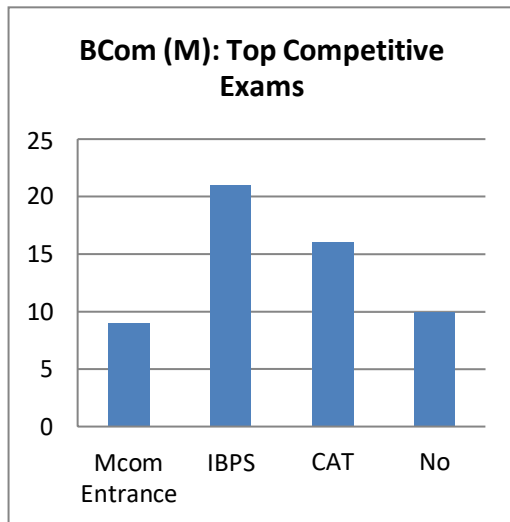


Interpretation : From the graph we can see that most of the male students are irregular because of doing preparation of competitive exam & farm work responsibility and most of the female students are irregular due to house work and no more interested in college.

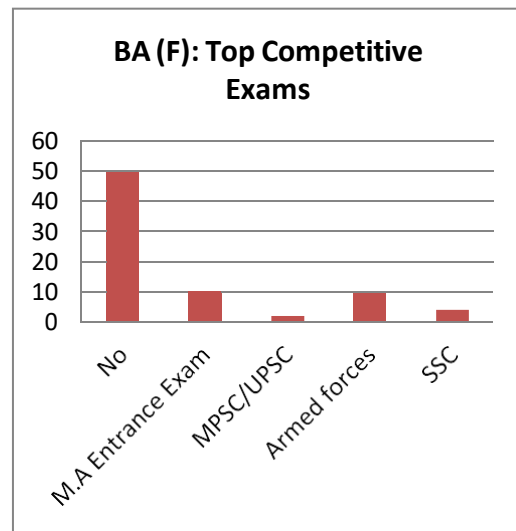
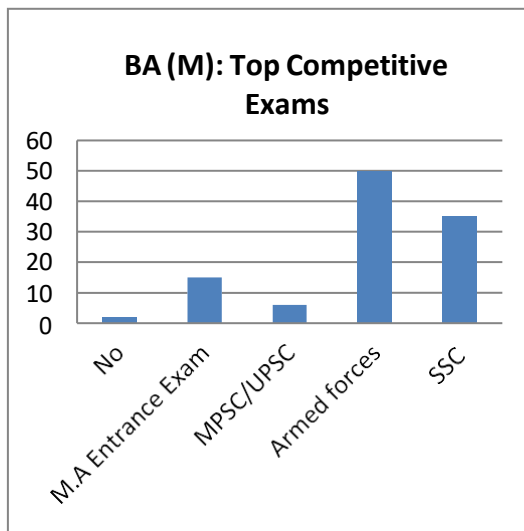
5) To identify which competitive exam are most of the student preparing across different streams and also to compare the approaches to competitive exam preparation employed by male and female students.



1) From the graph we can conclude that most of the male B.Sc students are preparing for Armed forces and M.Sc entrance and SSC exams



2) From the graph we can conclude that most of the male B.Com students are preparing IBPS and CAT exams.



3) From the graph we can conclude that most of the male B.A students are preparing Armed Forces and SSC exams and female students are not interested in competitive exams.

4) Also we can see that across all streams female students are less interested than male students in competitive exams.

6) To check dependency between the gender and attendance percentage across different streams.

H_0 : The relationship between gender and attendance percentage is independent

H_1 : The relationship between gender and attendance percentage is dependent

B.Sc :

Sr. No.	(O_i)	(E_i)	$\frac{(O_i - E_i)^2}{E_i}$
1	25	31.38318	1.29830562
2	21	14.61682	2.78753854
3	48	41.61682	0.97905014
4	13	19.38318	2.10207825
Total			7.16697255

The test statistic is,

$$\chi^2 = 3.8414$$

$$\begin{aligned} \text{Calculated } \chi^2 &= \sum \frac{(O_i - E_i)^2}{E_i} \\ &= \mathbf{7.1669} \end{aligned}$$

$$\text{Cal } \chi^2 > \text{Tab } \chi^2$$

Therefore we reject the Null hypothesis at 5% level of significance. We conclude that, the data provides that for B.Sc streams gender is dependent on the attendance percentage.

B.Com :

Sr. No.	(<i>O_i</i>)	(<i>E_i</i>)	$\frac{(O_i - E_i)^2}{E_i}$
1	37	38.34783	0.04737257
2	19	17.65217	0.10291283
3	26	24.65217	0.07369067
4	10	11.34783	0.16008662
Total			0.38406269

The test statistic is,

$$\chi^2 = 3.8414$$

$$\text{Calculated } \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$
$$= \mathbf{0.3840}$$

$$\text{Cal } \chi^2 < \text{Tab } \chi^2$$

Therefore we accept the Null hypothesis at 5% level of significance.

We conclude that, the data provides that for B.Com stream gender is independent on the attendance percentage.

B.A :

Sr. No.	(<i>O_i</i>)	(<i>E_i</i>)	$\frac{(O_i - E_i)^2}{E_i}$
1	27	32.7971	1.02467547
2	46	40.2029	0.83591946
3	35	29.2029	1.15078937
4	30	35.7971	0.93880185
Total			3.95018615

The test statistic is,

$$\chi^2 = 3.8414$$

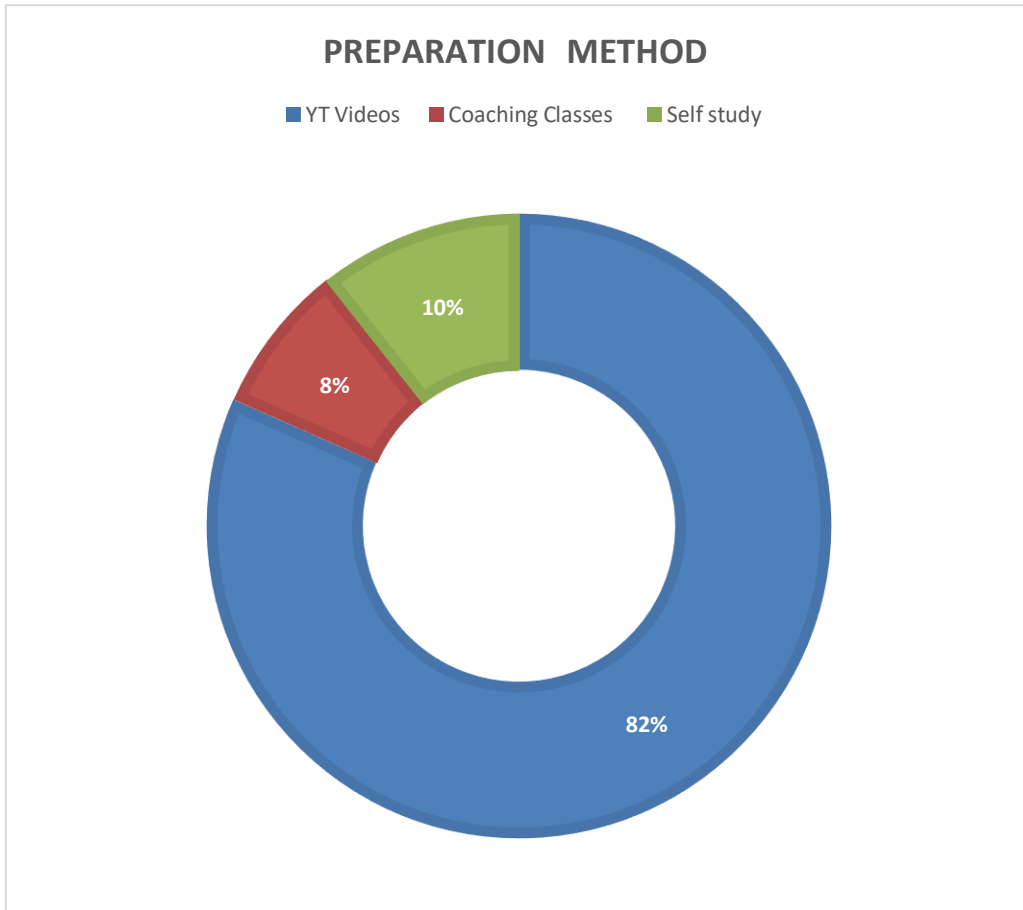
$$\text{Calculated } \chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

$$= 3.9501$$

$$\text{Cal } \chi^2 > \text{Tab } \chi^2$$

Therefore we reject the Null hypothesis at 5% level of significance. We conclude that, the data provides that for B.A stream gender is dependent on the attendance percentage.

7) To identify the preparation method to prepare for competitive exams



Interpretation : From graph we can observe that most of the students prefers Youtube videos to prepare for competitive exams

OVERALL CONCLUSION

- Most preferred stream in TY across B.A, B.Com and B.Sc streams are ‘Geography’, ‘Advanced Accounting’ and ‘Chemistry’ respectively.
- For BSc and B.Com student pursuing higher study is dependent on gender and for B.A student it is independent on gender.
- BSc students are more regular to college and B.A students are more irregular to college.
- Most of the male students are irregular because of doing preparation of competitive exam & farm work responsibility and most of the female students are irregular due to house work and no more interested in college.
- Most of the students are preparing for
 - I) BSc: Armed forces and MSc entrance and SSC exams
 - II) B.Com: IBPS and CAT exams
 - III) B.A: Armed Forces and SSC exams
- Female students are less interested than male students in competitive exams.
- For BSc and B.A streams gender is dependent on the attendance percentage and for B.A stream gender is dependent on the attendance percentage.
- Students prefers ‘Youtube videos’ to prepare for competitive exams

LIMITATION

- This study is limited to 347 respondent. So finding and suggestion given on the basis of the study cannot be assumed to entire population.
- This study also restricts itself to the geographical area of the Kavathe Mahankal tehsil.

SCOPE

- The study can provide insights and information to make informed decisions about their academic journey.
- The study can raise the awareness of potential biases and advocating for equitable educational experiences.
- The study highlight the need for motivating female students to participate in competitive exams.

BIBLIOGRAPHY

- Fundamental of statistics by Gupta and Kapoor.
- 100 statistical test by Gopal K. Kanji

QUETIONARRIE

<https://docs.google.com/forms/d/e/1FAIpQLSdxR6R7M2O357cUKkVF4v2oGrgXmpA6OYZdY7OISlflj1CxQ/formResponse?pli=1>

THANK

YOU...

B. Sc. III Students Projects (2022-23)

Sr. No.	Name of the Students	Title of the Project	Name of Mentor
1	Athavale Akanksha Suresh	Statistical Perception Towards The Online Shopping	Mr. V. V. Koshti
	Mulla Anjum Maula		
	Bhosale Dhanashri Sanjay		
2	Kakade Manish Mahavir	Statistical Analysis Of Milk Production	Mr. G. D. Satpute
	Pawar Vikas Pradip		
	Sathe Tejas Tatyasaheb		
	Babar Suraj Dinkar		
	Borakar Someshwar Ramesh		
3	Purohit Tanaya Dhananjay	Statistical analysis of demand of Bakery Products	Miss. P. M. Patil
	Mali Shital Ananda		
	Londhe Dipali Uddhav		
	Jagatap Pranali Keshev		
4	Desai Vikrat Anandrao	The statistical analysis of Bike User's	Mr. M. K. Patil
	Patil Sairaj Nandkumar		
	Hippapkar Dayanand R.		
	Ghubade Aniket Mangesh		
	Nadaf Shakil Sardar		
5.	Deshmukh Rutuja Sunil	Customer Perception towards the Himalaya Products	Miss. S. S. Patil
	Kadam Pooja Balkrishna		
	Patil Prajwal Shivaji		

A
PROJECT REPORT
ON
“THE STATISTICAL ANALYSIS OF BIKE USER’S “

Submitted to,
Shikshn Prasarak Sanstha’s
Padambhushan Vasantrodada Patil Mahavidyalaya
Kavathe Mahankal ,Sangli
Affiliated to Shivaji University , Kolhapur

DEPARTMENT OF STATISTICS

By,

Mr. Vikrant Anandrao Desai.

Mr . Sairaj Nandkumar Patil.

Mr . Shakil Sardar Nadaf.

Mr. Aniket Mangesh Ghubde.

Mr. Dayanand Ramchandra Hipparkar.

Under the guidance of,

Prof. Dr. Patil M.K.

Prof. Koshti V.V.

Assi. Prof. Satpute G.D.

Assi . Prof. Patil S.S.

Assi . Prof. Patil P.M.

Through,

Hon. Principle, Prof. Dr. Patil M.K. ,

Padambhudhan Vasantrodada Patil Mahavidyalaya,
Kavathe Mahankal

For the academic year 2022-23
Padambhushan Vasantrodada Patil Mahavidyalaya ,Kavathe Mahankal

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for **B.Sc. –III**. They have presently submitted their project entitled “**THE STATISTICAL ANALYSIS OF BIKE USER'S**”.

This work is genuine, original and the project or the part of it has not been submitted for award of any other degree.

Sr.No.	Exam No.	Name	PRN
1	27570	Vikrant Anandrao Desai.	2020005582
2	27582	Sairaj Nandkumar Patil.	2020005692
3	27580	Shakil Sardar Nadaf.	2020006457
4	27573	Dayanand Ramchandra Hipparkar.	2020031505
5	27572	Aniket Mangesh Ghubde.	2020045620



Teacher in charge



Examiner



Dr. M. K. Patil

Head of Department

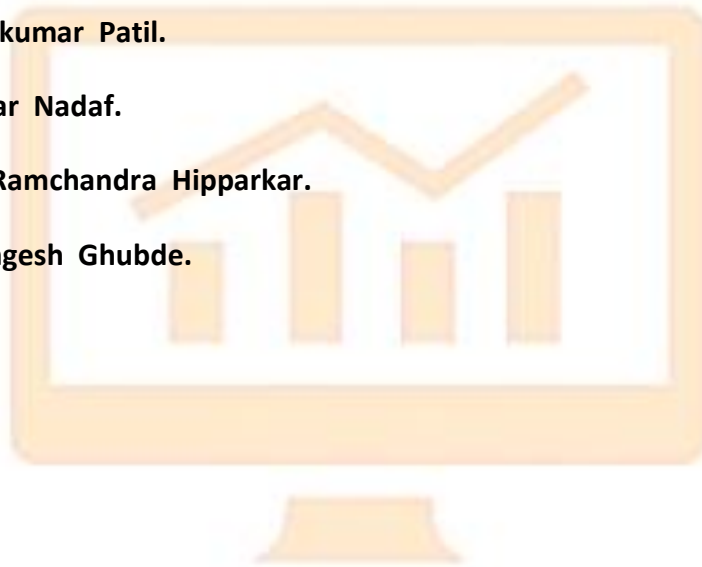
DECLARATION

We hereby declare that the project entitled “**THE STATISTICAL ANALYSIS OF BIKE USER’S** “ with special reference to **ALL TEACHER’S OF STATISTICS DEPARTMENT**.

It is original work carried out by us at the Department of Statistics.

This project or part of it has not been summated previously for any other degree of any purpose.

- ◆ **Mr.Vikrant Anandrao Desai.**
- ◆ **Mr.Sairaj Nandkumar Patil.**
- ◆ **Mr.Shakil Sardar Nadaf.**
- ◆ **Mr.Dayanand Ramchandra Hipparkar.**
- ◆ **Mr.Aniket Mangesh Ghubde.**



Place: Kavathe Mahankal

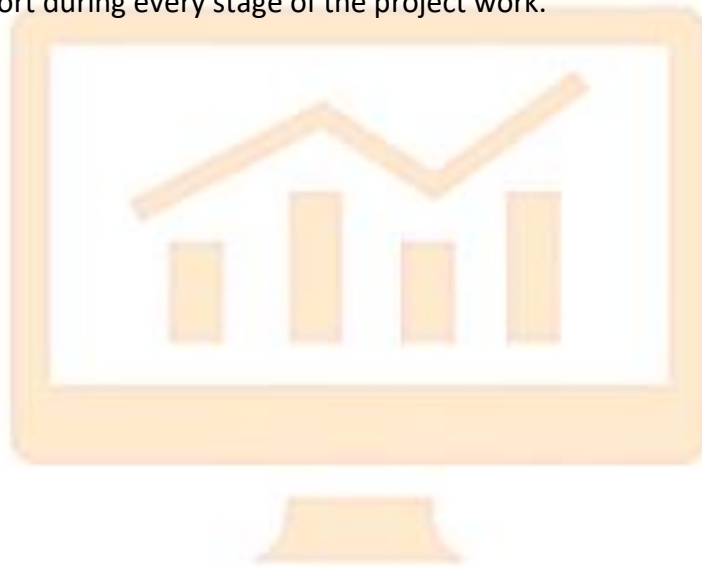
Date:

ACKNOWLEDGEMENT

A truly independent project is contradiction in terms. Every project involves contributions of many people, this project also bears imprints of many people and it is a pleasure to acknowledge all of them.

We express our hearty thanks and deepest gratitude to **Dr.Patil M.K.** who gave backing and his valuable guidance for this project work. We further thank to **Prof.Koshti V.V. , Assi.Prof.Satpute G.D. , Assi.Prof.Patil S.S. , Assi. Prof. Patil P.M.** Who gave us their encouragement ,kind support and constant inspiration during our project , work further we are thankful to **“BIKE USER’S”** who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.



INTRODUCTION

Automotive Industry, globally, as well as in India, is one of the key sectors of the economy due to its strong forward and backward linkages. In a Global Competitiveness Survey of 104 countries, India ranked only 3th in a automobile industry.

At present, people give more importance to the bike is a very important vehicle used by the students and other citizens for travelling and for various purposes, various model of bike are in the market in new features are available but in more price, and the available bike are most preferred as in a middle class person. some people are uses a most popular bike model for a publicity.

Now a days the life style of the people is different. so they are using various bike we have to study to the which type of bike are preferred, and it is useful in various purpose. And the most reliable bike and the which features are like the people to offered a bike. Now a day use bike in a larger proportion so to some awareness about the services of bike and Thus, bike has become very popular among the students. And awareness about the services and government rules

The motive of project to study not only the uses of bike but also that the person faces the problem while the no follow the government rules on bike . Also to study the risk factors from the not awareness about the document of bike like, licenses, insurance, etc .

The study of the consumer preference towards the bike purchases . The study was undertaken to know about the annual expenditure spent on bike.

CONTENTS

CONTENTS.....	PAGE NO.
1) CERTIFICATE.....	02
2) DECLARATION.....	03
3) ACKNOWLEDGEMENT.....	04
4) INTRODUCTION	05
5) CONTENTS.....	06
6) OBJECTIVES.....	07
7) SAMPLING METHODOLOGY	08
8) TOOLS & TECHNIQUES USED FOR ANALYSIS.....	10
9) DATA ANALYSIS.....	11
10) MAJOR OF FINDINGS.....	26
11) LIMITATIONS.....	27
12) SCOPE FOR FUTURE.....	27
13) REFERENCE.....	28
14) QUESTIONNAIRE.....	29

OBJECTIVES

- To study the Gender & Age group wise distribution of bike users
- To analyze expenditure pattern of various villages in Kavthe Mahankal Tahasil. (Region wise daily expenditure on fuel.)
 - To study the consumption pattern of fuel.
 - To study of ratings of bike owners for qualities of bike .
 - To study the most preferred bike in both areas .
 - To test the equality of proportion of bike in urban and rural area.
 - To study the equality of proportion of female bike user in urban and rural area.
 - To Study the dependency between occupation and use of helmet.
 - To study the dependency between factor and bike accident.
 - To study the consumption of petrol in urban and rural area are equal.
- To check monthly many spend on bike is significantly differ Of not in 5 villages.
 - To analysis opinion about new rule of R.T.O.
 - To cheak the health problems caused by prolonged use of bikes in people .
 - Accuption of costumer & to study of pattern of finance for buying a bike.
- To test if there is any correlation between gender of bike users and health problems caused due to prolonged driving.

Sampling Methodology

- **Data collection and sample size determination:**

For whole project, our area under consideration is in Sangli.

We consider the Kavthe Mahankal tahasil from each stratum using convenient sampling we select tehasil.

My area under consideration for whole project is Kavthe Mhankal Tahasil. There are 60 villages in kavthe mahankal.

I consider a particular type of villages as a stratum and select the sample size from each stratum using convenience methods.

- **Sample Size Determination:**

The sample size is determined by the **Yamane's (1967)** formula. The formula is given as,

$$n = \frac{N}{(1 + N * e^2)}$$

Where,

n = Sample size

N = Population size

e = Level of precision

Here,

The total Population of Kavthe Mahankal = **N = 2,00,000** Then our sample size becomes, and **e=0.05**

$$\begin{aligned} n &= \frac{N}{(1+N*e^2)} \\ &= \frac{200000}{(1+200000*0.05^2)} = 399.20 \approx 399 \end{aligned}$$

We got sample size **2,00,000** in all villages in kavathe mahankal tahasil. We done survey in Kavathe Mahankal tahasil , as per sampling scheme we got sample size **399**.

Village name	No. of household	Total no. Of male	Total no. Of female	population
Dhalgaon	1071	3162	2959	6014
Kokale	921	2604	2460	4975
Nagaj	996	2901	2738	5540
Kadamwadi	182	505	500	1005
Kavathe Mahankal	3733	10084	9741	19477

Total number of household in kavthe mahankal tahasil	
Total	6903
Rural	3110
Urban	3733

Sample size of rural and urban area :

Rural = 201.096

Urban = 197.904

Table : Collection of sample in different villages.

Village	No. of household	Sample size
Dhalgaon	1071	70
Kokale	921	50
Nagaj	996	50
Kadamwadi	182	31
Kavathe Mahankal	3733	198

We fill each questionnaire by asking the question to the respondent. We collect the 399 collection, enter all data in excel and analyze it.

Tools and techniques used for analysis

➤ Statistical techniques:

1. Tabulation and graphical representation:

Pie chart, bar chart, multiple bar diagram, sub-divided bar diagram, etc.

2. Proportion test:

It is used to test specific value of population proportion.

3. The Kruskal-Wallis rank sum test of K population.

It is used to test if K random samples could have come from K populations with the same mean.

Let, R_j be the rank sum of the j^{th} sample. N_j be the size of the j^{th} sample and N be the combined sample

Test statistics :

$$H = \left[\frac{12}{N(N+1)} \sum_{i=1}^n \frac{R_i^2}{N_i} \right] - 3(N+1)$$

This follows a χ^2 distribution with $(k-1)$ degree of freedom. The null hypothesis of equal means is rejected when H exceeds the critical value.

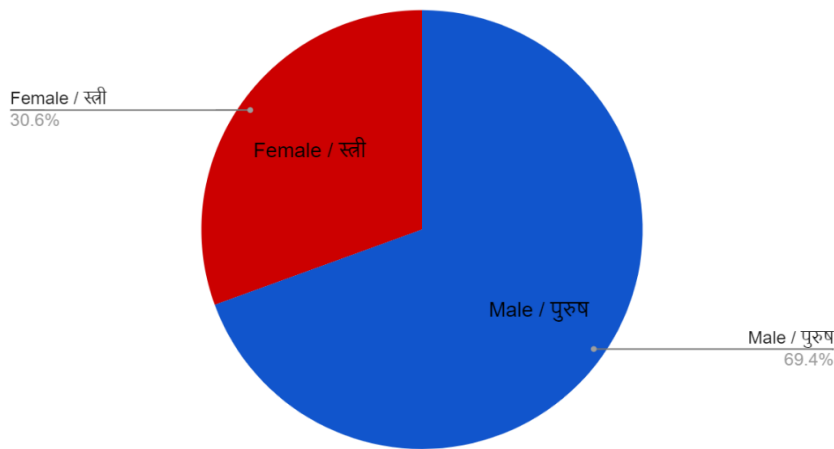
Data Analysis

Objective: To study of Gender and Age group wise representation.

Distribution of respondents according to various aspect:

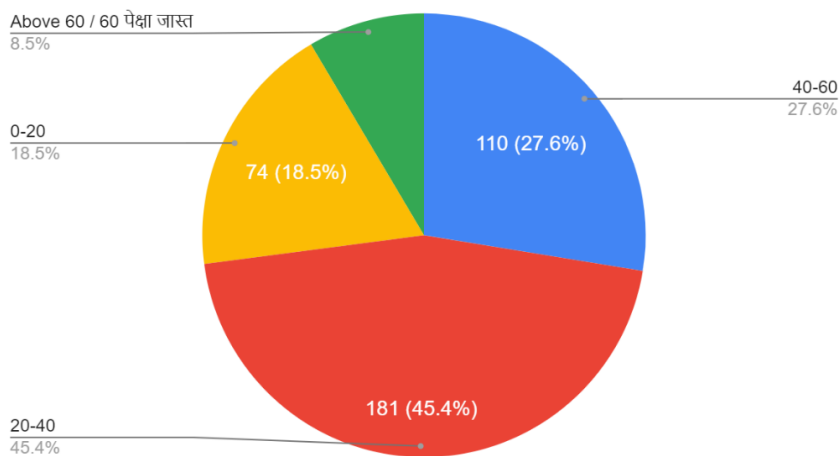
1. Gender wise distribution of male and female bike user's

Count of Gender



2. Age group wise representation

Age



Conclusion:- From figure we can say that 20-60 old people can use bike in larger scale.

❖ **Objective** : To analyze expenditure pattern of various villages in Kavthe Mahankal Tahasil Region wise daily expenditure on fuel.

Region	No. of sample	Petrol (liter/day)		Electricity (unit/day)	
		Average km	Average Price	Average km	Average Price(1km=0.15 paise)
Dhalgaon	70	49.5024	87.4560	42.5	375
Kokale	50	48.2469	85.2378	41.5	6.225
Nagaj	50	49.1548	86.8418	40.5	6.075
Kadamwadi	31	48.95	86.471	40	6
Kavathe Mahankal	198	52.1297	92.0975	44.5	6.675
Rural	201	48.9635	86.5038	41.125	6.1687
Urban	198	52.1297	92.1275	44.5	6.675

Table:- Showing the region wise daily expenditure pattern of various village Kavthe Mhankal tahasil.

Now, we focus on the only petrol users, and we find out the daily expenditure on petrol. We get average expenditure on petrol is 86.5038 in rural area & also we get average expenditure on petrol is 92.1275 in rural area . We have also find out the daily expenditure on electricity. We get average expenditure on electricity is 6.1687 in rural area & also we get average expenditure on electricity is 6.675 in uraban area.

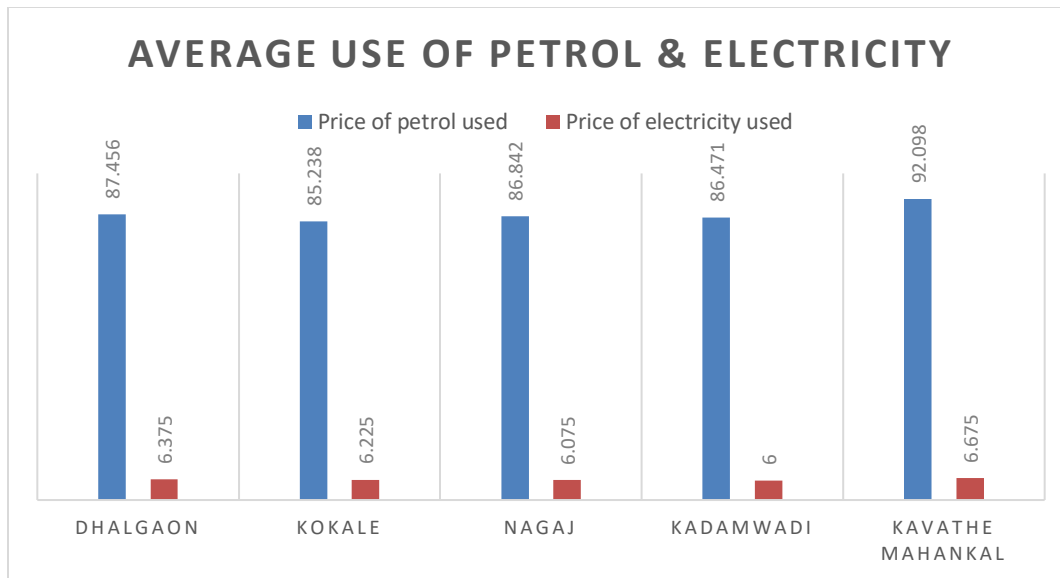


Figure:- Show the average use of Petrol and Electricity in rural and urban area.

Conclusion:

The expenditure on fuel is larger in urban area as compared to rural area. Electric vehicles are mostly not used in both areas but electric vehicles consume more electricity in urban areas than in rural areas.

❖ **Objective:** To study of consumption of pattern of fuel.

Region	No. Samples	Of	Petrol	Electricity
Dhalgaon	70		43	27
Kokale	50		38	12
Nagaj	50		37	13
Kadamwadi	31		26	5
K Mahankal	198		143	55
Urban	198		143	55
Rural	201		144	57

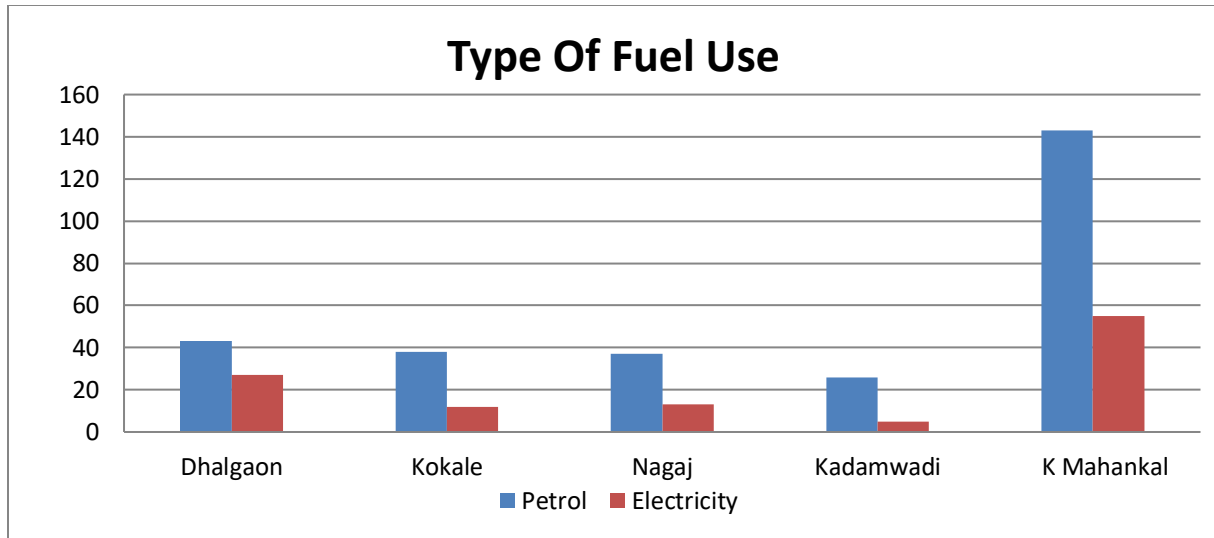
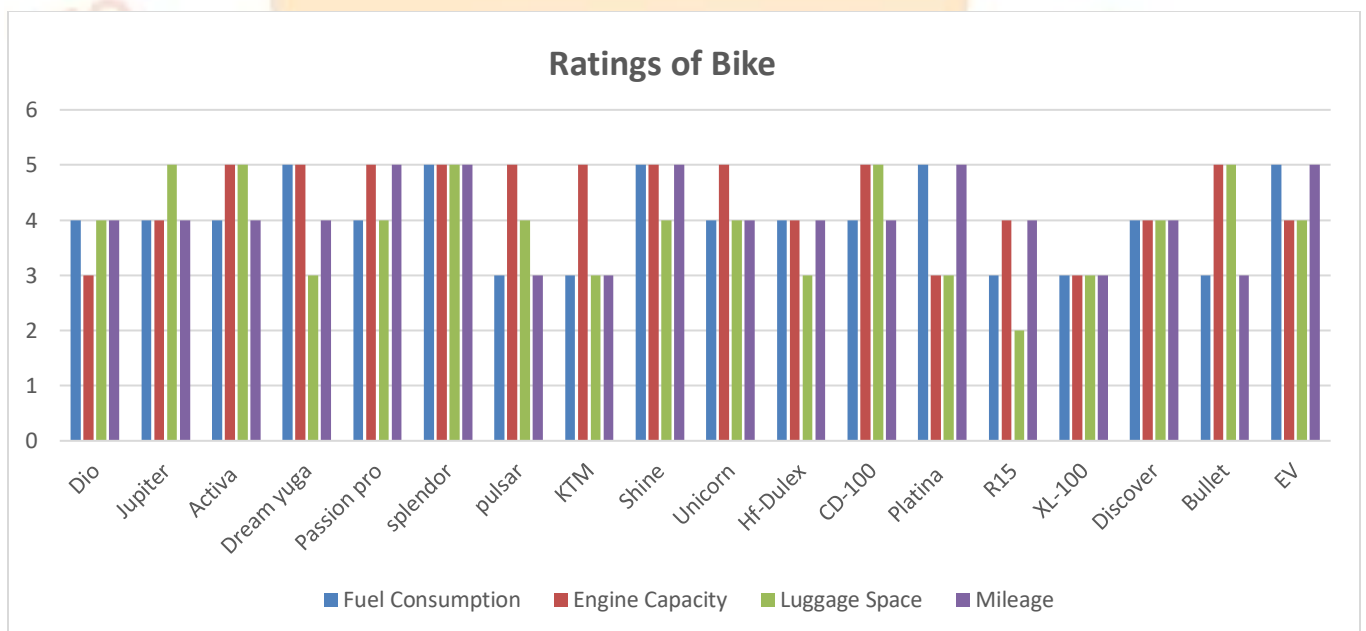


Figure 1

Conclusion: Most of the people uses petrol & very few peoples uses electricity.

❖ **Objective :** The rating of bikes.

Here we will see the ratings given bike owners for there bike qualities like Fuel consumption of bike , Luggage space in bike ,Engine capacity of bike & Mileage of bike.



Rating to the bike to 5 out of 5. The people give the more star to splendor bike. Splendor bike is getting ratings 5 in all qualities. People are also want to use Electric Vehicle in future.

Conclusion: From the above figure, we conclude that, most of the people give best rating to splendor bike.

❖ **Objective:** Most preferred bike in both areas.

Here we will check the most preferred bike in both rural & urban areas.

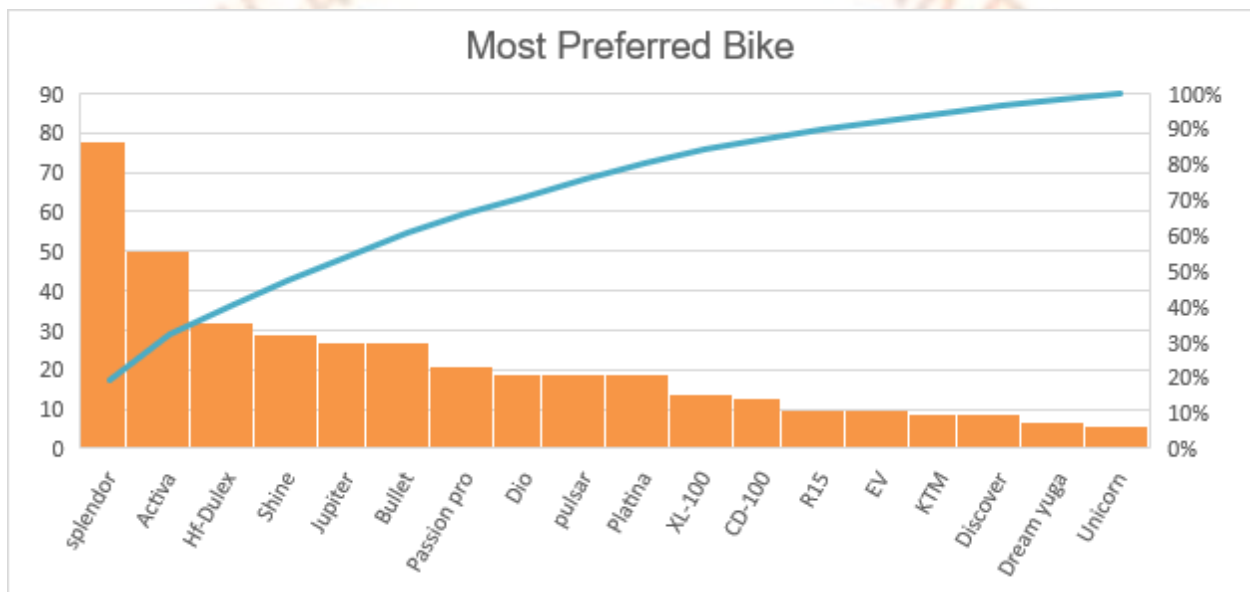


Figure shows most preferred bike in both areas.

Conclusion: From the above graph, we conclude that the most preferred bike is splendor as compared with other bikes and the second number gives to activa. This is done in both the regions the splendor can be preferred in rural as urban areas. Now electric vehicle is also coming in competition and make space in market.

❖ **Objective :** To test the equality of proportion of bike in urban area and rural area

Hypothesis:

H₀: The proportion of bike users in urban and rural area are equal.

H₁: The proportion of bike users in urban and rural area are not equal.

Urban Area N ₁	Rural area N ₂	No. of bike users In X ₁	No. of Bike users In X ₂	P ₁	P ₂
198	201	195	199	0.9949	0.9850

Calculation:

$$P = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

$$P = \frac{198 \cdot 0.9949 + 201 \cdot 0.9850}{198 + 201}$$

$$P = 0.9899$$

$$Q = 1 - P$$

$$Q = 1 - 0.9899$$

$$= 0.0101$$

Test statistics:

$$Z = \frac{P_1 - P_2}{\sqrt{PQ \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

$$Z = \frac{0.9949 - 0.9850}{\sqrt{0.9899 \cdot 0.0101 \cdot \left(\frac{1}{198} + \frac{1}{201} \right)}}$$

$$Z = 0.9888$$

Here,

cal. $Z = 0.9888$ and the table value at 5% l.o.s is $Z = 1.96$

Conclusion: We can accept the H_0 at 5% of l.o.s., thus the proportion of bike users in urban and rural area are equal. They are the proportion of bike user are differ by 0.1.

❖ **Objective:** To test the equality of proportion of female in urban and rural area.

Table gives the proportion of females:

Females in urban area	Females in rural area	No. of female bike users in urban area	No. of female bike users in rural area	P ₁	P ₂
87	35	83	34	0.9770	0.9714

Calculation:

$$P = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

$$P = \frac{87 * 0.9770 + 35 * 0.9714}{87 + 35}$$

$$= 0.9754$$

$$Q = 1 - P$$

$$Q = 1 - 0.9754$$

$$= 0.0246$$

Test statistics:

$$Z = \frac{P_1 - P_2}{\sqrt{PQ \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

$$Z = \frac{0.9770 - 0.9714}{\sqrt{0.9754 * 0.0246 * \left(\frac{1}{87} + \frac{1}{35} \right)}}$$

$$Z = 0.1806$$

Here cal. $Z = 0.1806$ and the table value at 5% l.o.s is $Z = 1.96$

Conclusion:

We can accept the H_0 at 5% of l.o.s. thus the proportion of female in rural and urban area are equal.

❖ **Objective:** To check the dependency between occupation and helmet use.

From their occupation is known their education or awareness about use of helmet.

Occupation /use helmet	Yes	no	Total
Student	124	7	131
Farmer	45	5	50
Businessman	76	3	79
Private job	75	5	80
Government job	58	1	59
Total	378	21	399

Hypothesis: **H₀:** To check occupation and use of helmet are independent.

H₁: To check occupation and use of helmet are dependent.

Calculation: Now,

Let, E_{ij} expected frequency of the observation in the observation in ith and jth column. Thus we have

$$E_{11} = \frac{131 \cdot 378}{399} = 124.1053 \quad E_{12} = \frac{131 \cdot 21}{399} = 6.8947,$$

$$E_{21} = \frac{50 \cdot 378}{399} = 47.3684 \quad E_{22} = \frac{50 \cdot 21}{399} = 2.6315$$

$$E_{31} = \frac{79 \cdot 378}{399} = 74.8421 \quad E_{32} = \frac{79 \cdot 21}{399} = 4.1579$$

$$E_{41} = \frac{80 \cdot 378}{399} = 75.7895 \quad E_{42} = \frac{80 \cdot 21}{399} = 4.2105$$

$$E_{51} = \frac{59 \cdot 378}{399} = 55.8947 \quad E_{52} = \frac{59 \cdot 21}{399} = 3.1053$$

The chi-square statistic is $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

$$\chi^2 = 4.2551$$

Table value of $\chi^2 = 9.4877$

Here the calculated value of chi-square is greater than the tabulated value of chi-square.

Conclusion:

Here the calculated value of chi-square is greater than tabular value then we reject the null hypothesis.

Therefore we say that occupation and use of helmet is dependent. hence we conclude that the occupation of peoples can use the helmet in large proportion than other.

❖ **Objective:** To check dependency between factor and bike accident.

(factor affecting on bike accident)

Hypothesis:

H₀ : To check the factor are independent on bike accident.

H₁ : To check the factor are dependent on bike accident.

Now table:

Factor/person	yes	no	total
Speed	60	30	90
Drink and drive	84	29	113
No follow rule	89	31	120
Other	44	32	76
total	277	122	399

We can get expected frequency ,

$$E_{11} = \frac{90 \times 277}{399} = 62.4812$$

$$E_{12} = \frac{90 \times 122}{399} = 27.518$$

$$E_{21} = \frac{113 \times 277}{399} = 78.4486$$

$$E_{22} = \frac{113 \times 122}{399} = 34.5514$$

$$E_{31} = \frac{120 \times 277}{399} = 83.3083$$

$$E_{32} = \frac{120 \times 122}{399} = 36.6917$$

$$E_{41} = \frac{76 \times 277}{399} = 52.7619$$

$$E_{42} = \frac{76 \times 122}{399} = 23.2381$$

The chi-square statistic is $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

$$\text{Cal } \chi^2 = 7.6375$$

And the tabular value $\chi^2 = 7.8147$

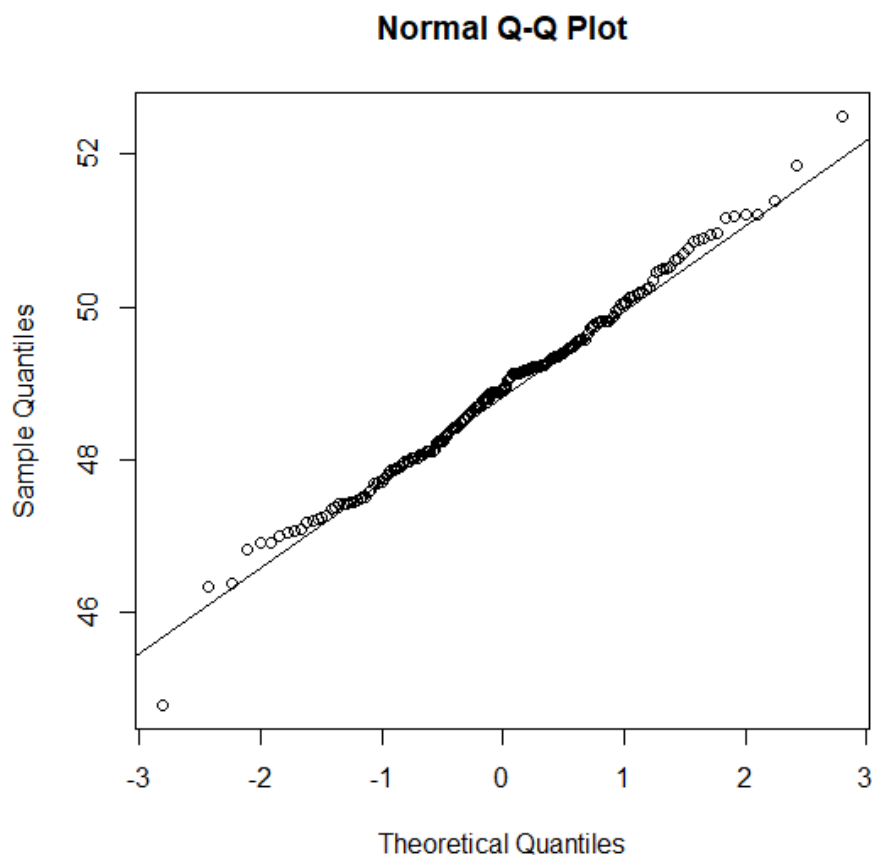
Here the calculated value of chi-square is greater than tabular value,

Conclusion: The calculated value of chi-square are greater than tabular value of chi-square. Therefore we can reject the null hypothesis, hence the accident are dependent on factors.

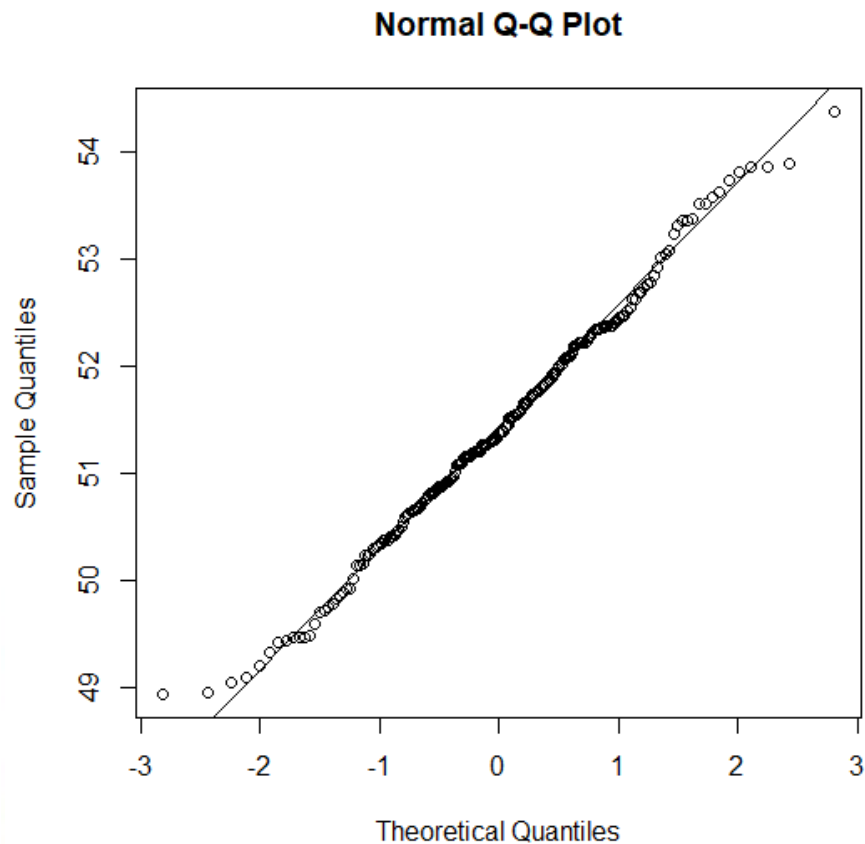
From above conclusion we can say that the factors are responsible for an accident.

❖ **Objective:** TO check whether the consumption of fuel as petrol in rural are and urban area is equal.

We consider the consumption of petrol in rural area and urban area as two sample.



Normal probability plot for urban area. Since the given data is distributed linearly over the qqplot we can say given data not follows normal distribution.



Normal probability plot for rural area. Since the given data is distributed linearly over the qqplot we can say given data not follows normal distribution.

We found that, the normality assumptions is not followed by these two sample

So, we can use the Kruskal-Wallis rank sum test.

We found that, we reject the null hypothesis of equal mean.

Conclusion: The consumption of petrol in rural and urban area is not same.

- ❖ **Objective:** To check monthly many spend on bike is significantly differ Of not in 5 villages.

Hypothesis:

H_0 : The money spend on bike in different villages with identical median.

H_1 : The money spend on bike in different village with median are not equal.

Given data is not normal because we take the Kruskal-Wallis rank sum test.

Dhalgaon	K.Mahankal			Kokale	Nagaj	Kadamwadi
11.6666	8.33333	6.66666	16.666	15.0011	6.66666	11.3333
7.2289	3.2598	7.2541	7.154	16.6666	7.2354	13.3587
6.66666	10.0021	10.8333	13.3333	7.4587	10.8333	
7.2354	13.3333	3.3258	3.2654	8.66666	3.4789	
10.8333	3.2365	7.33333	10.0214	7.2365	7.33333	
3.4789	15.0002	3.4790		9.33333	3.2514	
7.33333	16.6666	11.3333		3.2547	9.33333	
3.2514	7.2587	3.2145		8.254	3.25847	
11.3333	11.6666	14.0325		12.3365	8.2554	
3.5614	7.8547	15.0114			12.365	
14.2389						
15.4147						
Ri^2	=781			=301.5	=228	=90
=335.5						
$\frac{Ri^2}{ni} = 9380$	=24398			=10100	=5198.4	=4050

The test statistics:

$$H = \left[\frac{12}{N(N+1)} \sum_{i=1}^n \frac{Ri^2}{Ni} \right] - 3(N + 1)$$

$$H = 9.30$$

Tabular value $H = 9.48$

Here the tabular value is greater than calculated value.

Conclusion:

Thus we can accept the null hypothesis and conclude that the many spend on bike in different villages are identical median. It means that the money spend on bike in different villages is same.

❖ **Objective:** To analyze opinion about new rule of R.T.O.

(RULE:- According to motor act 1988 of the central government the use of helmet is mandatory, When you crack the rule then you have to pay Rs. 1000 fine.)

Consider the following table:

Opinion about rule	frequency	percentage
Right	377	0.9449
Wrong	22	0.0551

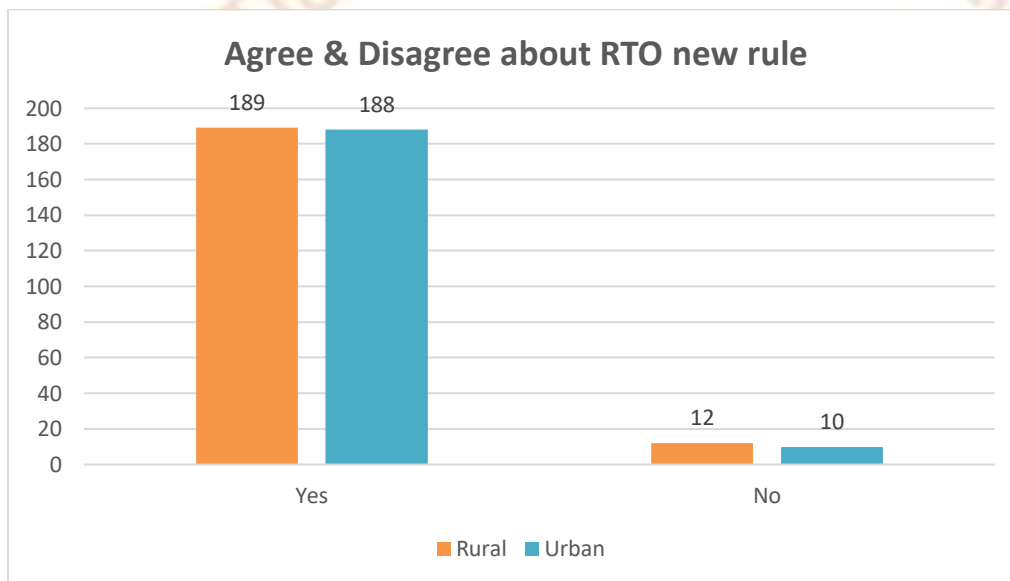


Figure: Showing percentage of respondent who are in favor and against the rule .

Conclusion:

Showing the percentage of respondents who in favor and who are in against the new rule.

About 5.51% people are against this rule. This indicates that, this rule is appropriate according to people.

About 94.49% people are favor in rule. This indicates that, this rule appropriate according to people.

Now, we wanted to see whether the native place of the respondent affects on the opinion. To do this, consider area

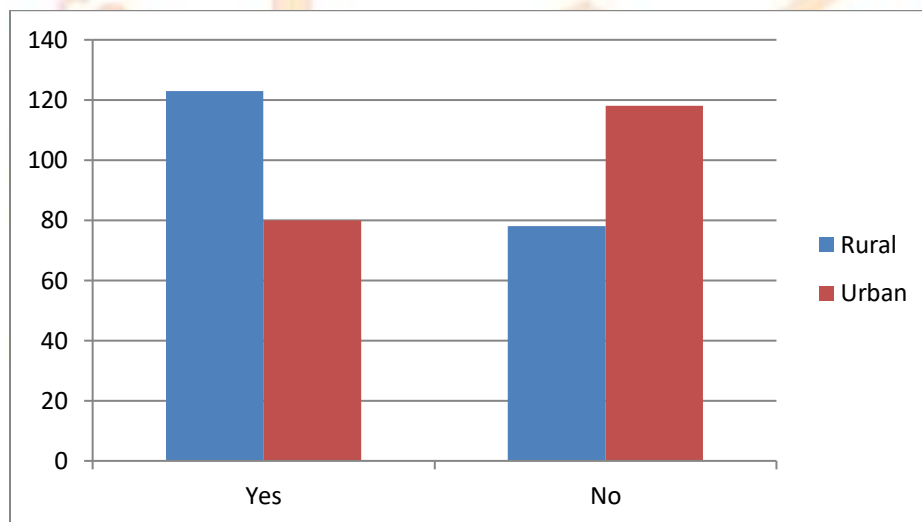
About 100% respondents in urban area and 100% respondents in rural area are respondent. And 94.03% of people are in favor on rule. And 94.95% of people are in favor on rule. We conclude that, most of the people in urban and rural are favor in rule.

❖ **Objective:** To check the health problems caused by prolonged use of bikes in people .

Here we will check health problems caused by prolonged use of bike in both rural & urban areas.

Table:

Region	Yes	No	Total
Rural	123	78	201
Urban	80	118	198
Total	203	196	399



Conclusion:

Here we see health problems cause of bike use is rural area greater than urban area.

We got in rural area 61.19% people's are having health problems cause of bike use & We also got in urban area 40.40% people's are having health problems cause of bike use.

❖ **Objective:** Accupation of costumer & to study of pattern of finance for buying a bike.

Here we will check the accupation of costumer & Check ther payment method used by the costumer while purchasing the bike.

Table:

Row Lables	Businessman	Farmer	Govt.Job	Private Job	Student	Grand Total
Any Other	8	5	6	10	16	45
Bank loan	40	19	30	34	41	164
Finance Company	18	15	12	19	46	110
Self Financeing	13	11	11	17	28	80
Grand Total	79	50	59	80	131	399

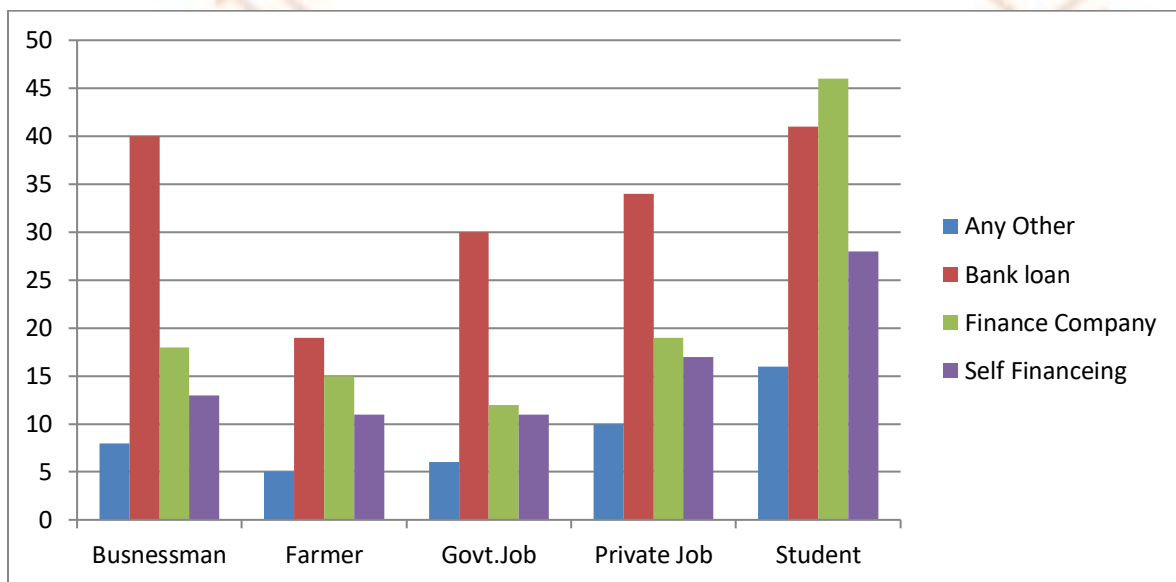


Figure: Showing respondents accuption & payment method while purchasing the bike.

Conclusion: In this graph is showing respondents accuption & payment method while purchasing the bike.

We conclude that peoples from all types of occupation pay through Bank Loans and use Finance Companies while buying a bike. And few peoples use Self Fanancing and Other methods.

Objective: To check correlation between gender of bike users and health problems caused.

Hypothesis:

H₀: p-value <= α: There is no correlation between gender and health problems .

H₁: p-value > α: There is a correlation between gender and health problems.

Where, **p-value** is ratio of expected test statistic and calculated test statistic, and α is level of significance which is 0.05 (i.e. 5%)

Conclusion:

The p-value is greater than level of significance, hence we accept null hypothesis and reject alternate hypothesis.

And here we conclude that there is no correlation between gender of bike users and health problems caused due to prolonged driving.



Major of Findings

- ✦ The most preferred bike is splendor.
- ✦ And the second most preferred bike is Shine.
- ✦ The splendor is used in rural area because it is comfortable, low maintenance.
- ✦ The people gives approximately 5 star to splendor bike.
- ✦ The most of the people use petrol bike, as compared to electric bike.
- ✦ The expenditure on fuel is larger in urban area as compared to rural area. And average expenditure on petrol is Rural = RS. 76.5038 (For fuel price of Rs.110/litre.)
Urban = RS. 92.1275
- ✦ Consumption of petrol in urban area larger as compared with rural area.
- ✦ The proportion of female bike user in urban and rural area are same.
- ✦ The occupation and use of helmet are dependent. Hence the many people they working are use helmet.
- ✦ The accident are dependent on factors.
- ✦ The consumption of petrol in urban and rural are is not same.
- ✦ The moeny spend on the different villages is same.
- ✦ The people are favor in rule, hence the rule are appropriate.
- ✦ The above we can see that the people in urban and rural area are favor in rule.
- ✦ Mostly peoples in kavathe mahankal tahasil is agree with R.T.O. rules
- ✦ Now people are preferring electric vehicles but the pricec of electric vehicles are high.
- ✦ All types of occupation pay through Bank Loans and use Finance Companies while buying a bike. And few peoples use Self Fanancing and Other methods.

LIMITATION

- Sample size(399) is very small to represent the Kavthe Mahankal Tahsil.
- About the electric bike it has small number of observation. So we could not analyze the hypothesis about it.
- In this project, we haven't considered the impact of season on two wheeler bike.

SCOPE FOR FUTURE

- This type of project can be carried out at state or national level.
- This type of project in state level or national level it can used two wheeler company's for launching a new bike on market.
- The analysis and finding can be made more accurate if the data on large scale is available.
- This type of study used in industries for their future in bike industries.
- For ex. The most preferred bike is splendor so that the industries can manufacture these bikes. And their part.

REFERENCE

1. Discrete Multivariate Analysis and Methods

- Yvonne M. Bishop
- Stephen E. Fienberg
- Paul W. Holland

2. Fundamental of mathematical statistics

- S. C. Gupta
- V. K. Kapoor



QUESTIONNAIRE

Section A:

• **Primary information**

1) Email / Name

.....

2) Gender

a) Male

b) Female

3) What is your age?

a) 0-20

b) 20-40

c) 40-60

d) Above 60

4) Occupation

a) Student

b) Farmer

c) Businessman

d) Govt. Job

e) Private Job

5) Region

a) Rural

b) Urban

6) Income (Per year)

a) Below 50 thousand

b) 1-2 lakh

c) 2-4 lakh

d) Above 4 lakh

Section B:

• **Secondary information**

7) Which two wheeler you can use currently?

a) Geared Bike

b) Gearless motor bike

c) Gearless Mopeds

8) How much you average travel per day on your bike?

a) 0-50 km

b) 50-100km

c) 100-150km

d) Above 150

9) Which company's bikes would you like to buy?

a) Hero Honda

b) Yamaha

c) TVS

d) Bajaj

e) Honda

f) Royal Enfield

g) Hero

10) Which model do you use or name of your bike?

.....

11) What is engine capacity of your bike?

a) Less than 100CC

b) 100-125CC

c) Above 125CC

12) Which feature of a bike prefers most of people?

a) Speed

b) Power

c) Mileage

d) Design

e) Brand

f) Pick-up

g) Comfort

13) Where do people want the servicing of his motor bike to be done?

a) Showroom

b) Road Mechanics

14) What is the ranking you would give to fuel consumption of your bike?

a) *

b) **

c) ***

d) ****

15) What is the ranking you would give to engine capacity of your bike?

a) *

b) **

c) ***

d) ****

- 16) What is the ranking you would give to luggage space of your bike?
a) * b) ** c) *** d) ****
- 17) What is the ranking you would give to mileage of your bike?
a) * b) ** c) *** d) ****
- 18) You are use an electric bike or e-vehicle?
a) Yes b) No
- 19) Can you afford electric bike in future?
a) Yes b) No
- 20) What is your expected price to afford an electric bike?
a) 45,000-50,000 b) 50,000-55,000
c) 55,000-65,000 d) Above 65,000
- 21) What is your Source of finance?
a) Bank loan b) Finance Company c) Self financing d) Other
- 22) Which attributes do you like most in your two wheeler?
a) Luggage space b) Fuel efficiency c) Engine power
d) Resale value e) Comfort f) Out look
- 23) In future which brands bike you would like to use?
a) Hero b) Bajaj c) Suzuki d) Royal Enfield
e) KTM f) Honda g) Yamaha h) E-vehicle
- 24) Do you know about the government rule on use helpmate?
(Rule=According to motor act 1988 of a central government, the use of helpmate is mandatory. When you crack the rule then you have pay 500rs fine.)
a) Yes b) No
- 25) You are agreeing government rules?
a) Yes b) No
- 26) You payable a fine for broken a rule of bike or road?
a) Yes b) No
- 27) Any health problem caused by use bike for long time?
a) Yes b) No

✚ This is the Google form link of our questionnaire.

Link :- <https://forms.gle/u4PseFhRBGBBYewU7>

✚ This is the QR code of collected data.



Thank You



A
PROJECT REPORT
ON
“STATISTICAL ANALYSIS OF MILK PRODUCTION”

Submitted to,

Shikshn Prasarak Sanstha's

Padambhushan Vasanttraodada Patil Mahavidyalaya

Kavathe Mahankal, Sangli

Affiliated to Shivaji University, Kolhapur

DEPARTMENT OF STATISTICS

By,

Mr. Pawar Vikas Pradip.

Mr. Babar Suraj Dinkar.

Mr. Sathe Tejas Tatyasaheb.

Mr. Kakade Manish Mahavir.

Mr. Borkar Someshwar Ramesh.

Under the guidance of,

Prof. Dr. Patil M.K. Patil

Prof. Koshti V.V.

Asst. Prof. Patil S.S.

Asst. Prof. Patil P.M.

Through,

Asst. Prof. Satpute G.D.

Padambhushan Vasanttraodada Patil Mahavidyalaya, Kavathe
Mahankal

For the academic year 2022-23
Padambhushan Vasant Rao Dada Patil Mahavidyalaya, Kavathe Mahankal

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for **B.Sc. –III**. They have presently submitted their project entitled “**THE STATISTICAL ANALYSIS OF MILK PRODUCTION**”.

This work is genuine, original and the project of their part of it has not been submitted for any other degree.

Sr.No.	ExamNo.	Name	PRN
1	27570	Mr.Pawar Vikas Pradip.	2020005582
2	27582	Mr.Babar Suraj Dinkar.	2020005692
3	27573	Mr.Sathe Tejas Tatyasaheb.	2020031505
4	27572	Mr.Kakade Manish Mahavir.	2020045620
5	27580	Mr.Borkar Someshwar Ramesh.	2020006457



Teacher in charge



Examiner



Head of Department

DECLARATION

We hereby declare that the project entitled “**THE STATISTICAL ANALYSIS OF MILK PRODUCTION**” with special reference to **All TEACHER’S OF STATISTICS DEPARTMENT**.

It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been summated previously for any other degree of any purpose.

◆ **Mr.Pawar Vikas Pradip.**

◆ **Mr.Babar Suraj Dinkar.**

◆ **Mr.Sathe Tejas Tatyasaheb.**

◆ **Mr.Kakade Manish Mahavir.**

◆ **Mr.Borkar Someshwar Ramesh.**

Place: Kavathe Mahankal

Date:

ACKNOWLEDGEMENT

A truly independent project is contradiction in terms. Every project involves contributions of many people, this project also bears imprints of many people and it is a pleasure to acknowledge all of them.

We express our hearty thanks and deepest gratitude goes to **Dr. M.K.Patil** who made valuable guidance and co-operation during our project work. We are rather inspired by the guidance of **Mr. Koshti sir, Miss. S.S. Patil madam, Mr. G.D. Satpute Miss. Patil. P.M.**, who put us their encouragement, kind support, and constant inspiration during our project work further we are thankful to all of our college students who gave us the necessary information to carry this project.

Finally, we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.

INTRODUCTION

Milk Production has been an important part of the Agricultural scenario for thousands of years. India being A predominantly agrarian economy has about 70 per Cent of its population living in villages, where livestock Play a crucial role in the socio-economic life. In rural area all people are connected directly and indirectly to farming. Animal husbandry is the most used side business of farmer. India is not only one of the top producers of Milk in the world, but also the largest consumer of milk in the world.

Milk sector is instrumental in bringing Socio-economic transformation in India. It has created A lot of employment opportunities and also provides Improved nutritional benefits. In order to ascertain possible to find the expected price of healthy cows by considering somefactors.

CONTENTS

CONTENTS.....	PAGENO.
1) CERTIFICATE.....	02
2) DECLARATION.....	03
3) ACKNOWLEDGEMENT.....	04
4) INTRODUCTION.....	05
5) CONTENTS.....	06
6) OBJECTIVES.....	07
7) SAMPLINGMETHODOLOGY.....	08
8) TOOLS &TECHNIQUESUSEDFORANALYSIS.....	10
9) DATA ANALYSIS.....	11
10) MAJOROFFINDINGS.....	26
11) LIMITATIONS.....	27
12) SCOPEFORFUTURE.....	27
13) REFERENCE.....	28
14) QUESTIONNARIE.....	29

OBJECTIVES

- To construct an equation for predicted price of cow by their age, fat test and average quantity of milk.
- To study which factors affecting on milk production.
- To study involvement of people in milk production
- To investigate whether the quantity of milk and animal feed dependent or not.
- To investigate whether the quantity of milk and animal feed dependent or not.
- To study proportion of people happy with dairy business in cow & buffalo husbandry is same or not.
- To study variation between price of milk among different diaries.
- To analyze month wise variation in milk production for calving.

Sampling Methodology

As the main aim of our project, “Statistical analysis of dairy production” we consider different breeds of cows & buffaloes, age, milk yield, fat test, various feed, price of animal you feel, etc. So we decided to collect all primary data from villages Deshing, Agalgaon, Raywadi, Dhalgaon, Kadamwadi in Kavathe Mahankal Tehsil, which provide us necessary and sufficient data within less time and more accuracy.

So we divide the selected area ad each village, we selected a sample of size 340 by convenient sampling method of the data is collected by filling questionnaires from each people by interview method.

Tools and techniques used for analysis

- **Statistical techniques:**

The following are some statistical tools that lead us towards the appropriate results and findings.

1. **Graphical representation:**

To represent our data graphically, we use the multiple bar diagram, pie chart method, etc. To construct the appropriate divided bar diagram of given data we used Microsoft Excel.

Bar Chart:

A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

Pie chart:

A pie chart displays data, information, and statistics in an easy-to-read ‘pie-slice’ format with varying slice size telling you how much of one data element exists. The bigger the slice the more of that particular data was gathered.

2. Proportion test:

To test the population proportion of two samples.

Hypothesis:

H₀: Two population proportions are same. ($p_1 = p_2$)

Vs

H₁: Two population proportions are not same. ($p_1 \neq p_2$)

Under H₀,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

The test statistic is given by,

Let α be the level of significance and Z_α being the critical value.

Conclusion:

If calculated $|Z| > \text{tabulated } Z$, we reject H₀ at α % l.o.s otherwise we accept it.

3. Test for independence of attributes

To test the independence of attributes, here data are in the form of cross tables.
Where, A_i = Educational Methods/Problems.

B_j = Mode of education

Hypotheses:

H₀: Attributes A_i and B_j are independent of each other. vs

H₁: Attributes A_i and B_j are dependent of each other.

Under H₀ the test statistic is,

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Statistic χ^2 has $(R-1)(C-1)$ d. f.

The value of χ^2 is obtained with the help of formula. To make decision about H_0 , Calculated value of χ^2 is compared with table value of χ^2 for $(R-1)(C-1)$ d. f and α % level of significance (l.o.s)

Conclusion:

If $\text{cal. } \chi^2 \geq \text{tab. } \chi^2$, then we reject H_0 at α % l.o.s. Otherwise we accept H_0 .

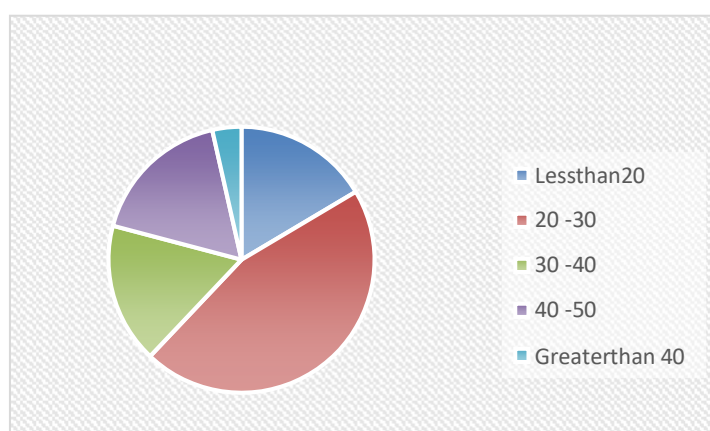
Statistical Analysis

➤ Objectives:

1] To study the involvement of people in milk production by their age and profession & Age wise)

1. Age wise

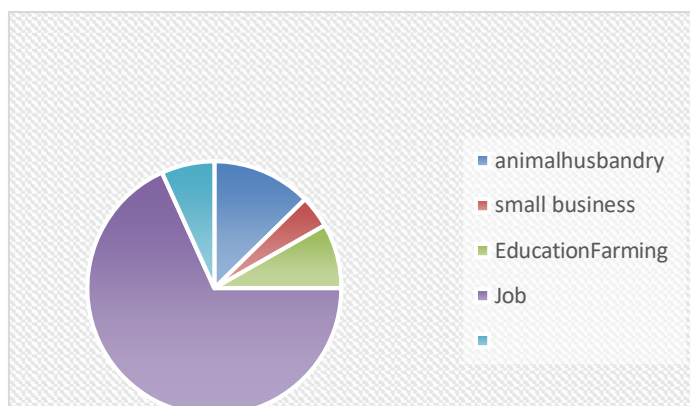
Age	Total
Lessthan 20	56
20 -30	155
30-40	58
40-50	59
Greaterthan40	12



comment: From the above diagram we conclude that, people in age group 20-30 is most active in animal husbandry.

2. Profession Type

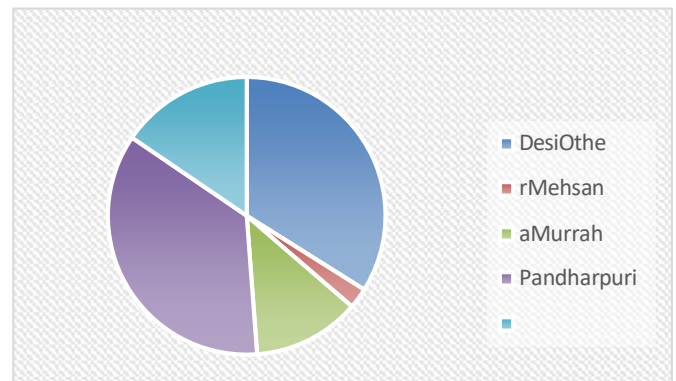
Profession	Count
animal husbandry	43
small business	14
Education	28
Farming	232
Job	23



Conclusion: From the above chart we conclude that, proportion of farmer is greater.

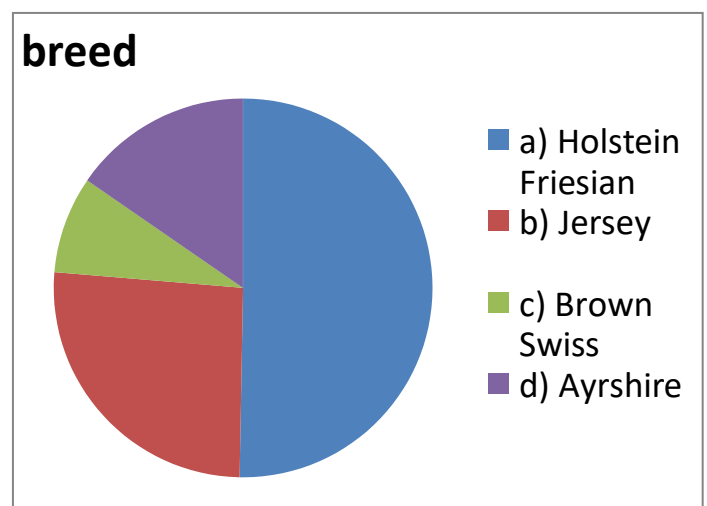
2] To study which breed is mostly used in milk production

Breed	Total
Desi	57
Other	4
Mehsana	21
Murrah	60
Pandharpuri	26



Comment: From the above chart, we conclude that ,murha and mhaisana is most preferd buffalo in buffalo husbandry.

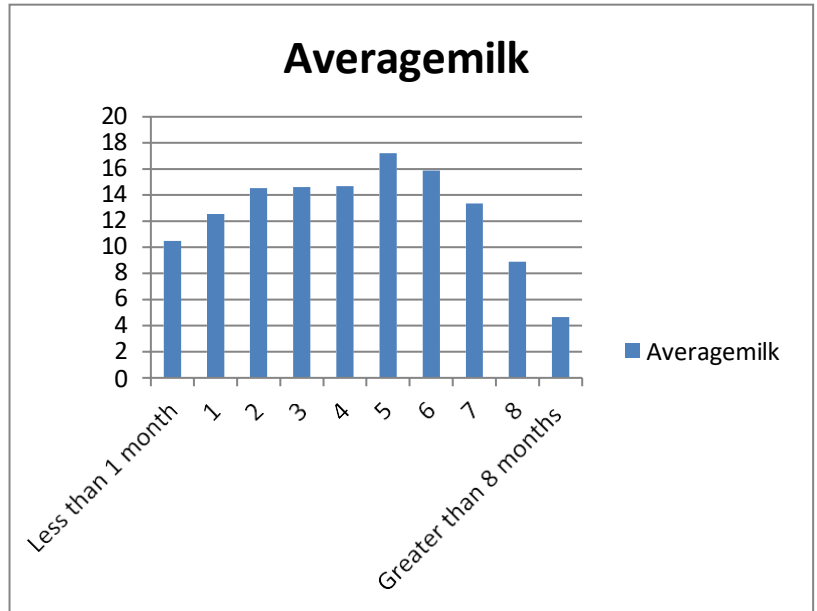
Breed	Total
a)HolsteinFriesian	85
b) Jersey	44
c)BrownSwiss	14
d) Ayrshire	26



Comment: Fromtheabovediagramweconcludethat ,theHolsteinFriesian isthe most common used cow in animal husbandry

3)To analyze monthwise variation in milk production

Lactation month	Average milk
Less than 1 month	10.5
1	12.5645
2	14.5353
3	14.61538462
4	14.68421053
5	17.215
6	15.88888889
7	13.36363636
8	8.909090909
Greater than 8 months	4.666666667



Conclusion: From both the graphs we say that, cow & buffalo gives us more milk in 3 to 6 months of lactation period.

5) To study proportion of people happy with dairy business in cow & buffalo husbandry is same or not.

H₀: the proportion of people happy in. V/S

H₁: the proportion of people happy in milk production.

Let, $\alpha = 5\%$ be the level of significance.

Animal	Cow	Buffalo
Total no. of animal	169(N1)	171(N2)
No. of people happy with this business	140(X1)	123(X2)
P	0.8284(P1)	0.7193(P2)

Calculation:

$$P = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

$$= \frac{169 \cdot 0.8284 + 171 \cdot 0.7193}{169 + 171}$$

$$P = 0.7735$$

$$Q = 1 - P$$

$$Q = 1 - 0.7735$$

$$= 0.226$$

$$Z = \frac{P_1 - P_2}{\sqrt{PQ \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

$$Z = \frac{0.8284 - 0.7193}{\sqrt{0.7735 * 0.2265 * \left(\frac{11}{171169} + \right)}}$$

$$Z = 2.403$$

$$\text{CalZ} = 2.4030$$

Let, alpha be the 5% level of significance. Tab. Z = -1.645

Result: Here, the calculated value of Z is less than the tabulated value of Z. Hence, we accept H_0 at a 5% level of significance.

Conclusion: We conclude that, the proportion of people happy with dairy business in cow & buffalo husbandary is not same.

6) To check type of animal feed for the different quantity of milk are dependent or not

Attributes:

A: Variety of animal feed B:

Quantity of milk

H_0 : Animal feed and quantity of milk are independent
 H_1 : Animal feed and quantity of milk are not independent

Observation Table:1

Class	Freq.	Legum meal	Goli meal	Sarki meal	Maize lime	Total
0 to 6	3	5	10	6	10	31
6 to 10	15	11	24	7	23	65
10 to 14	59	25	191	47	85	348
14 to 18	52	36	136	55	76	303
18 to 22	30	36	46.5	37	43	162.5
22 to 26	10	12	21	11	20	64
Total	169	125	428.5	163	257	973.5

Observation Table 2 :

Sr.no.	Oij	Eij	(Oij-Eij)/Eij
1	5	3.980483	0.261128
2	11	8.346174	0.843835
3	25	44.68413	8.671198
4	36	38.90601	0.217059
5	36	20.86543	10.97773
6	12	8.217771	1.740771
7	10	13.6451	0.973736
8	24	28.61068	0.743023
9	191	153.1772	9.339279
10	136	133.3698	0.05187
11	46.5	71.52671	8.756675
12	21	28.17052	1.825183
13	6	5.19055	0.126231
14	7	10.88341	1.385676
15	47	58.2681	2.179068
16	55	50.73344	0.358808
17	37	27.20853	3.523637
18	11	10.71597	0.007528
19	10	8.183873	0.403027
20	23	17.15973	1.987719
21	85	91.87057	0.513818
22	76	79.99076	0.1991
23	43	42.89933	0.000236
24	20	16.89574	0.570348
25			55.65668

Now, calculated χ^2 is given by,

Here, $m=6$ & $n=4$

m

$$\chi^2 = \sum \sum ((O_{ij} - E_{ij}) / E_{ij})^2 = 55.6567$$

$$i=1j=1$$

The table value of χ^2 with $(m-1)(n-1)$ d.f. at 5% l.o.s. is,

$$\text{Tab}\chi^2(15,0.05) = 24.9958$$

Result:

Here, calculated $\chi^2 > \text{table}\chi^2$.

There fore we reject H_0 , we conclude that Animal feed and quantity of milk are not independent.

7] To study the correlation between water and quantity of milk:

We have to study the correlation between water and quantity of milk. For this we use Karl Pearson's coefficient of correlation,

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

$$r_1 = 0.680085$$

$$r_2 = 0.659303$$

where, r_1 = correlation coefficient between quantity of milk and the water they drink for buffalo

where, r_2 = correlation coefficient between quantity of milk and the water they drink for cow

Result: Here, we see that there is a positive correlation between quantity of milk and the water they drink. There is a linear relationship between quantity of milk and the water they drink .

Conclusion: As they drink more water, they will have milk

8] To derive an equation for predicted price of cow:

- For finding expected price of healthy cows we use regression model.

➤ RegressionStatistics	
➤	
➤ MultipleR	➤ 0.992797471
➤ R Square	➤ 0.985646818
➤ AdjustedR Square	➤ 0.968262492
➤ StandardError	➤ 9097.605572
➤ Observations	➤ 64

RegressionStatistics	Coefficients
MultipleR	0
lactationmonth	516.0351794
Age	-3621.55974
Milkinliter	4797.394054
Fattest	5135.773803

We observed that the R- Square value is .9856 that mean the regression equation is goodfit.

Resultant regression equation for healthy cows is

$$\text{ExpectedPrice} = 516 * \text{lactationmonth} - 3622 * \text{age} + 4797 * \text{milkinlitre} + 5136 * \text{fattest}$$

Conclusion:-

From this project we conclude that Murha & Holstein Freinsen are reared in large numbers. Peoples of age group 20 to 40 is most active on animal husbandry as well as the fixed output was that farmers raise more livestock. During the 2nd,3th,4th,5th months of lactation, animals produce more milk. From this project we observe that quantity of milk, fat test and age are affected factors on price of that animal and fodder, animal feed, water affected on milk production. We saw correlation between milk and water and animals that give more milk drink more water as well as animal who eat more feed produces more milk. That means water & animal feed is affected of their milk quantity and fat test

LIMITATIONS

- 1)** We have to limit the scope of this research according to the result we obtain from the questionnaires.
- 2)** We consider the whole population is normal but in reality, it may be or may not be normal.
- 3)** The time constraint was one of the major limitations.
- 4)** We collect data through but people not give us 100% correct information.

SCOPE

- 1)** A well-planned Study can be taken including some characteristics.
- 2)** It can be extended to a greater region for better estimates.

Field Project of B. Sc. III Students (2021-22)

Sr. No.	Name of the Students	Title of the Project	Name of Mentor
1	Koshti Akanksha Vijay.	The impact of covid-19 on social welfare.	Mr. V. V. Koshti
	Shinde Utkarsha Ananda.		
	Suryavanshi Rutuja Sanjay.		
	Jadhav Radhika Vinayak.		
	Gidde Pooja Mahadev.		
2	Jagdale Anuradha Subhash.	Effect of online education system on quality of education.	Miss. S.S. Bhosale
	Ganeshwade Pragti Balaso.		
	Patil Nikita Nemgonda.		
	Kadam Pooja Nayku.		
	Chougule Pramila Parsu.		
3	Patil Suraj Govind.	Statistical Analysis of effect of spending time on computer/mobile platforms and on screen activities on real life.	Mr. A. P. Patil
	Kumbhar Ashitosh Sahadev.		
	Desai Nikhil Aanadrao.		
	Pawar Onkar Vilas.		
	Gurav Pankaj Keshav.		
4	Suryawanshi Arati Jaysing	Role of media on human development.	Mr. G. D. Satpute
	Pawar Mahadevi Machhindra.		
	Shinde Nisha Sanjay.		
	Savale Prajakta Popat.		
	Gidde Ashwini Dilip.		

*A
Project Report
On*

“THE IMPACT OF COVID-19 ON SOCIAL WELFARE”

Submitted to

**Padmabhushan Vasantodada Patil Mahavidyalaya,
KavatheMahankal.
DEPARTMENT OF STATISTICS**

By,

1. Miss. Koshti Akanksha Vijay
2. Miss. Shinde Utkarsha Ananda
3. Miss. Suryavanshi Rutuja Sanjay
4. Miss. Jadhav Radhika Vinayak
5. Miss. Gidde Pooja Mahadev

Under the Guidance of,

**Prof. Dr. Patil M. K.
Prof. Koshti V. V.**

**Prof. Patil A. P.
Prof. Bhosale S. S.
Prof. Patil S. S.
Prof. Satpute G.D.**

Through,

**The Principal,
Padmabhushan VasantodadaPatil
Mahavidyalaya, Kavathe Mahankal.**

For the academic year 2021-2022

**Padmabhushan Vasantodada Patil Mahavidyalaya, Kavathe
Mahankal**

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for the B. Sc. Part III. They have presently submitted their project entitled “THE IMPACT OF COVID-19 ON SOCIAL WELFARE”

This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

1. Miss. Koshti Akanksha Vijay
2. Miss. Shinde Utkarsha Ananda
3. Miss. Suryavanshi Rutuja Sanjay
4. Miss. Jadhav Radhika Vinayak
5. Miss. Gidde Pooja Mahadev



V. V. Koshti (Asst. Prof.)
Teacher in charge



Examiner



Prof. (Dr.) M. K Patil
Head of Department

DECLARATION

We hereby declare that the project entitled “THE IMPACT OF COVID-19 ON SOCIAL WELFARE”.

It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been submitted previously for any other degree of any other purpose.

Sr. No.	Name	PRN
1.	Koshti A. V.	2019012610
2.	Shinde U. A.	2019010723
3.	Suryavanshi R. S.	2019012477
4.	Jadhav R. V.	2019012550
5.	Gidde P. M.	2019012515

Place : Kavathe Mahankal

Date :

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude to Dr. Patil M. K. who gave backing and his valuable guidance for this project work. We further thank Prof. Koshti V.V. ,Prof. Bhosale S.S., Prof. Patil A.P., Prof. Satpute G.D. and Prof. Patil S.S. , who gave us their encouragement, kind support, and constant inspiration during our project. We are thankful to all the people in our taluka who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.

CONTENTS

Sr. No.	Name	Page No.
1	Introduction	6
2	Objective	9
3	Data collection and Sampling Method	10
4	Statistical tools used	11
	1) Graphical representation	
	2) Paired T test	
	3) Z test for population proportions	
5	Analysis of Data	14
6	Summary	32
7	Bibliography and Questionnaire	33

INTRODUCTION

We are facing a global health crisis unlike any in the history of the World — one that is killing people, spreading human suffering, and upending people's lives. But this is much more than a health crisis. It is a human, economic and social crisis. The corona virus disease (COVID-19), which has been characterized as a pandemic by the World Health Organization (WHO), is attacking societies at their core.

The COVID-19 outbreak affects all segments of the population and is particularly detrimental to members of those social groups in the most vulnerable situations, continues to affect populations, including people living in poverty situations, older persons, persons with disabilities, youth, and indigenous peoples.

Older persons are particularly susceptible to the risk of infection from COVID-19, especially those with chronic health conditions such as hypertension, cardiovascular disease and diabetes.

Older persons are not just struggling with greater health risks but are also likely to be less capable of supporting themselves in isolation. Although social distancing is necessary to reduce the spread of the disease, if not implemented correctly, such measures can also lead to increased social isolation of older persons at a time when they may be in most need of support.

In terms of employment, youth are disproportionately unemployed, and those who are employed often work in the informal economy or gig economy, on precarious contracts or in the service sectors of the economy, that are likely to be severely affected by COVID-19.

What is Social Welfare?

Welfare, or commonly social welfare, is a type of government support intended to ensure that members of a society can meet basic human needs such as food and shelter. Welfare may also encompass efforts to provide a basic level of well being through free or subsidized *social services* such as healthcare, education, vocational training, and public housing. In a welfare state, the state assumes responsibility for the health, education, and welfare of society, providing a range of social services such as those described.

The Central Government of India's social programmes and welfare expenditures are a substantial portion of the official budget, and state and local governments play roles in developing and implementing social security policies. Additional welfare

measure systems are also uniquely operated by various state governments. The government uses the unique identity number (Aadhar) that every Indian possesses to distribute welfare measures in India. As of 2020, the government's expenditure on social programme and welfare (direct cash transfers, financial inclusion, benefits, health and other insurances, subsidies, free school meals, rural employment guarantee), was approximately fourteen lakh crore rupees (\$192 billion), which was 7.3% of gross domestic product (GDP).

Indicators of Social Welfare:

The indicators used to measure Social Welfare are Poverty rate, Employment rate, People suffering from mental health problems, Physical health, Life expectancy, Energy consumption, Cleanliness and hygiene, Financial background, Social Security

Government schemes:

Government schemes associated with the Indicators of Social Welfare mentioned above are PM Jan Dhan Yojana, National Rural Employment Guarantee Scheme, National Mental Health Programme, PM Jan Arogya Yojana, National Health Policy PM Jeevan Jyoti Bheema Yojana, Bachat Lamp Yojana, Swachha Bharat Abhiyaan, National Social Assistance Scheme and Public Provident Fund.

About Our Project:

Our prime objective while carrying out the project was to bring out the real effects of COVID-19 on some of the important indicators of Social Welfare and to check which of the indicators had the most impact on the lives of the people. While collecting the data, it was our aim to collect the data in such an order that we would be able to compare the data before, in-between and post COVID-19.

We sampled 463 families from various villages of Kavathe Mahankal Tehsil belonging to different social sections of society (white/ saffron/ yellow ration-card holders). As we went forward with our project, we observed that the impact of COVID-19 varied among the families of different social sections. Families with income earners belonging to the service sector had comparatively lesser impact than those with the breadwinner of the families belonging to the working class.

We discussed the impact COVID-19 had on the health of the subjects and which

measures were taken by them to fight the virus. The measures include physical exercises, yoga, healthy diet and home remedies/ ayurvedic medicines.

In order to check whether COVID-19 brought about awareness among the people regarding the government schemes, we asked them about various schemes and noted whether they had the knowledge of the scheme before COVID or did COVID brought about this awareness.

The intention behind this project is to analyze the grave flexibility of the social welfare system and its capability to tackle unbearable situations like COVID-19, because we are spending hefty amounts to get to Mars but one unprecedented 80 nm sized virus could shut the earth down, and make us question its reliability.

This project will provide all the aspects of the social welfare system and the major indicators of the same. It will provide an insight on the impact the COVID-19 had on the lives of the people from Kavathe Mahankal Tehsil.

OBJECTIVES OF STUDY

1. To study whether the income and the expenditure of the families under study is affected due to COVID-19 pandemic.
2. To study whether the health problems are increased during the COVID-19 pandemic.
3. To know whether the awareness about the health precautions was increased during COVID-19 pandemic.
4. To know whether the awareness about the various Government Social Welfare schemes was increased during COVID-19 pandemic
5. To study whether the electricity consumption of the families under study is increased due to COVID-19 pandemic.
6. To investigate whether the COVID-19 pandemic results in the drop-out of the students.
7. To study whether the awareness about the cleanliness was brought about by the COVID-19 pandemic.
8. To investigate whether the COVID-19 pandemic results in unemployment.

Source of Data & Sampling Method

The main objective of this project is, “To understand the impact of COVID-19 on Social Welfare”. We have sampled people from different sections of the society to ensure that conclusions apply to various situations. Also this project is intended to know about the current financial and physical state of our subjects.

We designed a questionnaire to fulfill the objectives behind this project accurately. The data collection procedure was done using one-to-one approach. We sampled 463 families from Kavathe Mahankal Tehsil and asked them each question in our questionnaire carefully.

We sampled subjects using simple random sampling without replacement (SRSWOR) method. The collection of sampling has been taken through direct interviewing with the head of the family.



STATISTICAL TOOLS

Following are some statistical tools used towards the appropriate results and finding.

• *Graphical Representation*

To represent our data graphically, we use the multiple bar diagram, pie chart method etc. To construct the appropriate divided bar diagram of given data we used Microsoft Excel.

- i. ***Pie Chart:*** A pie chart displays data, information and statistics in an easy-to-read 'pie-slice' format with varying slice sizes telling you how much of on data element exists. The bigger the slice the more of that particular data was gathered.
- ii. ***Bar Chart:*** A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

• *Paired t test*

To test the difference between the set of before and after the pandemic observations, we use Paired-t test.

Here data is in the form of before and after type.

Where, X_i = values of the variable before pandemic and

Y_i = values of the variable after pandemic. ($d_i = Y_i - X_i$)

H_0 : $\mu_d = 0$ ($\mu_{y-x} = 0$)

i.e. The difference between the means is insignificant.

Vs

H₁: $\mu_d \neq 0$ ($\mu_{y-x} \neq 0$)

i.e. the difference between the means is significant.

Under H₀ the test statistics is,

$$t = \frac{\bar{d}}{\frac{S_d}{\sqrt{n}}}$$

where $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$

$$S_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}$$

Statistic t has (n-1) d .f.

The value of t is obtained with the help of formula. To make decision about H₀, Calculated value of t is compared with table value of t for (n-1) d. f at 5 % level of significance (l.o.s)

Conclusion:

If cal. $t \geq$ tab. t, then we reject H₀ at 5 % l. o. s. Otherwise we accept H₀.

• **Test for population proportion:**

1. One Sample Proportion Test:

To test the hypothesis that the population proportion P has a specified value P₀.

H₀: Population proportion is equal to P₀ (P=P₀) .

Vs

H₁: Population proportion is not equal to P₀. (P ≠ P₀)

Under H₀, the test statistic is given by,

$$Z_{calc} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Let α be the level of significance and Z_α being the critical value.

Conclusion:

If calculated $|Z| >$ tabulated Z , we reject H_0 at α % l.o.s otherwise we accept it.

2. Two sample proportion test:

To test the population proportion of two samples.

H₀: Two proportions are same. ($P_1 = P_2$) .

Vs

H₁: Two population proportions are not same. ($P_1 \neq P_2$)

Under H_0 , the test statistic is given by,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Let α be the level of significance and Z_α being the critical value.

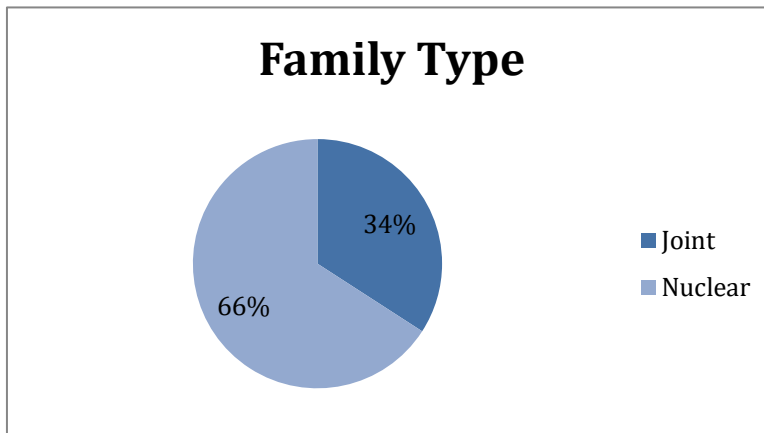
Conclusion:

If calculated $|Z| >$ tabulated Z , we reject H_0 at α % l.o.s otherwise we accept it.

STATISTICAL ANALYSIS

1. To get the basic idea about the sample under study.

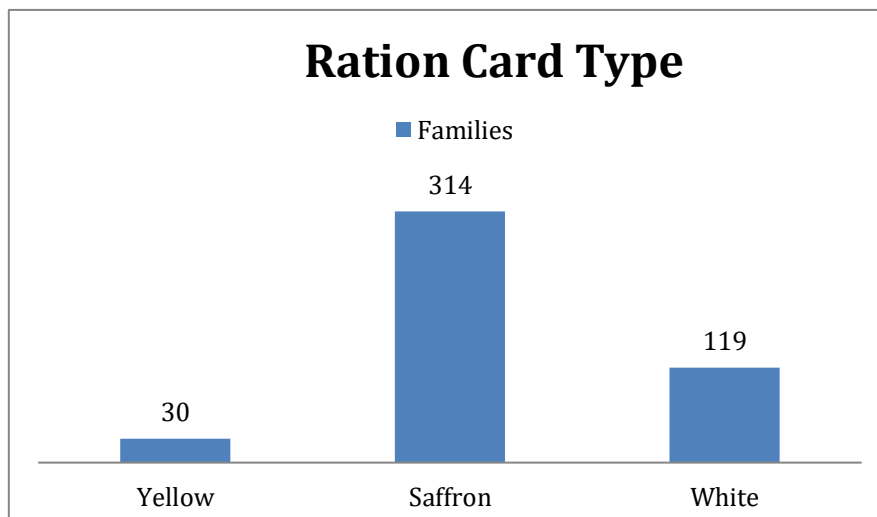
i) Is your Family type joint or nuclear?



Conclusion:

From the above graph we conclude that, 66% of families are nuclear while the remaining 34% are joint families.

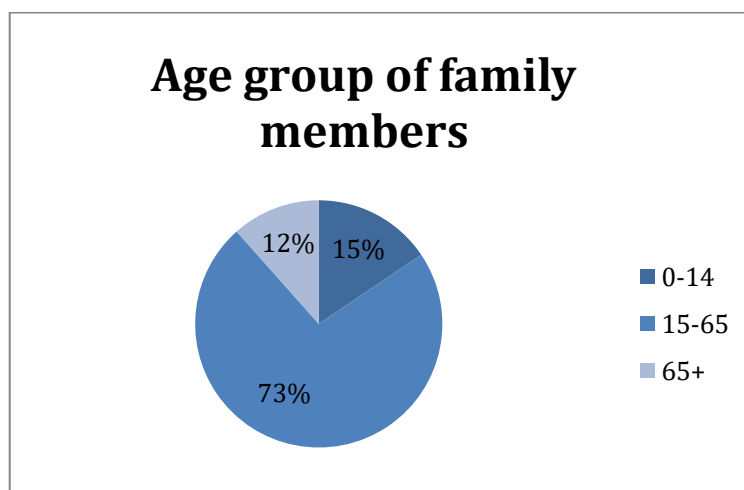
ii) What color ration card does your family hold?



Conclusion:

From the above graph we conclude that, most of the families hold saffron Ration card, followed by white and yellow respectively.

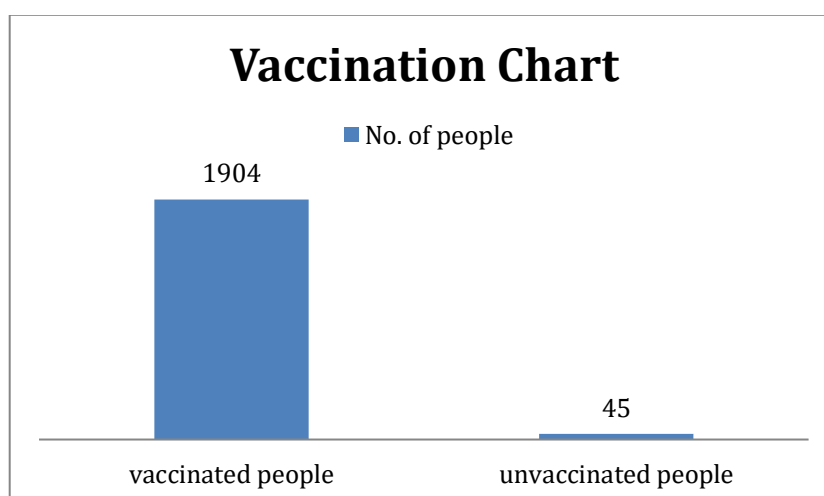
iii) What is the age group of each family member in your family?



Conclusion:

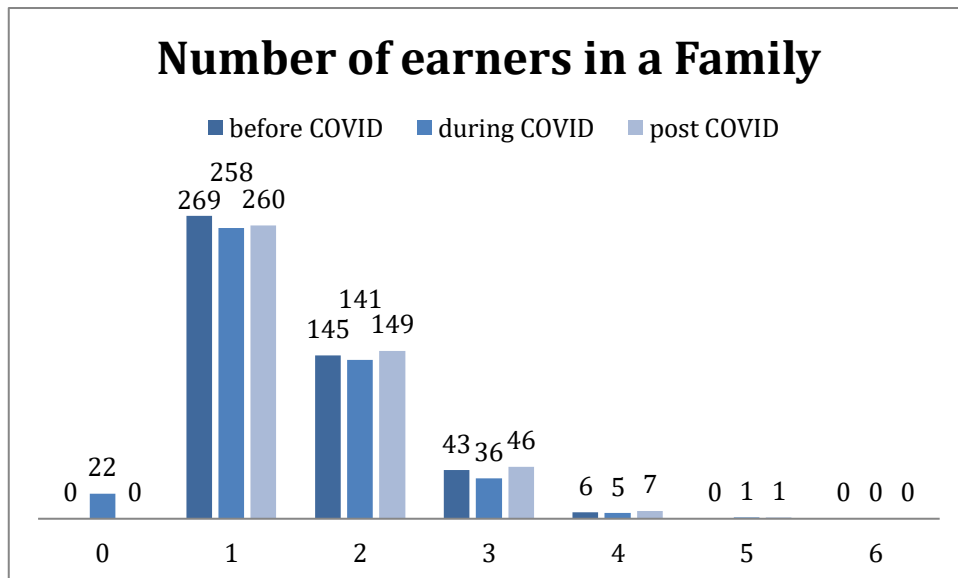
From the above graph we conclude that, almost $\frac{3}{4}$ th of the total people falls under 15-65 age group followed by 0-14 age group and 65+ age group respectively.

iv) Did you get vaccinated for Covid-19?



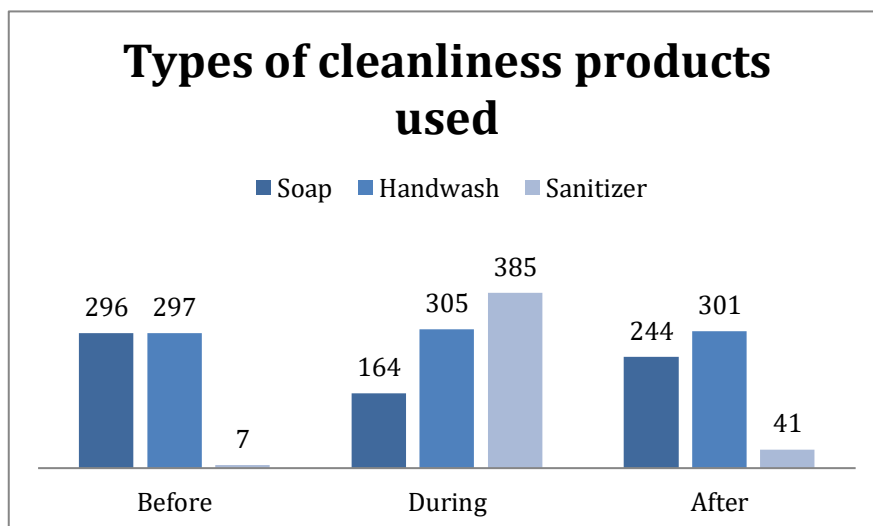
Conclusion: From the above graph we can conclude that, the number of vaccinated people is significantly greater than that of unvaccinated people.

- v) Total number of earners in a family before COVID, during COVID and post COVID



Conclusion: From the above graph we can conclude that, the number of earners in a family decreased during COVID, but the number increased again post COVID.

- vi) Total number of families using various types of cleanliness products, before COVID, during COVID and post COVID



Conclusion: From the above graph we can conclude that,

- a. The use of soap decreased during COVID. It again increased after COVID.
- b. The use of handwash remained almost the same throughout the pandemic.
- c. More families used sanitizer during COVID, but the number falls after the COVID.

2. To test whether the income is affected due to COVID-19 pandemic:

Here we use Paired t-test of two samples for mean to test whether the mean difference between the income of families before and during COVID is zero or not.

H₀: There is no significant difference between the income of families before and during COVID-19.

i.e. $\mu_x = \mu_y$ **against**

H₁: There is significant difference between the income of families before and during COVID-19.

i.e. The income of families decreased during the Pandemic .

i.e. $\mu_x > \mu_y$

Observation Table:

	<i>Before COVID</i>	<i>During COVID</i>
Mean	41544.44276	37250.70626
Variance	1751798098	1752170363
Observations	463	463
Pearson Correlation	0.975975266	
Hypothesized Mean Difference	0	
Df	462	
t Stat	10.06969797	
P(T<=t) one-tail	5.37523E-22	
t Critical one-tail	1.648158493	
P(T<=t) two-tail	1.07505E-21	
t Critical two-tail	1.965111939	

Under H₀, The test statistics is

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

where $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$

$$s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1}$$
$$= 10.0696$$

Tabulated value of t_{462} at 5% level of significance = 1.6481

Result: Here, Calculated value of t is 10.0696 > tabulated value of t is 1.6481, hence we reject H_0 .

Conclusion: From above, we conclude that the income of families before COVID was more than that of during COVID.

3. To test whether the expenditure is affected due to COVID-19 pandemic:

Here we use Paired t-test of two samples for mean to test whether the mean difference between the expenditure of families before and during COVID is zero or not.

H_0 : There is no significant difference between the expenditure of families before and during COVID-19.

i.e. $\mu_x = \mu_y$ against

H_1 : There is significant difference between the expenditure of families before and during COVID-19.

i.e. The income of families decreased during the Pandemic .

i.e. $\mu_x > \mu_y$

Observation Table:

	Variable 1	Variable 2
Mean	15844.49244	9506.695464
Variance	179725765.1	64701275.42
Observations	463	463
Pearson Correlation	1	
Hypothesized Mean Difference	0	
Df	462	
t Stat	25.43101281	
P(T<=t) one-tail	3.64498E-90	
t Critical one-tail	1.648158493	
P(T<=t) two-tail	7.28996E-90	
t Critical two-tail	1.965111939	

Under H_0 , The test statistics is,

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

$$\text{where } \bar{d} = \frac{\sum_{i=1}^n d_i}{n}$$

$$s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1} = 25.4310$$

Tabulated value of t_{462} at 5% level of significance = 1.6481

Result: Here, Calculated value of t is $25.4310 >$ tabulated value of t is 1.6481, hence we reject H_0 .

Conclusion: From above, we conclude that the expenditure of families is significantly decreased during COVID.

4. To test whether the awareness about cleanliness is increased due to COVID-19 pandemic:

Here we use Paired t -test of two samples for mean to test whether the mean difference between the number of times people from a family wash their hands before and during COVID is zero or not.

H_0 : There is no significant difference between the cleanliness habit of families before and during COVID-19.

i.e. $\mu_x = \mu_y$ **against**

H_1 : There is significant difference between the cleanliness habit of families before and during COVID-19.

i.e. The awareness about cleanliness of families increased during the Pandemic

i.e. $\mu_x < \mu_y$

Observation Table:

	<i>Before COVID</i>	<i>During COVID</i>
Mean	5.332613391	9.64362851
Variance	6.560124541	36.95281105
Observations	463	463
Pearson Correlation	0.44651671	
Hypothesized Mean Difference	0	
Df	462	
t Stat	-17.04749062	
P(T<=t) one-tail	3.29011E-51	
t Critical one-tail	1.648158493	
P(T<=t) two-tail	6.58022E-51	
t Critical two-tail	1.965111939	

Under H_0 , The test statistics is,

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

where $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$

$$s_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1} = 25.4310$$

Tabulated value of t_{462} at 5% level of significance = 1.6481

Result: Here, Calculated value of t is $25.4310 >$ tabulated value of t is 1.6481, hence we reject H_0 .

Conclusion: From above, we conclude that there is a significant increase in the number of times people wash their hands during COVID as against that of before COVID.

5. To test whether the electricity consumption is influenced due to COVID-19 pandemic:

Here we use Paired t-test of two samples for mean to test whether the mean difference between the electricity bills a family received before and during COVID is zero or not.

H₀: There is no significant difference between the electricity consumption of families before and during COVID-19.

i.e. $\mu_x = \mu_y$ **against**

H₁: There is significant difference between the electricity consumption of families before and during COVID-19.

i.e. The electricity consumption is increased during the Pandemic .

i.e. $\mu_x < \mu_y$

Observation Table:

	<i>Before COVID</i>	<i>During COVID</i>
Mean	510.8855292	597.2354212
Variance	123009.3873	173513.5527
Observations	463	463
Pearson Correlation	0.83896029	
Hypothesized Mean Difference	0	
Df	462	
t Stat	-8.196458335	
P(T<=t) one-tail	1.22986E-15	
t Critical one-tail	1.648158493	
P(T<=t) two-tail	2.45972E-15	
t Critical two-tail	1.965111939	

Under H_0 , The test statistics is

$$t = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$$

where $\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$

$$S_d^2 = \frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n-1} = 25.4310$$

Tabulated value of t_{462} at 5% level of significance = 1.6481

Result: Here, Calculated value of t is $25.4310 >$ tabulated value of t is 1.6481, hence we reject H_0 .

Conclusion: From above, we conclude that there is a significant increase in the number of times people wash their hands during COVID as against that of before COVID.

6. To test whether the proportion of awareness among people about health precautions is influenced due to COVID-19 pandemic:

a) Physical Exercise:

Here we use two-proportion Z-test of to determine whether the proportion of people doing Physical exercises before COVID is different from that of post COVID.

H_0 : The proportion of awareness about physical exercise before and after the pandemic is same. **i.e. ($p_1 = p_2$) against**

H_1 : The proportion of awareness about physical exercise after the pandemic is increased. **i.e. ($p_1 < p_2$)**

Observation Table:

	Before COVID	Post COVID
No. of people exercised	141	330
No. of people didn't exercise	322	133
Total	463	463
P	0.304535637	0.712743
\hat{p}	0.508639309	
\hat{q}	0.491360691	
Z-stat	-12.4237054	
p-value	9.71735E-36	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$= -12.4237$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is $-12.4237 <$ tabulated value of Z is -1.6448 . Hence we accept H_0 .

Conclusion: We conclude that, there is a significant difference between the number of people who did physical exercises before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about Physical exercises.

b) Yoga:

Here we use two-proportion Z-test of to determine whether the proportion of people doing Physical exercises before COVID is different from that of post COVID.

H_0 : The proportion of awareness about yoga before and after the pandemic is same. **i.e. ($p_1 = p_2$) against**

H_1 : The proportion of awareness about yoga after the pandemic is increased. **i.e. ($p_1 < p_2$)**

Observation Table:

	Before COVID	Post COVID
No. of people practiced Yoga	190	263
No. of people didn't practice Yoga	273	200
Total	463	463
P	0.410367	0.568035
\hat{p}	0.489201	
\hat{q}	0.510799	
Z-stat	-4.79898	
p-value	7.97E-07	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$= -4.7989$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is -4.7989 < tabulated value of Z is -1.6448. Hence we accept H₀.

Conclusion: We conclude that, there is a significant difference between the number of people who practiced yoga before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about We can say that, COVID-19 brought about awareness among the people about Yoga.

c) Ayurvedic Remedies:

Here we use two-proportion Z-test of to determine whether the proportion of people using Ayurvedic remedies before COVID is different from that of post COVID.

H₀: The proportion of awareness about Ayurvedic remedies before and after the pandemic is same. **i.e. (p₁ = p₂) against**

H₁: The proportion of awareness about Ayurvedic remedies after the pandemic is increased. **i.e. (p₁ < p₂)**

Observation Table:

	Before COVID	Post COVID
No. of people used Ayurvedic remedies	129	333
No. of people didn't use Ayurvedic remedies	334	130
Total	463	463
P	0.278618	0.719222
\hat{p}	0.49892	
\hat{q}	0.50108	
Z-stat	-13.4077	
p-value	2.72E-41	

Under H₀ Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -13.4077$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is -13.4077 < tabulated value of Z is -1.6448. Hence we accept H₀.

Conclusion: We conclude that, there is a significant difference between the number of people who used Ayurvedic remedies before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about Ayurvedic remedies.

d) Diet:

Here we use two-proportion Z-test of to determine whether the proportion of people following a balanced diet before COVID is different from that of post COVID.

H₀: The proportion of awareness about a balanced diet before and after the pandemic is same. **i.e. (p₁ = p₂) against**

H₁: The proportion of awareness about a balanced diet after the pandemic is increased. **i.e. (p₁ < p₂)**

Observation Table:

	Before COVID	Post COVID
No. of people followed a balanced diet	298	434
No. of people didn't follow a balanced diet	165	29
Total	463	463
P	0.643629	0.937365
\hat{p}	0.790497	
\hat{q}	0.209503	
Z-stat	-10.9822	
p-value	2.33E-28	

Under H₀ Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$= -10.9822$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is -10.9822 < tabulated value of Z is -1.6448. Hence we accept H₀.

Conclusion: We conclude that, there is a significant difference between the number of people who followed a balanced diet before COVID and that of after the COVID-19 pandemic. The number of people followed a balanced after the COVID has increased significantly.

7. To test whether the proportion of awareness about Government Schemes to aid social welfare (PM Jan Dhan Yojana, PM Jan Arogya Yojana, PM Ujwala Yojana and Swachha Bharat Abhiyan) is influenced due to COVID-19 pandemic:

a) PM Jan Dhan Yojana:

Here we use two-proportion Z-test of to determine whether the proportion of people aware about PM Jan Dhan Yojana before COVID is different from that of post COVID.

H₀: The proportion of awareness about PM Jan Dhan Yojana before and after the pandemic is same. i.e. (p₁ = p₂) **against**

H₁: The proportion of awareness about PM Jan Dhan Yojana after the pandemic is increased. i.e. (p₁ < p₂)

Observation Table:

	Before COVID	Post COVID
No. of people aware about PM Jan Dhan Yojana	197	391
No. of people unaware about PM Jan Dhan Yojana	266	72
Total	463	463
P	0.425485961	0.844492
\hat{p}	0.634989201	
\hat{q}	0.365010799	
Z- statistic	-13.2422021	
p- value	2.50296E-40	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -13.2422$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is $-13.2422 <$ tabulated value of Z is -1.6448 . Hence we accept H_0 .

Conclusion: We conclude that, there is a significant difference between the number of people who were aware about PM Jan Dhan Yojana before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about PM Jan Dhan Yojana.

b) PM Jan Arogya Yojana:

Here we use two-proportion Z-test of to determine whether the proportion of people aware about PM Jan Arogya Yojana before COVID is different from that of post COVID.

H_0 : The proportion of awareness about PM Jan Arogya Yojana before and after the pandemic is same. i.e. ($p_1 = p_2$) **against**

H_1 : The proportion of awareness about PM Jan Arogya Yojana after the pandemic is increased. i.e. ($p_1 < p_2$)

Observation Table:

	Before COVID	Post COVID
No. of people aware about PM Jan Arogya Yojana	165	357
No. of people unaware about PM Jan Arogya Yojana	298	106
Total	463	463
P	0.356371	0.771058
\hat{p}	0.563715	
\hat{q}	0.436285	
Z- statistic	-12.7227	
p- value	2.21E-37	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = -12.7227$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is $-12.7227 <$ tabulated value of Z is -1.6448 . Hence we accept H_0 .

Conclusion: We conclude that, there is a significant difference between the number of people who were aware about PM Jan Arogya Yojana before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about PM Jan Arogya Yojana.

c) PM Jan Ujwala Yojana:

Here we use two-proportion Z-test of to determine whether the proportion of people aware about PM Jan Ujwala Yojana before COVID is different from that of post COVID.

H_0 : The proportion of awareness about PM Jan Ujwala Yojana before and after the pandemic is same. **i.e. ($p_1 = p_2$) against**

H_1 : The proportion of awareness about PM Jan Ujwala Yojana after the pandemic is increased. **i.e. ($p_1 < p_2$)**

Observation Table:

	Before COVID	Post COVID
No. of people aware about PM Jan Ujwala Yojana	172	357
No. of people unaware about PM Jan Ujwala Yojana	291	106
Total	463	463
P	0.37149	0.771058
\hat{p}	0.571274	
\hat{q}	0.428726	
Z- statistic	-12.2844	
p- value	5.49E-35	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$= -12.2844$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is -12.2844 < tabulated value of Z is -1.6448. Hence we accept H_0 .

Conclusion: We conclude that, there is a significant difference between the number of people who were aware about PM Jan Ujwala Yojana before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about PM Jan Ujwala Yojana.

d) Swachha Bharat Abhiyan:

Here we use two-proportion Z-test of to determine whether the proportion of people aware about Swachha Bharat Abhiyan before COVID is different from that of post COVID.

H_0 : The proportion of awareness about Swachha Bharat Abhiyan before and after the pandemic is same. **i.e. ($p_1 = p_2$) against**

H_1 : The proportion of awareness about Swachha Bharat Abhiyan after the pandemic is increased. **i.e. ($p_1 < p_2$)**

Observation Table:

	Before COVID	Post COVID
No. of people aware about Swachha Bharat Abhiyan	259	397
No. of people unaware about Swachha Bharat Abhiyan	204	66
Total	463	463
P	0.559395	0.857451
\hat{p}	0.708423	
\hat{q}	0.291577	
Z- statistic	-9.97817	
p- value	9.5E-24	

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$= -9.9782$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is $-9.9782 <$ tabulated value of Z is -1.6448 . Hence we accept H_0 .

Conclusion: We conclude that, there is a significant difference between the number of people who were aware about Swachha Bharat Abhiyan before COVID and that of after the COVID-19 pandemic. We can say that, COVID-19 brought about awareness among the people about Swachha Bharat Abhiyan.

8. To test whether the proportion of people facing various health problems is increased due to COVID-19 pandemic:

Here we use two-proportion Z-test of to determine whether the proportion of people facing various health problems before COVID is different from that of post COVID.

H_0 : The proportion of people facing various health problems before and after the pandemic is same. i.e. ($p_1 = p_2$) **against**

H_1 : The proportion of people facing various health problems after the pandemic is increased. i.e. ($p_1 < p_2$)

Observation Table:

Sr. No.	Health Problems	Before	During	p_1	p_2	\hat{p}	\hat{q}	Z-stat	Decision
1	BP	149	154	0.321814	0.332613	0.327214	0.672786	-0.3502	accept H_0
2	Diabetes	99	104	0.213823	0.224622	0.219222	0.780778	-0.39715	accept H_0
3	Weight gain	2	48	0.00432	0.103672	0.053996	0.946004	-6.68846	reject H_0
4	Weight loss	0	22	0	0.047516	0.023758	0.976242	-4.74715	reject H_0
5	diff in breathing	0	27	0	0.058315	0.029158	0.970842	-5.2736	reject H_0
6	loss of appetite	0	28	0	0.060475	0.030238	0.969762	-5.37336	reject H_0
7	Weakness	0	13	0	0.028078	0.014039	0.985961	-3.63113	reject H_0
8	joint pain	11	17	0.023758	0.036717	0.030238	0.969762	-1.15144	accept H_0
9	Stress	10	70	0.021598	0.151188	0.086393	0.913607	-7.01821	reject H_0
10	Loneliness	0	36	0	0.077754	0.038877	0.961123	-6.12015	reject H_0

Under H_0 , test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, for health problems such as BP, Diabetes and Joint pain, calculated value of $Z >$ tabulated value of Z (-1.6448). Hence we accept H_0 . For remaining health problems such as weight gain, weight loss, difficulty in breathing, loss of appetite, weakness, stress and loneliness, calculated value of $Z <$ tabulated value of Z (-1.6448). Hence we reject H_0 .

Conclusion: We conclude that, there is no significant difference between the proportion of people suffering health problem such as BP, diabetes and joint pain before COVID and that of after the COVID-19 pandemic as against the people suffering from the remaining health problems.

9. To test whether the proportion of students leaving education is affected by COVID-19 Pandemic :

Here we use test for single population proportion to determine whether the proportion of people students leaving education is affected by COVID-19 Pandemic.

H_0 : The proportion of students leaving education due to COVID is 5%.

i.e. ($p = 0.05$) **against**

H_1 : The proportion of students leaving education due to COVID is less than 5%. i.e. ($p_1 < 0.05$)

Observation Table:

No. of students leaving education	11
No. of students continuing education	605
total number of students	616
\hat{p}	0.017857
Z-stat	-3.6604

Under H_0 , test statistic is,

$$Z_{calc} = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = -3.6604$$

Tabulated value of Z at 5% level of significance = -1.64485

Result: Here, calculated value of Z is $-3.6604 <$ tabulated value of Z (-1.6448). Hence we reject H_0 .

Conclusion: We conclude that, the proportion of students leaving education is less than 5%. We can say that, education is not affected by COVID-19 pandemic.

SUMMARY

From this project, we come to a conclusion that everything is affected by COVID-19 pandemic to some extent or other in Kavathe Mahankal Tehsil, which is the area under consideration of this study.

The following are indicators of Social welfare with impact of COVID-19 pandemic on them:

- i) The income of families was significantly decreased during COVID-19 pandemic, though we observe that, the income of permanent employees remained same.
- ii) There was a decrease in the expenditure of families during COVID-19 pandemic since hotels, cinema halls, functions, etc were all closed during lockdown.
- iii) No new cases of major health problems like BP, diabetes, joint pain were emerged but some other minor health problems were more dominate than ever during COVID-19 pandemic. People were more aware about the health precautions and cleanliness during COVID-19 pandemic.
- iv) COVID-19 pandemic brought about awareness about various Government Social Welfare Schemes.
- v) Electricity consumption was significantly increased during COVID-19 pandemic.

To sum up, we can say COVID-19 pandemic had a strong impact on indicators of Social Welfare. The bright side of this pandemic is that, it prepared people to face any such pandemic in future, if emerges.

BIBLIOGRAPHY

1.	Basic Statistics	B. L. Agarwal	New Age International (2006)
2.	Mathematical Statistics	Gupta & Kapoor	Sultan Chand & Sons (2017)

3. www.wikipedia.com

Questionnaire

Data Collection

A
PROJECT ON

“EFFECT OF ONLINE EDUCATION SYSTEM ON QUALITY OF EDUCATION”



Submitted to

Padmabhushan Vasantodada Patil Mahavidyalaya, KavatheMahankal.
DEPARTMENT OF STATISTICS

By,

Miss. Jagadale Anuradha Subhash
Miss. Ganeshwade Pragati Balaso
Miss. Patil Nikita Nemgonda
Miss. Kadam Pooja Nayaku
Miss. Chougule Pramila Parasu

Under the Guidance of,

Prof. Dr. Patil M. K.

Prof. Koshti V. V.

Prof. Patil A. P.

Prof. Bhosale S. S.

Prof. Satpute G.D.

Through,

The Principal,
Dr.M.K.Patil sir

PADMABHUSHAN VASANTODADA PATIL MAHAVIDYALAYA,
KAVATHEMAHANKAL.

For the academic year 2021-22

CERTIFICATE

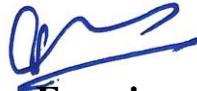
This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for the B. Sc. Part III. They have presently submitted their project entitled “**EFFECT OF ONLINE EDUCATION SYSTEM ON QUALITY OF EDUCATION**”

This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

1. Miss. Jagadale Anuradha Subhash
2. Miss. Ganeshwade Pragati Balaso
3. Miss. Patil Nikita Nemgonda
4. Miss. Kadam Pooja Nayaku
5. Miss. Chougule Pramila Parasu



Teacher in-charge



Examiner



**Dr.M.K.Patil
Head Of Department**

DECLARATION

We hereby declare that the project entitled “**EFFECT OF ONLINE EDUCATION SYSTEM ON QUALITY OF EDUCATION.**”

It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been submitted previously for any other degree of any other purpose.

Sr.no	Roll no	Name	PRN
1	33	Jagadale A.S	2019012426
2	35	Ganeshwade.P.B	2019015797
3	34	Patil N.N	2019012389
4	36	Kadam P.N.	2019015827
5	19	Chougule P.P	2018010762

Place KavatheMahankal

Date :

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude to Dr. Patil M. K. who gave backing and his valuable guidance for this project work. We further thank to Prof. Koshti V.V. Prof. Bhosale S.S. Prof. Patil A.P. and Prof.Satpute G.D Who gave us their encouragement, kind support, and constant inspiration during our project, work further we are thankful to all of our college students who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immerse support during every stage of the project work.

CONTENTS

Sr. No.	Name	Page No.
1	Introduction	6
2	Objective	8
3	Data collection and Sampling Method	9
4	Statistical tools used	10
	4.1) Graphical representation	
	4.2) Paired t-test	
	4.3) Chi Square test for independence of two attributes. (2*2 contingency table)	
	4.4) Z test for population proportion of two samples.	
5	Analysis of Data	12
6	Summary	19
7	Bibliography and Questionnaire	20

INTRODUCTION

OFFLINE EDUCATION SYSTEM



ONLINE EDUCATION SYSTEM



As the first wave of COVID-19 entered India in early 2020, schools were closed in mid-March, just weeks before the end of the 2019-20 academic years.

The vast majority of the country's 26 crore school children have not set foot in a school since then, depending instead on various forms of distance education.

The availability of digital education – whether via live, synchronous teaching on platforms like Zoom, or through recorded lectures, emails, WhatsApp or educational apps - was largely dependent on whether schools, teachers and parents had access to the necessary infrastructure. In many States, teachers came to school and taught in their own empty classrooms, using their blackboards and lab facilities, while

facing a computer screen that communicated the lessons to their students at home.

Our aim is to bring out real effect of online education on quality of education system. And what students/ learners think about online education, what major issues they are facing to educate themselves through online method. We sampled students from various colleges.

We also discussed **Online v/s Offline education** with students, about what challenges they are facing to learn online, because none of them have been through this kind of time .During sampling we found that, it's really hard to learn because of some technical issues, network connection & lack of teacher students interaction.

The intention behind this project is to analyze the grave flexibility of educational system and its infrastructure to tackle unbearable situation like this.

This project will provide all the aspect of students who are being educated by online scheme, what are the ups and downs of online education, comparison of online learning with traditional learning through various options.

Conclusions can be drawn by statistical graphs and various statistical tests.

OBJECTIVES OF STUDY

1. To compare academic performance of students who studied from both online education and offline education.
2. To check advantages and disadvantages of online education system.
3. To check students are satisfied or not with online education system.
4. To analyse effect of online education on students progress.
- 5 . To detect popular platform preferred for online teaching
6. To measure effect of online teaching on rural and urban students.
7. To find online exams are conducted strictly or not and are there lots of chances for cheating or not.

SOURCE OF DATA & SAMPLING

METHOD

The main objective of this project is, “To find how online education affect the quality of education”. Our area under consideration for thr whole project is Miraj city and respective villages. We have sampled students from different educational phases, to ensure that conclusions apply to various situations. Also this project is intended to know about the current methods that are students are using for online teaching-learning.

We designed a questionnaire to fulfill the objectives behind this project accurately. The data collection procedure was done using Google forms and by one-to-one approach.

Here we use convenient sampling method,which is easy and accessible for data collection.We sampled 409 students from various villages of Miraj tehsil.

We sampled subjects using simple random sampling without replacement (SRSWOR) method. The collection of sampling has been taken through online questionnaire using Google forms and direct interviewing with students.

STATISTICAL TOOLS

Following are some statistical tools us towards the appropriate results and finding.

• **Graphical Representation**

To represent our data graphically, we use the multiple bar diagram, pie chart method etc. To construct the appropriable divided bar diagram of given data we used Microsoft Excel.

- i. **Pie Chart:** A pie chart displays data, information and statistics in an easy-to-read 'pie-slice' format with varying slice sizes telling you how much of on data element exists. The bigger the slice the more of that particular data was gathered.
- ii. **Bar Chart:** A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

• **Paired t-test:-**

To test the difference between the online and offline marks, we use Paired-t test. Here data is in the form of online and offline marks.

Where, X_i =values of the online marks and

Y_i = values of the offline marks. ($d_i = Y_i - X_i$)

H_0 : The difference between the means is insignificant.

V_s

H1: The difference between the means is significant.

Under Ho the test statistics is

$$t = \bar{d} / (sd / \sqrt{n})$$

Statistic t has (n-1)d.f

The value of t is obtained with the help of formula.. To make decision about Ho, Calculated value of t is compared with table value of t for (n-1) d. fat 5 % level of significance (1.0.s)

Conclusion:-

If cal. $t > \text{tab. } t$, then we reject Ho at 5 % 1. o. s. Otherwise we accept Ho.

• ***Test for population proportion***

To test the population proportion of two samples.

Hypothesis:

H₀: Two population proportions are same. ($p_1 = p_2$)

Vs

H₁: Two population proportions are not same. ($p_1 \neq p_2$)

Under H₀, the test statistic is given by,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Let α be the level of significance and Z_α being the critical value.

Conclusion:

If calculated $|Z| >$ tabulated Z , we reject H_0 at α % l.o.s otherwise we accept it.

•Chi square test using 2*2 contingency table:-

For any attributes A & B having two classes on each test statistic for testing independence of A & B is given as

$$x^2_{cal} = \frac{N(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)} \sim x^2(1)$$

If A & B are any two attributes having two classes on each then 2*2 contingency table for observed values on attributes A & B is given as,

	B1	B2	Total
A1	a	b	a+b
A2	c	d	c+d
Total	a+c	b+d	N

Here, observed cell frequencies a,b,c,d with $N=a+b+c+d$ be the total class frequency.

We want to test two attributes A & B are independent to each other.

i.e H_0 : A & B are independent to each other

Vs

H_1 : A& B are dependent on each other

Test statistic is,

$$x^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)} \sim x^2(1)$$

Result:-

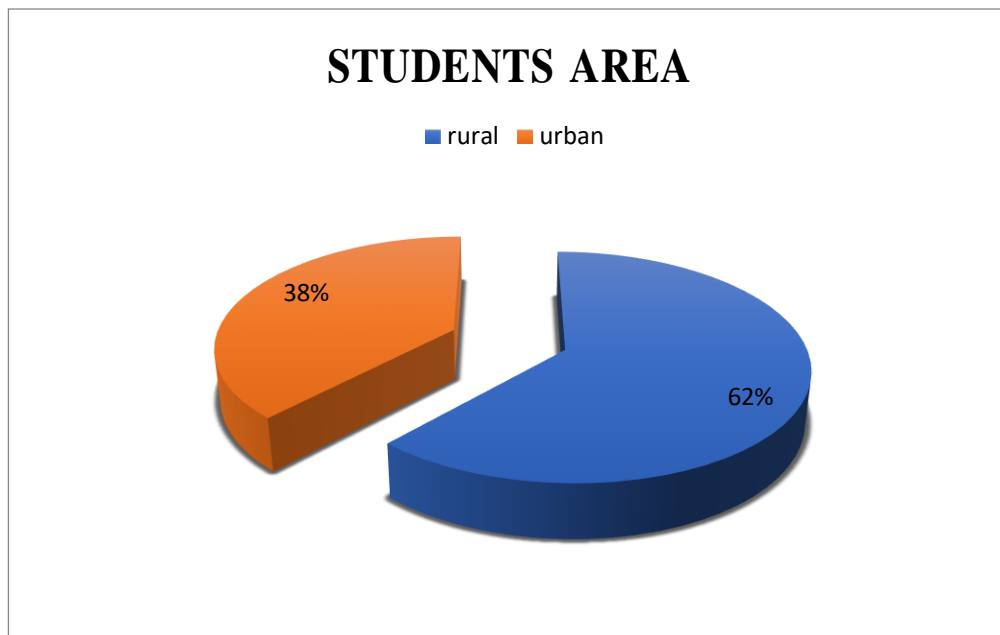
Reject H_0 if $x^2_{cal} > x^2(1)$,a otherwise accept H_0 .

i.e two attributes A & B are independent on each other.

STATISTICAL ANALYSIS

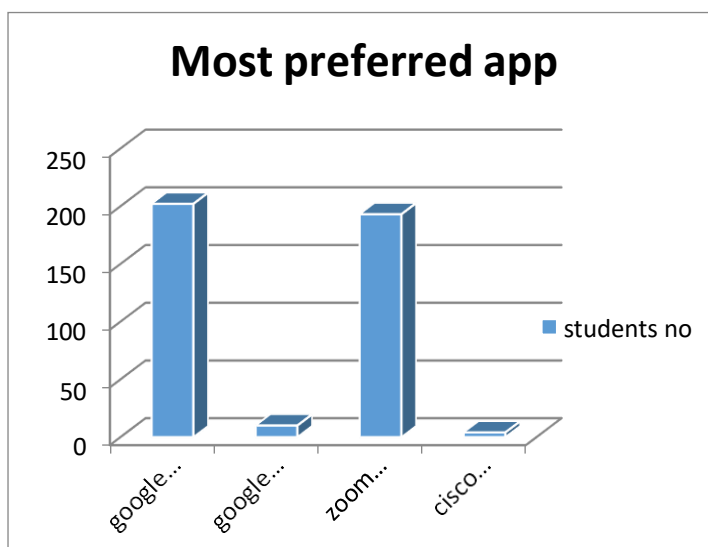
1. TO GET THE BASIC IDEA ABOUT CURRENT ONLINE INFRASTRUCTURE THROUGH VARIOUS STATISTICAL GRAPHS.

I] FROM WHICH AREA THE STUDENTS BELONG?



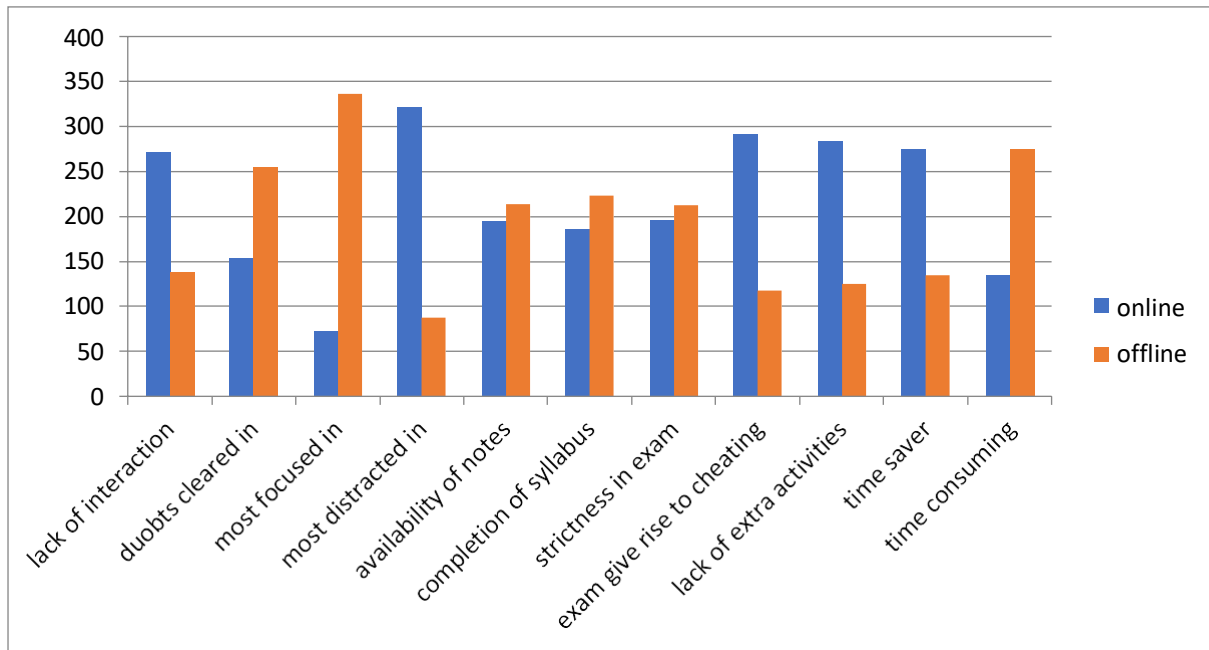
Conclusion:- According to data, most of the students belongs from rural area .

II) MOST PREFERRED PLATFORM DURING ONLINE LEARNING:-



Conclusion:- From above graph we will conclude that, zoom meeting & Google meet are most preferred platform for online teaching.

III) TO CHECK ADVANTAGES AND DISADVANTAGES OF ONLINE EDUCATION.

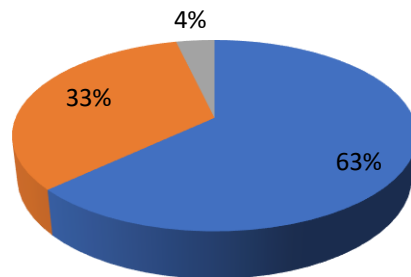


Conclusion:- From above graph we concluded that, there is lack of teacher Students interaction, no doubts cleared, most distracted mode, less availability of notes, incomplete syllabus, less strictness in exam which promotes cheating and lack of extra curricular activities are the disadvantages of online education mode But it is time saver which is only one advantage of online education

IV. STUDENTS CHOICE WHEN THEY GET BORED DURING ONLINE LECTURES.

STUDENT'S CHOICE WHEN GET BORED.

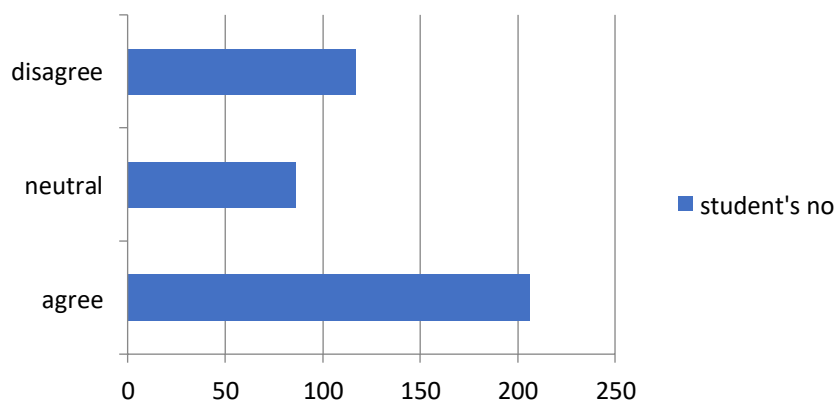
■ social media ■ sports fun ■ T.V.



Conclusion: According to above pie chart, we concluded that most of the students use social media ,when they get bored during online lectures.

V. TO CHECK ONLINE EXAM PROMOTES CHEATING OR NOT.

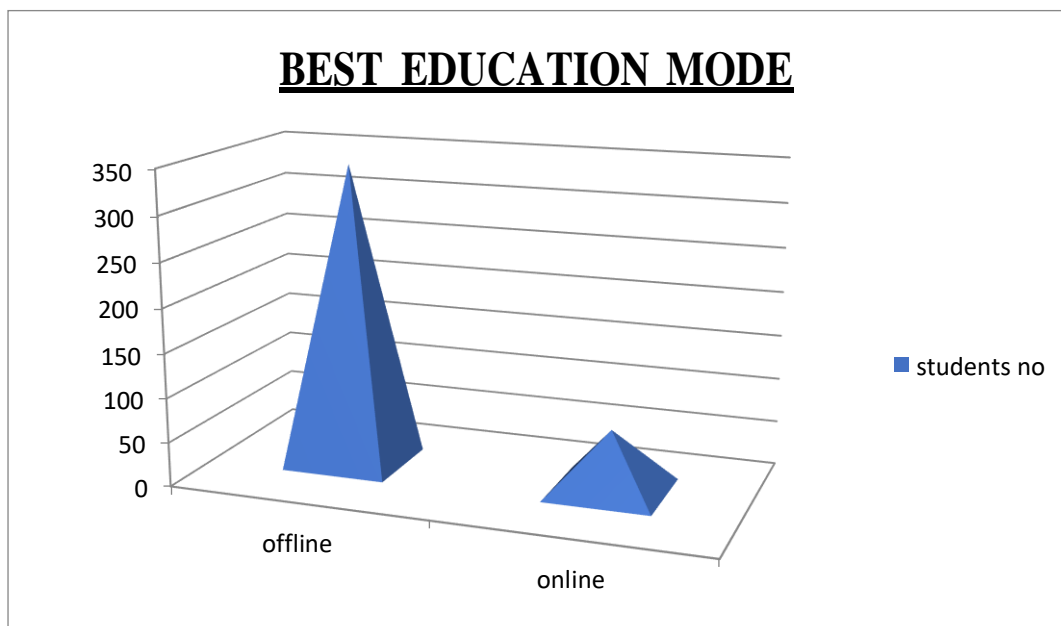
Online exam promotes cheating





Conclusion: From above graph it is clear that, according to most of students online exam mode promotes cheating.

VI) ACCORDING TO STUDENTS WHICH IS BEST EDUCATION MODE?



Conclusion:- From above graph it is conclude that , offline mode is best education mode than online mode.

2. TO TEST WHETHER THE MARKS ARE INCREASED OR NOT DURING ONLINE EXAMS AS COMPARED TO OFFLINE.

Here ,we use paired t- test of two samples for mean to test whether the mean Difference Between online marks and offline marks is zero or not.

Ho: There is no significant difference between online and offline marks.

i.e. $\mu_x = \mu_y$

H1: There is significant difference between online and offline marks
i.e $\mu_x > \mu_y$

<i>Column1</i>	<i>70</i>	<i>60</i>
Mean	80.8488	70.60611
Variance	80.26043	111.081
Observations	409	409
Pearson Correlation	0.537562	
Hypothesized Mean Difference	0	
df	408	
t Stat	21.85612	
P(T<=t) one-tail	5.72E-71	
P(T<=t) two-tail	1.14E-70	

Under Ho, the test statistics is,

$$t = \frac{\bar{d}}{(Sd / \sqrt{n})}$$

$$= 21.8561$$

Tabulated value of t₄₀₈ at 5% level of significance = 1.6481

Result: Here, calculated value of t is 21.8561 > tab value of t is 1.6481, hence we reject Ho.

Conclusion :- there is significant difference between online and offline marks

i. e. Marks significantly increases during online exam.

3. TO STUDY THE DEPENDENCE OF MODE OF EDUCATION AND FOCUS AND DISTRACTION OF STUDENTS .

Here we use chi square test for independence of attributes to test hypothesis . to check whether the modes of education and focus and distraction of students is dependent or not

Ho: Student's focus and distraction are independent upon mode of education.

Vs

H1: Student's focus and distraction are dependent upon mode of education.

Column1	Focused	Distracted	Total
Online	73	321	394
Offline	336	88	424
Total	409	409	818

Test Statistics is,

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(a+c)(c+d)(b+d)} \sim$$

$$\chi^2(1)$$

$$=301.1581$$

tab value of chi square at 5% level of significance =3.841

Result:

Here ,calculated value of chi square is $301.1581 >$ tabulated value of chi square is 3.841.

Hence we reject H_0 .

Conclusion:

Students focus and distraction are dependent upon mode of education and also concludes students are more distracted in online education.

4. TO STUDY THE PREFERENCE OF OFFLINE MODE BY RURAL AND URBAN STUDENTS

Here we want to check, what proportion of preference of offline mode by rural and urban students.

This is area wise comparison, according to their preference

H_0 : There is no significant difference between preference of offline mode by rural and urban students.

. vs

H_1 : There is significant difference between preference of offline mode by rural and urban students.

Observation Table:

	Best education system		
	Online	Offline	Total
Rural	43	210	253
Urban	26	130	156

Under H_0 Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P1^{\wedge} = 210/253 = 0.8300$$

$$P2^{\wedge} = 130/156 = 0.8333$$

$$p^{\wedge} = 0.8312$$

$$q^{\wedge} = 0.1688$$

$$z_{cal.} = -0.02816$$

Tabulated value of Z at 5% level of significance = 1.64485

Result: Here, Calculated value of Z is -0.02816 < tabulated value of Z is 1.6448. Hence we accept H_0 .

Conclusion: There is no significant difference between preference of

offline mode by rural and urban students.

& All students from rural and urban prefer offline mode.

5. TO STUDY THE PREFERENCE OF ONLINE MODE BY RURAL AND URBAN STUDENTS

Here we want to check, what proportion of preference of online mode by rural and urban students.

phase prefer which education method. This is area wise comparison, according to their preference .

H_0 : There is no significant difference between preference of online mode by rural and urban students.

vs

H1: There is significant difference between preference of online mode by rural and urban students.

Observation Table:

Column1	Online	Offline	Total
Rural	43	210	253
Urban	26	130	156

under Ho Test statistics is

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$P1^{\wedge}=43/253=0.1699$$

$$P2^{\wedge}=26/156=0.1666$$

$$p^{\wedge}=0.1687$$

$$q^{\wedge}=0.8313$$

$$z_{cal.}=0.127242$$

Tabulated value of Z at 5% level of significance = 1.64485

Result: Here, Calculated value of Z is 0.127242 < tabulated value of z is 1.6448. Hence we accept H0.

Conclusion: There is no significant difference between preference of online mode by rural and urban students.

All students from rural and urban gives less preference for online mode.

From both cases discussed above we concluded that,According to both rural and urban students the proportion of offline mode as best eduaction system is greater .

SUMMARY

Overall conclusion can be drawn from this project is, there is bad effect of online education on quality of education .

Reasons are as follows:

- i) Lack of interaction and all doubts are not cleared.
- ii) More distractions, hard to get focused.
- iii) Availability of proper notes is less and whole syllabus not completed.
- iv) No strictness in the online exam which gives rise to cheating.
- v) As it promotes cheating during exam online marks are more than offline marks.
- vi) Students make more use of social media during online lectures.
- vii) From both rural and urban area most of the students chose offline.

Whilst, online education could be game changer if implemented properly, as now it is in its early stages. It is time saver education system ,there are chances to learn google concepts, vast freedom to learn ,development of skill.

Sampled data suggest that, above conclusions are strongly valid to the population.

LIMITATIONS

1. Sample size is very small to represent the whole city .
2. We have to limit the scope of this research according to the result we obtain from the questionnaires.
3. This survey is limited Miraj tehsil in Maharashtra.
4. There is time limit for collecting data which is not sufficient.

SCOPE

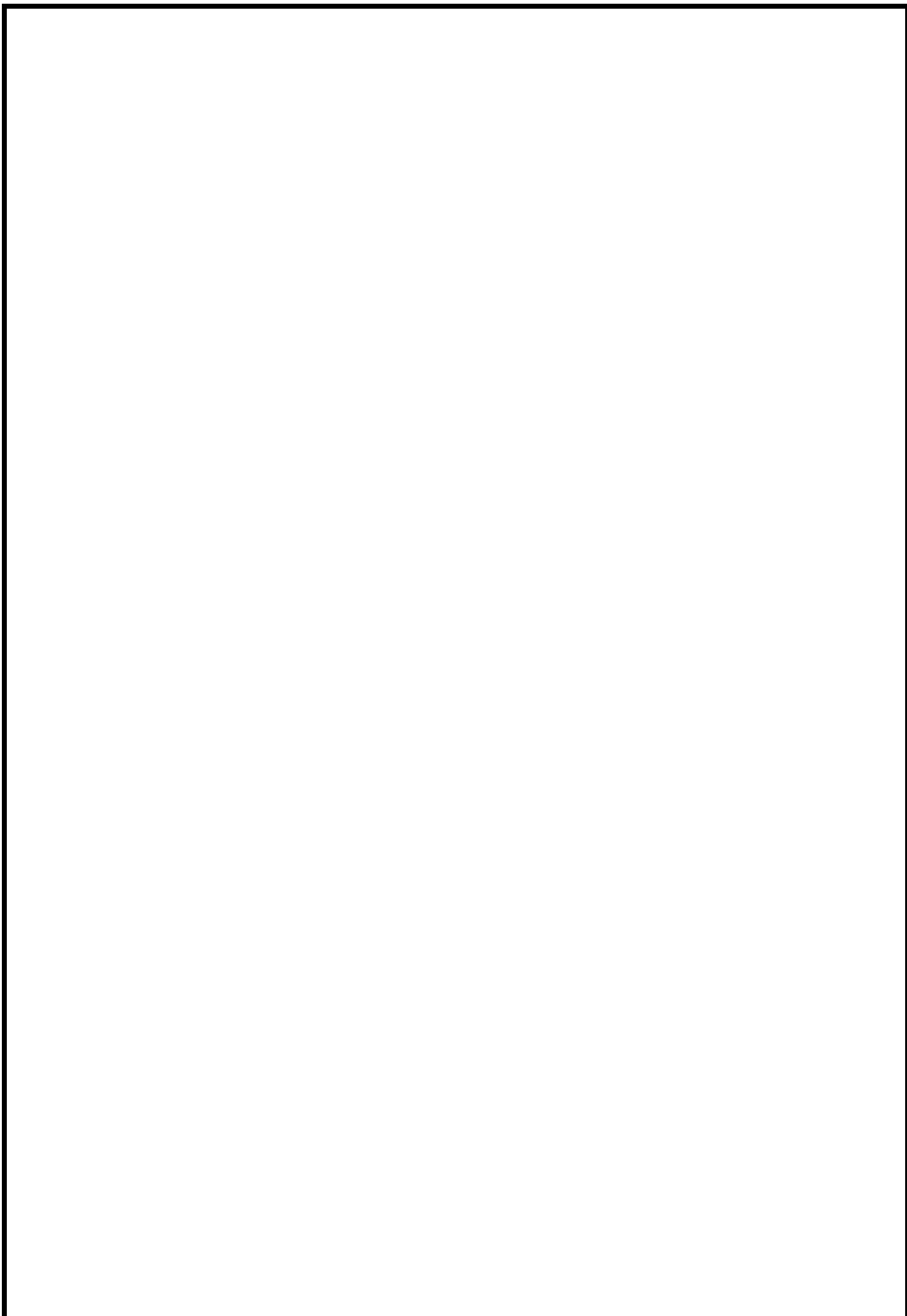
1. This project can be extended to the international level.
2. It can be extended to the greater region for better estimates.

BIBLIOGRAPHY

1. Basic Statistics, B. L. Agarwal, *New Age International* (2006)
2. Mathematical Statistics, Gupta & Kapoor, *Sultan Chand & Sons*
(2017)
3. www.pib.gov.in.
4. www.udiseplus.gov.in
5. www.mospi.nic.in

QUESTIONNAIRE

https://docs.google.com/forms/d/16XFeh7IQaXN-6_MC4Yrfb_42JdI-r2Y4W3_fduUmc34/edit?usp=drive



A

Project Report

On

“ROLE OF MEDIA ON HUMAN DEVELOPMENT”

Submitted to

Padmabhushan Vasantrodada Patil Mahavidyalaya,

Kavathe Mahankal.

DEPARTMENT OF STATISTICS

By,

Miss. Suryawanshi Arati Jaysing

Miss. Pawar Mahadevi Machhindra

Miss. Shinde Nisha Sanjay

Miss. Savale Prajakta Popat

Miss. Gidde Ashawini Dilip

Under the Guidance of,

Prof. Dr. Patil M. K.

Prof. Koshti V. V.

Prof. Patil A. P.

Prof. Bhosale S. S.

Prof. Satpute G.D.

Through,

The Principal,

Padmabhushan Vasantrodada Patil

Mahavidyalaya, Kavathe Mahankal.

For the academic year 2021-2022

**Padmabhushan Vasantrodada Patil Mahavidyalaya,
Kavathe Mahankal**

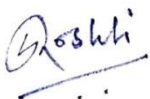
DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for the B. Sc. Part III. They have presently submitted their project entitled "ROLE OF MEDIA ON HUMAN DEVELOPMENT"

This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

1. Miss. Suryawanshi Arati Jaysing
2. Miss. Pawar Mahadevi Machhindra
3. Miss. Shinde Nisha Sanjay
4. Miss. Savale Prajakta Popat
5. Miss. Gidde Ashwini Dilip



Teacher in charge



Examiner



Dr. M. K. Patil

Head of Department

DECLARATION

We hereby declare that the project entitled "ROLE OF MEDIA ON HUMAN DEVELOPMENT".

It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been submitted previously for any other degree of any other purpose.

Sr. No.	Name
1.	Suryawanshi Arati Jayshing
2.	Pawar Mahadevi Machhindra
3.	Shinde Nisha Sanjay
4.	Savale Prajakta Popat
5.	Gidde Ashwini Dilip

Place : Kavathe Mahankal

Date:

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude to Prof. Dr. Patil M.K. who gave backing and his valuable guidance for this project work. We further thank Prof. Koshti V.V., Prof. Bhosale S.S., Prof. Patil A.P., and Prof. Satpute G.D. who gave us their encouragement, kindly support, and constant inspiration during our project. We are thankful to all the people in our taluka who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.

CONTENTS

Sr. No.	Name	Page No.
1	Introduction	6,7
2	Objective	8
3	Data collection and Sampling Method	9,10
4	Statistical tools used	
4.1	Graphical representation	11
4.2	Z test for population proportion of two samples	12
4.3	Chi-Square test for independence of attributes	13
4.4	Paired T test	14
5	Analysis of Data	15-26
6	Limitation and scope	27
7	Summary	28
8	Bibliography	29
9	Questionnaire	30,31

INTRODUCTION

Media is used in our development; there are many types of media. Some are the most popular mass media, print media, internet etc.

Media plays an important role in our day-to-day life. It plays a very pivotal role in our lives. Media plays the most crucial role in student life or the life of the current generation by getting them prepared and updated about their near future. Media helps increase our knowledge; it helps to get in touch with the entire world.

Peoples are using various types of media like TV, newspaper, Social media, magazine, radio etc.

1) Newspaper:

Newspaper is a mode of mass communication; it is very helpful in creating social awareness. Newspaper provides news about a country's economic situation, sports, entertainment, trade and commerce.

2) Social media:

Social media is the most updated version of media. Social media is organized as a way to interact with family and friends. Social media is an interesting based form of communication. There are many forms of social media. Some of the most popular social media websites with over a hundred million registered users include WhatsApp, Facebook, Telegram, YouTube, Hotstar, etc.

3) Television (TV):

Television is a source of information or communication and it plays a very significant role in everyone's life. Educational programming can develop young children's socialization and learning skills. News, current events and historical programming can help many young people more aware of other cultures and people.

We study effect of media in human development and which media more prefer human the motive of project to study not only the use of media but also that the persons are satisfied with these information also study the role of media during covid-19 period.

We have to study that which type of media is referred on large amount and it is useful for their mental growth also study the which media people want to see more.

OBJECTIVES OF STUDY

- What kind of news do people want to see more?
- To check which kind of news referred on large amount.
- To check which media that express the right kind of the information to you.
- To check the importance of media in Education.
- To check what effect does the news have on our minds.
- Role of media during covid-19 period according to you.
- The importance of media in our life.
- To study media influence our life.

Source of Data & Sampling Method

The main objective of this project is, “Role of media on human development”. Our area under consideration for the whole project is Kavathe Mahankal city and respective villages.

We designed a questionnaire to fulfill the objectives behind this project accurately. The data collection procedure was done using Google forms and by one-to-one approach.

Here we use convenient sampling method, which is easy and accessible for data collection. We sampled 403 people from various villages of Kavathe mahankal tahsil.

We sampled subjects using simple random sampling without replacement (SRSWOR) method. The collection of sampling has been taken through online questionnaire using Google forms and direct interviewing with people.

- **Sample size Determination:**

The sample size is determined by the Yamane's (1967) formula. The formula is given as,

$$n = \frac{N}{(1 + N * e^2)}$$

Where,

n = Sample size

N = Population size

e = Level of precision

Here,

The total Population of Kavthe Mahankal = N = 152327 Then our sample size becomes,

$$\begin{aligned} n &= \frac{N}{(1+N*e^2)} \\ &= \frac{152327}{(1+152327*0.05^2)} = 398.89 \approx 399 \end{aligned}$$

From above calculation we conclude that approximate sample size is 399 but we have taken 403 sample sizes.

STATISTICAL TOOLS

Following are some statistical tools used towards the appropriate results and finding.

• **Graphical Representation**

To represent our data graphically, we use the multiple bar diagram, pie chart method etc. To construct the appropriate divided bar diagram of given data we used Microsoft Excel.

I) **Pie Chart:** A pie chart displays data, information and statistics in an easy to-read 'pie-slice' format with varying slice sizes telling you how much of on data element exists. The bigger the slice the more of that particular data was gathered.

II) **Bar Chart:** A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

● Test for population proportion

To test the population proportion of two samples.

Hypothesis:

H_0 : Two population proportions are same. ($p_1 = p_2$)

V/s

H_1 : Two population proportions are not same. ($p_1 \neq p_2$)

Under H_0 , the test statistic is given by,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Let α be the level of significance and Z_α being the critical value.

Conclusion:

If calculated $|Z| >$ tabulated Z , we reject H_0 at α % l.o.s. otherwise, we accept it.

• Chi-Square test for independence of attributes:

The hypothesis that, the one factor independent of other or not, can be tested by the Chi-Square test.

The hypothesis that,

H_0 : Two attributes are independent of each other.

V/s

H_1 : Two attributes are dependent.

Under H_0 , the test statistic is,

$$\chi^2 = \sum \sum (O_{ij}^2 / E_{ij}) - N \sim \chi^2(m-1)(n-1) \text{d.f.}$$

Statistic χ^2 has $(m-1)(n-1)$ d.f in formula (a), E_{ij} is the expected frequency

Corresponding to (i,j) th cell of observed frequency O_{ij} .

Under the null hypothesis, the expected frequency is,

$$\begin{aligned} E_{ij} &= (\text{ith row total} * \text{j th column total}) / \text{sample size} \\ &= (R_i * C_j) / n \end{aligned}$$

Once the expected frequencies are calculated, the value of Chi-Square is obtained with the help of formula (a). To make a decision about H_0 , calculated value of Chi-Square is compared with the table attributes are not independent. Otherwise accept H_0

• Paired t-test

To test the difference between the set of before and after covid-19 media use in day-to-day life, we use paired t-test

Here data is in the form of before and after type

Where, X_i = value of media use before covid-19

Y_i = value of media use after covid-19 ($d_i = Y_i - X_i$)

H_0 : The difference between the means is insignificant.

V/s

H_1 : The difference between the means is significant.

Under H_0 the test statistics is

$$t = \frac{\bar{d}}{(sd/\sqrt{n})}$$

Statistic t has (n-1)d.f.

The value of t is obtained with the help of formula. To make decision about H_0 , Calculated value of t is compared with table value of t for (n-1) d. fat 5 % level of significance (l.o.s)

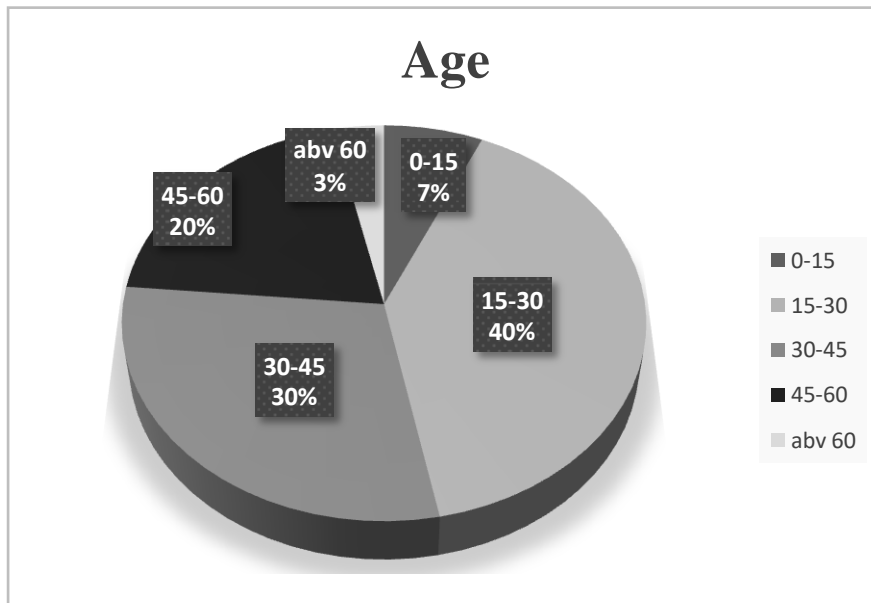
Conclusion:-

If cal. t > tab. t, then we reject H_0 at 5 % l. o. s. Otherwise we accept H_0 .

Data Analysis:

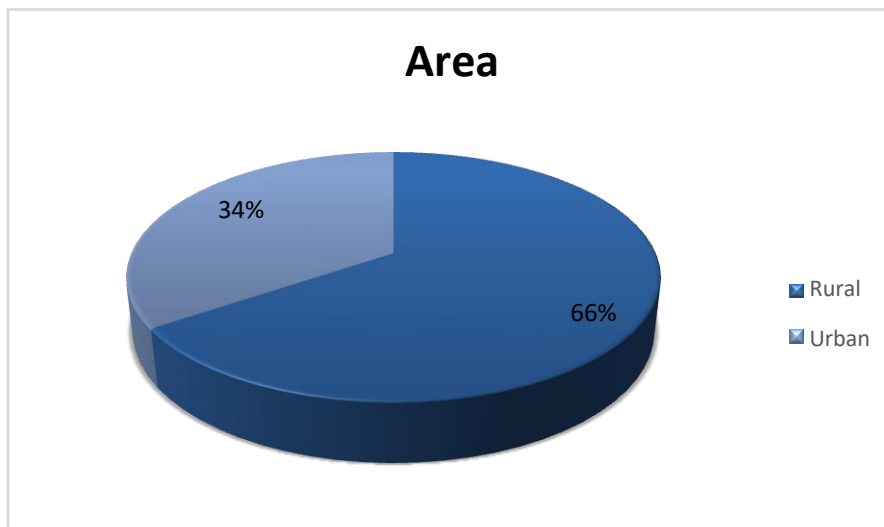
Distribution of respondents according to various aspect:

1. Age group wise representation:



Conclusion: From the above graph we conclude that 15-30 and 30-45 age groups give more response to that of other age groups.

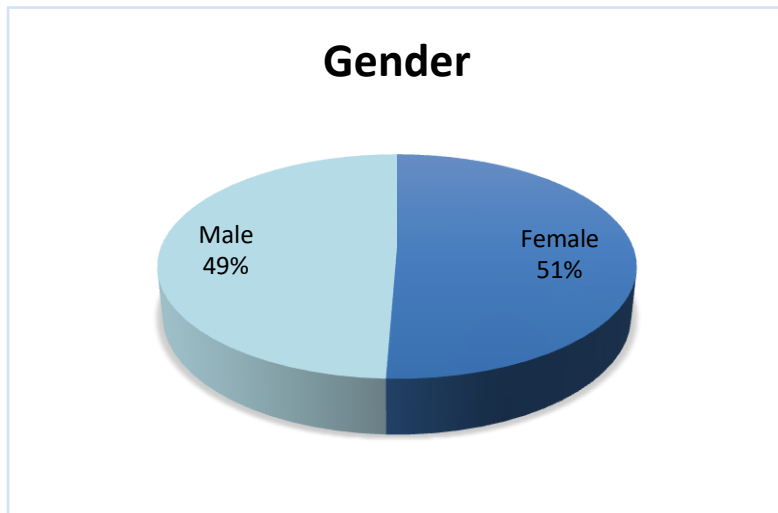
2. Area wise response of Urban and Rural



Conclusion: we conclude that rural area gives more response.

3. Gender wise distribution of male and female

Female	Male	Total
204	199	403



Conclusion: from above graph we conclude that approximately same responses given by male and female.

❖ **Objective:** To test the equality of proportion of female in urban and rural area.

Table gives the proportion of females:

Total Rural	Total Urban	female in R	female in U	p1	p2
238	165	125	79	0.52521	0.478788

Calculation:

H0: The proportions of female in rural and urban area are equal.

V/s: H1: The proportion of female in rural and urban area are not equal.

$$P = \frac{N1P1 + N2P2}{N1 + N2}$$

$$P = \frac{238 * 0.5252 + 165 * 0.4787}{238 + 165}$$

$$P = 0.5062$$

$$Q = 1 - P$$

$$Q = 1 - 0.5062$$

$$Q = 0.4938$$

Test statistics:

$$Z = \frac{P1 - P2}{\sqrt{PQ \left(\frac{1}{N1} + \frac{1}{N2} \right)}}$$

$$= \frac{0.52521 - 0.478788}{\sqrt{0.5062 * 0.4938 * \left(\frac{1}{238} + \frac{1}{165} \right)}}$$

$$Z = 0.9165$$

Here cal. |Z| = 0.9165 and the table value at 5% l.o.s is Z = 1.96

Conclusion:

We can accept the H0 at 5% of l.o.s. thus the proportion of female in rural and urban area are equal.

❖ **Objective:** To test the equality of proportion of male in urban and rural area.

Table gives the proportion of females:

Total Rural	Total Urban	male in R	male in U	p1	p2
238	165	113	86	0.47479	0.5212121

Calculation:

Hypothesis:

H_0 : The proportions of male in rural and urban area are equal.

V/s: H_1 : The proportion of male in rural and urban area are not equal.

$$P = \frac{N_1P_1 + N_2P_2}{N_1 + N_2}$$

$$P = \frac{238 * 0.4747 + 165 * 0.5212}{238 + 165}$$

$$= 0.4937$$

$$Q = 1 - P$$

$$Q = 1 - 0.4937$$

$$= 0.5062$$

Test statistics:

$$Z = \frac{P_1 - P_2}{\sqrt{PQ \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}$$

$$= \frac{0.4747 - 0.5212}{\sqrt{0.4937 * 0.5062 * \left(\frac{1}{238} + \frac{1}{165} \right)}}$$

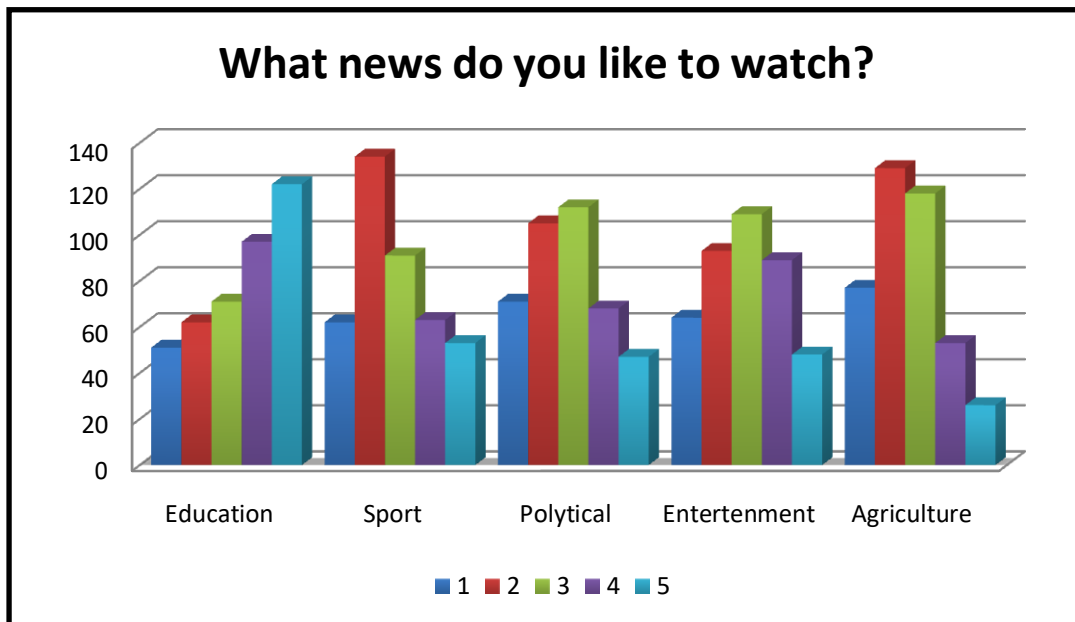
$$Z = -0.9165$$

Here cal. $|Z| = 0.9165$ and the table value at 5% l.o.s is $Z = 1.96$

Conclusion:

We can accept the H_0 at 5% of l.o.s. thus the proportion of male in rural and urban area are equal.

1) ***Objective:** What kind of news do people want to see more



Conclusion: we conclude that educational news people see more

2) ***Objective:** Which media that express the right kind of the information to you

❖ **Objective:** To check the dependency between Age and TV use.

Age/TV	Yes	No	Total
0-15	15	4	19
15 -30	45	9	54
30-45	44	14	58
45-60	40	14	54
above 60	8	2	10
Total	152	43	195

Calculation:

H_0 : to check Age and use of TV are independent.

v/s H_1 : to check Age and use of TV are dependent.

Now let E_{ij} expected frequency of the observation in the observation in i^{th} and j^{th} column. Thus we have

$$E_{11}=14.8103, E_{12}=4.1897, E_{21}=42.0923, E_{22}=11.9077$$

$$E_{31}= 45.2103, E_{32}=12.7897, E_{41}=42.0923, E_{42}=11.9077$$

$$E_{51}=7.7948, E_{52}=2.2051$$

The chi-square statistic is $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

$$\chi^2=1.5649$$

Table value of $\chi^2 = 9.4877$

Here the calculated value of chi-square is less than the tabulated chi-square.

Conclusion:

Here the calculated value of chi-square is less than tabular value then we accept the null hypothesis.

Hence we conclude that the age of people and use of TV is independent to each other.

❖ **Objective:** To check the dependency between Age and Social media use.

Age/Social media	Social Media		Total
	Yes	No	
Age			
0-15	6	1	7
15 -30	74	17	91
30-45	26	13	39
45-60	12	4	16
above 60	1	0	1
Total	119	35	154

Calculation:

H_0 : to check Age and use of Social media are independent.

V/sH_1 : to check Age and use of Social media are dependent.

Now let E_{ij} expected frequency of the observation in the observation in i^{th} and j^{th} column. Thus we have

$E_{11}=5.4090$, $E_{12}=1.5909$, $E_{21}=70.3118$, $E_{22}= 20.6818$,

$E_{31}= 30.1363$, $E_{32}=8.8636$, $E_{41}=12.3636$, $E_{42}=3.6363$,

$E_{51}=0.7727$, $E_{52}=0.2272$

The chi-square statistic is $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

$$\chi^2 = 3.9714$$

Table value of $\chi^2 = 9.4877$

Here the calculated value of chi-square is less than the tabulated chi-square.

Conclusion:

Here the calculated value of chi-square is less than tabular value then we accept the null hypothesis.

Hence we conclude that the age of people and use of Social media is independent to each other.

❖ **Objective:** To check the dependency between Age and Newspaper use.

Age/newspaper	Newspaper		Total
Age	Yes	No	
0-15	1	0	1
15 -30	15	2	17
30-45	17	5	22
45-60	8	3	11
above 60	3	0	3
Total	44	10	54

Calculation:

H_0 : to check Age and use of Newspaper are independent.

V/s H_1 : to check Age and use of Newspaper are dependent.

Now let E_{ij} expected frequency of the observation in the observation in i^{th} and j^{th} column. Thus we have

$E_{11}=0.8148$, $E_{12}=0.1851$, $E_{21}=13.8518$, $E_{22}= 3.1481$,

$E_{31}= 17.9259$, $E_{32}=4.0740$, $E_{41}=8.9629$, $E_{42}=2.0370$,

$E_{51}=2.4444$, $E_{52}=0.5555$

The chi-square statistic is $\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$

$$\chi^2 = 2.2399$$

Table value of $\chi^2 = 9.4877$

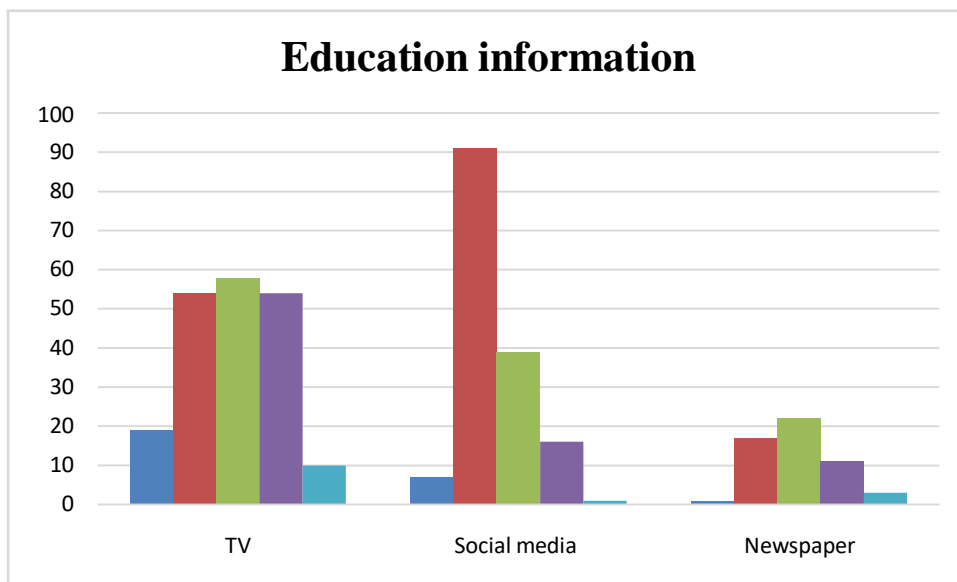
Here the calculated value of chi-square is less than the tabulated chi-square.

Conclusion:

Here the calculated value of chi-square is less than tabular value then we accept the null hypothesis.

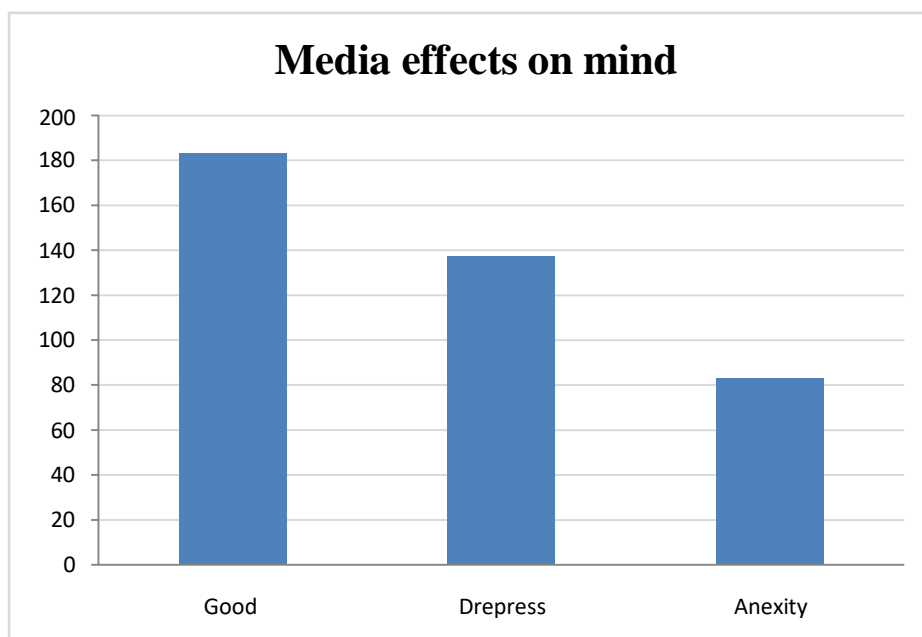
Hence we conclude that the age of people and use of Newspaper is independent to each other.

3) ***Objective:** Which media get most of the education information and are you satisfied



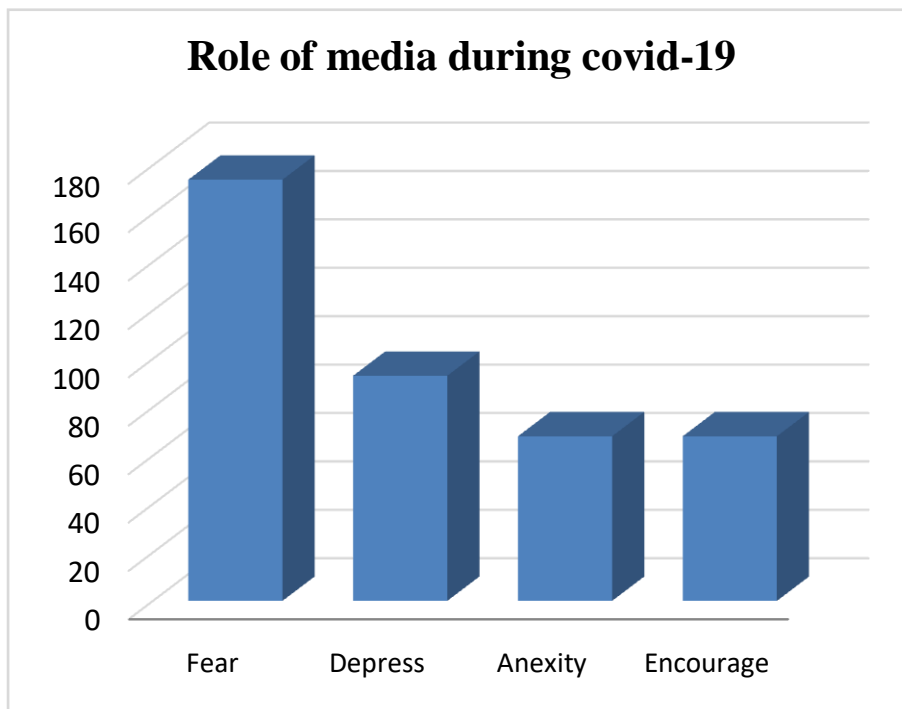
Conclusion: we conclude that the social media gives more educational information to peoples.

4) ***Objective:** What effect does the news have on own mind



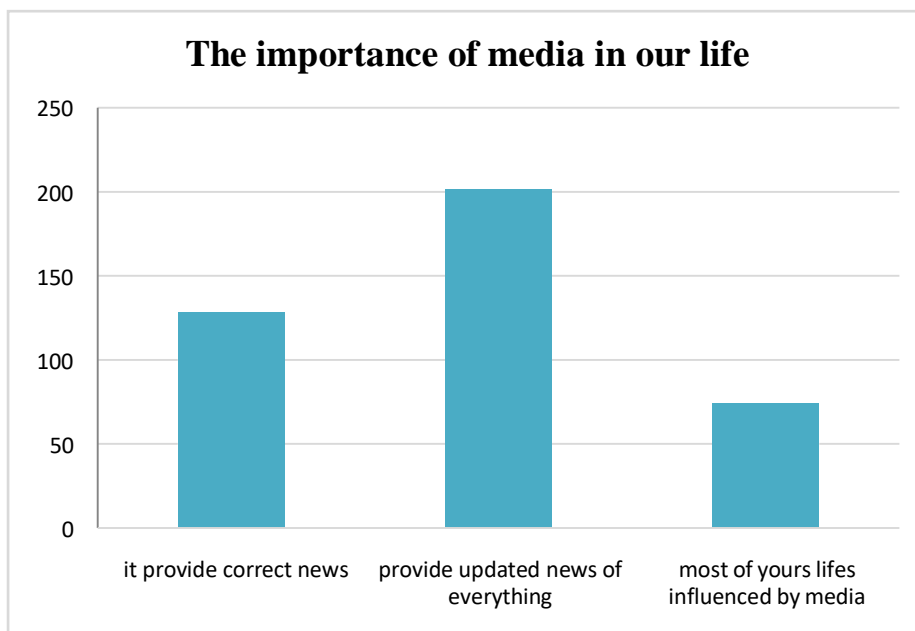
Conclusion: we conclude that effect of media is good on mind.

5) ***Objective:** Role of media during covid-19 period according to you



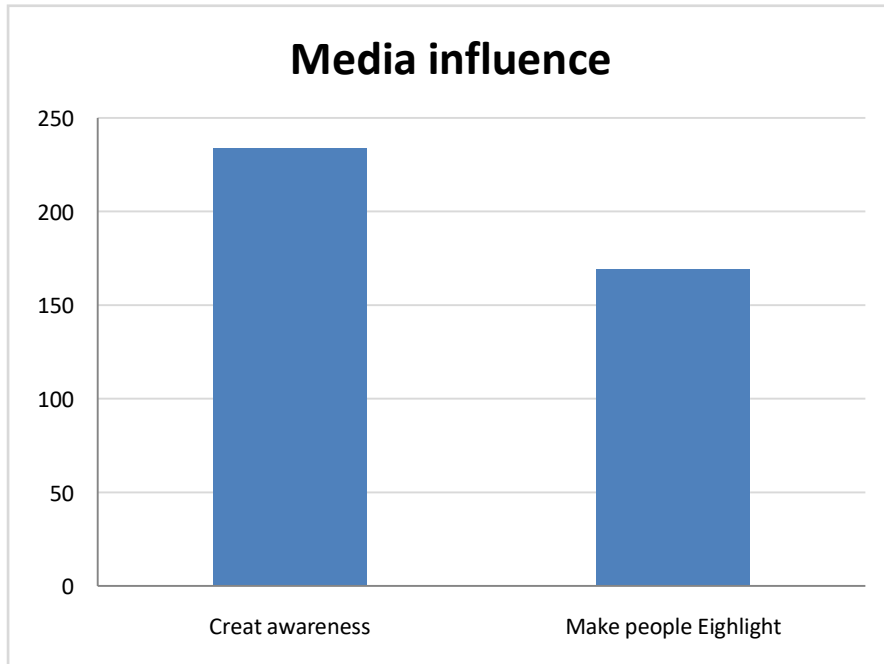
Conclusion: In Covid-19 pandemic news media became the reason of fear about virus among people.

6) ***Objective:** The importance of media in our life



Conclusion: In above graph we conclude that, according to most of the people media provide updated news of everything.

7) ***Objective**: How do media influence our life?



Conclusion: In above graph we conclude that, according to people media play an important role to create awareness.

8) ***Objective**: To compare use of media in before covid-19 and after covid-19

❖ **Objective**: use of media in before covid-19 and after covid-19

Here we use paired t-test of two samples for mean to test whether the difference between use of media before covid-19 and after covid-19

Hypothesis:

H_0 : there is no significant difference between using media before covid-19 and after covid-19

V/s

H_1 : there is significant difference between using media before covid-19 and after covid-19

i.e. $\mu_x = \mu_y$

V/s

i.e. $\mu_x > \mu_y$

Observation table:

t-Test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	35.80645	44.88834
Variance	248.2908	469.5771
Observations	403	403
Pearson Correlation	-0.02104	
Hypothesized Mean Difference	0	
d.f.	402	
t Stat	-6.73755	
P(T<=t) one-tail	2.8E-11	
t Critical one-tail	1.648653	
P(T<=t) two-tail	5.6E-11	
t Critical two-tail	1.965883	

Under H_0 , the test statistics is,

$$t = \bar{d} / (sd / \sqrt{n})$$

$$= -6.7375$$

Tabulated value of t_{402} at 5% level of significance = 1.6486

Result: Here calculated value of t is $-6.7375 >$ tab value of t is 1.6486, hence we reject H_0 .

Conclusion: There is significant difference between use of media before and after covid-19.

i.e. use of media significantly increases after covid-19.

LIMITATION

- 1) We have to limit the scope of this research according to the result we obtain from the questionnaires.
- 2) This survey is limited to Kavathe Mahankal.

SCOPE FOR FUTURE

- 1) This project can be extended to the international level.
- 2) We will-planned Study can be taken including some relevant Characteristics.
- 3) It can be extended to the greater region for better estimates.

Summary

Overall conclusion can be drawn from this project is media plays an important role in human development all around the world.

- 1) In this project people gives more preference to educational news than other news.
- 2) Social media can provide most educational information to people.
- 3) Trough media individuals get a platform to exercise their talents and improve themselves.
- 4) According to people all kind of media provide right kind of information.
- 5) Media gives good effect on their mind.
- 6) According to most of the people media provide updated news of everything
- 7) In Covid-19 pandemic news media became the reason of fear about virus among people.

BIBLIOGRAPHY

Basic Statistics. B. L. Agarwal

Mathematical Statistics. Kapoor & Gupta

Websites and search engines:-

1. <https://www.google.com>
2. <https://www.wikipedia.com>

Questionary

1) What is your name?

2) Gender

a) Male

b) Female

3) Where do you live?

a) Rural

b) Urban

3) What is your age?

a) 0-15

b) 15-30

c) 30-45

d) 45-60

e) Above 60

4) Rate the type of news according to your preference?

a) Education: 1 2 3 4 5

b) Sport: 1 2 3 4 5

c) Political: 1 2 3 4 5

d) Entertainment : 1 2 3 4 5

e) Agriculture: 1 2 3 4 5

5) What kind of media do you use watching/ hearing/reading for daily information?

a) TV

b) Newspaper

c) Social media

6) From which media you get most of the educational information?

a) Newspaper

b) TV

c) Social media

7) Are you satisfied with information from media which you have choose?

a) Yes

b) No

8) Do you get unbiased political news from media?

- a) Yes
- b) No

9) What effects does the news have on our minds?

- a) Good
- b) Depression
- c) Anxiety

10) How much times you spend on news before covid-19 and after covid-19?

- a) 0-30
- b) 30-60
- c) 60-90
- d) 90-120
- e) 120-150

11) Do you get the right information during covid-19?

- a) Yes
- b) No

12) Role of media during covid-19 period according to you?

- a) Encouraging
- b) Depressing
- c) Fear

13) According to you why media is considered as powerful tool?

- a) It provide correct news
- b) Provide updated news of everything
- c) Most of your lives is influenced by media

14) Media able to provide an impartial vision?

- a) Yes
- b) No

15) How do the media influence our life?

- a) Create awareness
- b) Makes people enlightened citizens

Project Work For B.Sc. III In Year 2020-21.

Sr. No.	Name of the Students	Title of the Project	Name of Mentor
1	Shinde Gurudas Ananda	To study the challenges of online education over traditional education	Dr. M. K. Patil
	Yamgar Bhavesh Bajirao		
	Kulkarni Abhishek Shrikant		
	Kulkarni Sushilkumar Jagdish		
2	Bhapkar Amruta Shivaji.	To study the relation between age, weight and blood pressure	Mr. V. V. Koshti
	Mujawar Swaleha Yunus		
	Patil Ankita Chavagounda.		
	Pawar Swati Pradip		
	Shaikh Simran Shoukat		
3	Kadam Komal Sukhdev	Roles of fertilizers on grapes production	Mr. A. P. Patil
	Kadam Gauri Suresh		
	Bhosale Seema Ajay		
	More Aishwarya Arvind		
4	Babar Aishwarya Sanjay	Statistical Analysis of effect of fertilizers and soil type on sugarcane	Miss. S.S. Bhosale
	Patil Bhagyashri Harishchandra		
	Chavan Sneha Shrikant		
	Patil Vaishnavi Vilas		
5	Shinde Jyoti Subhash	To study the time spent by students on social media	Miss. S. S. Patil
	Mujawar Simaran Nisar		
	Mali Pooja Sahadev		
	Mali Sonali Vithoba		
	Kore Pratiksha Jayant		

A
Project Report
On

**“TO STUDY THE CHALLENGES OF ONLINE EDUCATION
OVER TRADITIONAL EDUCATION”**

Submitted to

Padmabhushan Vasantodada Patil Mahavidyalaya,
KavatheMahankal.
DEPARTMENT OF STATISTICS

By,

Mr. Shinde Gurudas Ananda
Mr. Yamgar Bhavesh Bajirao
Mr. Kulkarni Abhishek Shrikant
Mr. Kulkarni Sushilkumar Jagdish

Under the Guidance of,

Prof. Dr. Patil M. K.
Prof. Koshti V. V.

Prof. Patil A. P.
Prof. Bhosale S. S.
Prof. Patil S. S.

Through,

The Principal,
Padmabhushan Vasantodada Patil
Mahavidyalaya, Kavathe Mahankal.

For the academic year 2020-2021

**Padmabhushan Vasantodada Patil Mahavidyalaya, Kavathe
Mahankal**

DEPARTMENT OF STATISTICS

CERTIFICATE

This is to certify that the following students are working under my guidance.

They have satisfactorily carried out and completed the project work for the B. Sc. Part III. They have presently submitted their project entitled “TO STUDY THE CHALLENGES OF ONLINE EDUCATION OVER TRADITIONAL EDUCATION”

This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

1. Mr. Shinde Gurudas Ananda.
2. Mr. Yamgar Bhavesh Bajirao.
3. Mr. Kulkarni Abhishekh Shrikant.
4. Mr. Kulkarni Sushilkumar Jagdish.

Dr. M. K. Patil

Teacher in charge

Dr. M. K. Patil

Examiner

Dr. M. K. Patil

Head of Department

DECLARATION

We hereby declare that the project entitled “TO STUDY THE CHALLENGES OF ONLINE EDUCATION OVER TRADITIONAL EDUCATION”.

It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been submitted previously for any other degree of any other purpose.

Sr. No.	Roll No.	Name	PRN	Sign
1.	1	Shinde G.A.	2018010208	
2.	2	Yamgar B.B	2018014653	
3.	3	Kulkarni A.S	2017045567	
4.	4	Kulkarni S.J	2018014505	

Place : Kavathe Mahankal

Date :

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude to Dr. Patil M. K. who gave backing and his valuable guidance for this project work. We further thank to Prof. Koshti V.V. Prof. Bhosale S.S. Prof. Patil A.P. and Prof. Patil S.S. Who gave us their encouragement, kind support, and constant inspiration during our project, work further we are thankful to all of our college students who gave us necessary information to carry out this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immerse support during every stage of the project work.

CONTENTS

Sr. No.	Name	Page No.
1	Introduction	6
2	Objective	8
3	Data collection and Sampling Method	9
4	Statistical tools used	10
	4.1) Graphical representation	
	4.2) Chi Square test for independence of two attributes.	
	4.3) Z test for population proportion of two samples.	
5	Analysis of Data	12
6	Summary	19
7	Bibliography and Questionnaire	20

INTRODUCTION

As the first wave of COVID-19 entered India in early 2020, schools were closed in mid-March, just weeks before the end of the 2019-20 academic years.

The vast majority of the country's 26 crore school children have not set foot in a school since then, depending instead on various forms of distance education.

The availability of digital education – whether via live, synchronous teaching on platforms like Zoom, or through recorded lectures, emails, WhatsApp or educational apps - was largely dependent on whether schools, teachers and parents had access to the necessary infrastructure. In many States, teachers came to school and taught in their own empty classrooms, using their blackboards and lab facilities, while facing a computer screen that communicated the lessons to their students at home.

Our aim is to bring out real effect of COVID-19 on educational sector and whether do we have necessary infrastructure to enhance online education in future? And what students/ learners think about online education, what major issues they are facing to educate themselves through online method. We sampled students from various colleges, universities and high-schools, from very beginning to, that is, Primary, to last phase of one's education, that is, Masters, Because there seems to be big difference between how college graduates, who have been already learning from various online sources and the school noob's who had never used such method for education till now.

We also discussed **Online v/s Offline education** with some school teachers and professors, about what challenges they are facing to teach online, because none of them have been through this kind of time .During sampling we found that, it's really hard for teachers to teach student below sixth or seventh grade and thereafter collaborative curve increases exponentially.

The intention behind this project is to analyze the grave flexibility of educational system and its infrastructure to tackle unbearable situation like this, because we are spending hefty amounts to get on Mars but one unprecedented 80 nm sized virus could shut the earth down, and make us question on the reliability.

This project will provide all the aspect of students who are being educated by online scheme, what are the ups and downs of online education, comparison of online learning with traditional learning through various options. Conclusions can be drawn by statistical graphs and various statistical tests.

OBJECTIVES OF STUDY

1. To get the basic idea about the current online infrastructure through various statistical graphs.
2. To compare the effectiveness of online and offline learning.
3. To study the dependence of method of learning and its mode of education.
4. To study the various educational issues and their dependence on mode of education.
5. To study the preference of junior students (below age 18) that of the senior students (above age 18) regarding to online and offline education.

Source of Data & Sampling Method

The main objective of this project is, “To compare the online education with traditional education”. We have sampled students from different educational phases, to ensure that conclusions apply to various situations. Also this project is intended to know about the current methods that are students and teachers are using for online teaching-learning.

We designed a questionnaire to fulfill the objectives behind this project accurately. The data collection procedure was done using Google forms and by one-to-one approach. The perspective of teachers towards online education has affected crucially to this project.

We sampled subjects using simple random sampling without replacement (SRSWOR) method. The collection of sampling has been taken through online questionnaire using Google forms and direct interviewing with students.

STATISTICAL TOOLS

Following are some statistical tools us towards the appropriate results and finding.

• *Graphical Representation*

To represent our data graphically, we use the multiple bar diagram, pie chart method etc. To construct the appropriable divided bar diagram of given data we used Microsoft Excel.

- i. ***Pie Chart:*** A pie chart displays data, information and statistics in an easy-to-read 'pie-slice' format with varying slice sizes telling you how much of on data element exists. The bigger the slice the more of that particular data was gathered.
- ii. ***Bar Chart:*** A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

• *Test for independence of attributes*

To test the independence of attributes, here data are in the form of cross tables. Where, A_i = Educational Methods/Problems

B_j = Mode of education

Hypotheses:

H_0 : Attributes A_i and B_j are independent of each other.

vs

H_1 : Attributes A_i and B_j are dependent of each other.

Under H_0 the test statistics is

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Statistic χ^2 has (R-1) (C-1) d .f.

The value of χ^2 is obtained with the help of formula. To make decision about H_0 , Calculated value of χ^2 is compared with table value of χ^2 for (R-1) (C-1) d. f and α % level of significance (l.o.s)

Conclusion:

If cal. $\chi^2 \geq \text{tab. } \chi^2$, then we reject H_0 at α % l. o. s. Otherwise we accept H_0 .

• **Test for population proportion:**

To test the population proportion of two samples.

Hypothesis:

H_0 : Two population proportions are same. ($p_1 = p_2$)

Vs

H_1 : Two population proportions are not same. ($p_1 \neq p_2$)

Under H_0 , the test statistic is given by,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Let α be the level of significance and Z_α being the critical value.

Conclusion:

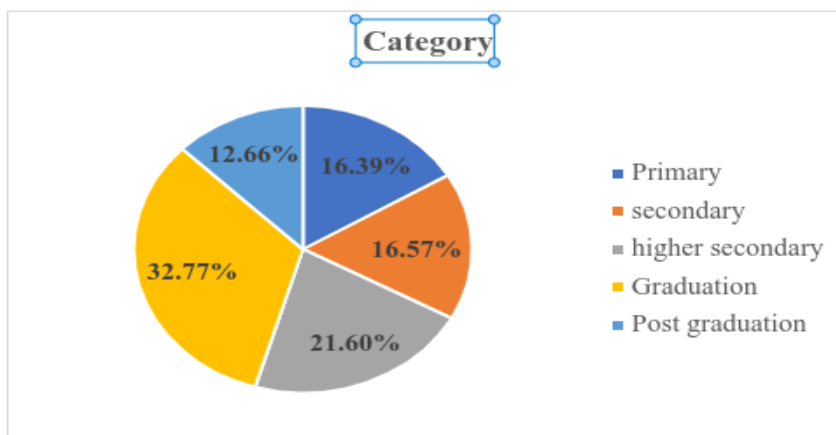
If calculated $|Z| > \text{tabulated } Z$, we reject H_0 at α % l.o.s otherwise we accept it.

STATISTICAL ANALYSIS

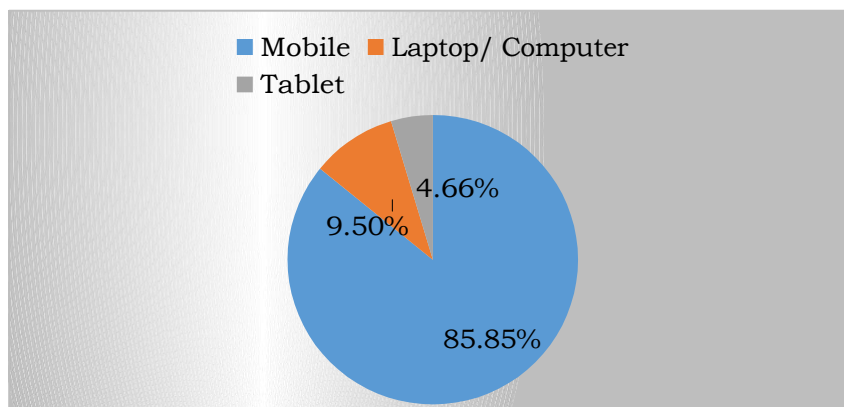
1. To get the basic idea about the current online infrastructure through various statistical graphs.

i) To which category do you belong?

This data are merely for sample and have done proportion test in further analysis.



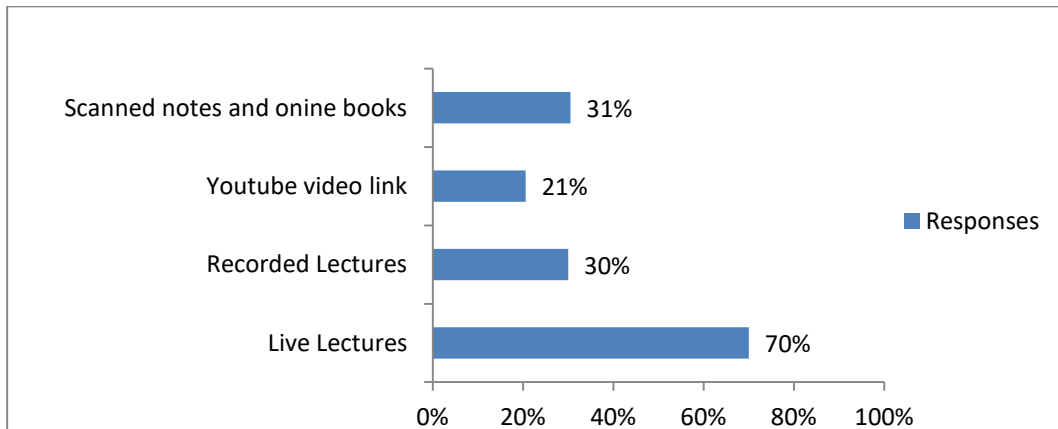
ii) Which device do you use for online learning?



Conclusion:

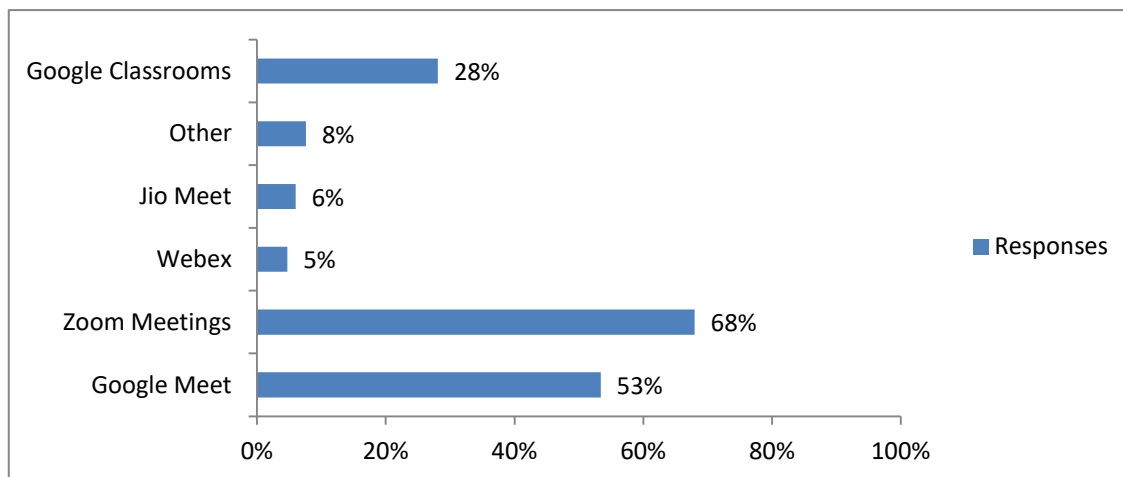
From the above graph we conclude that, most of the students are using Mobile as their primary learning device.

iii) Which method do you use for online learning?



Conclusion: From the above graph we can conclude that, **Live lectures** are popular among students and teachers.

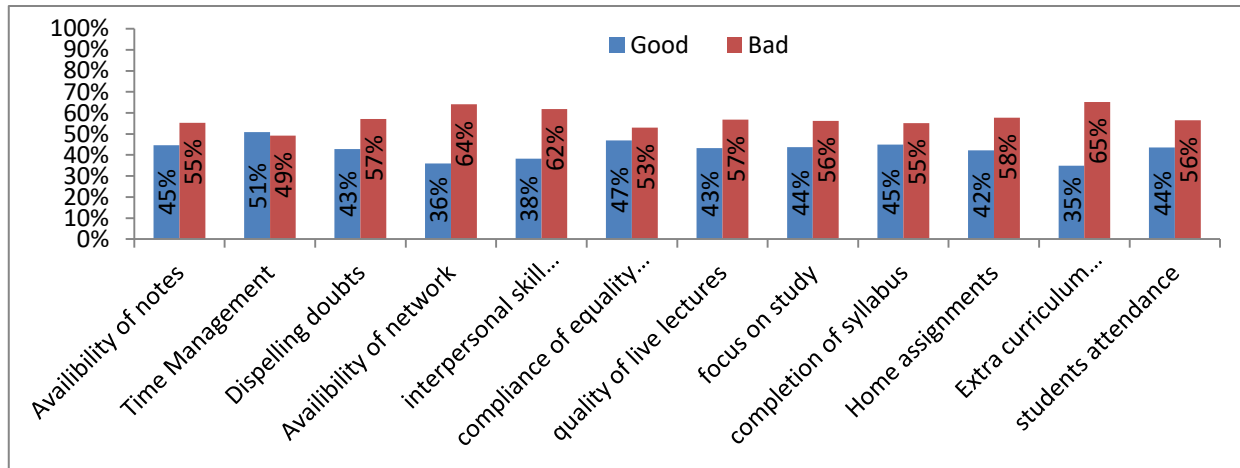
iv) Which application do you use for online learning?



Conclusion: **Zoom meetings** and **Google meet** are popular online learning platforms among all.

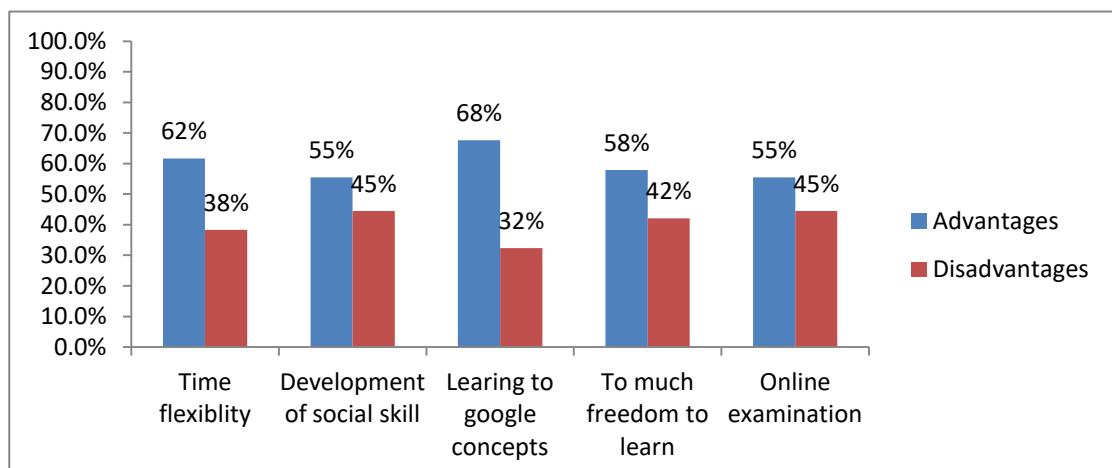
2. To compare the effect of online and offline learning.

i) Please answer by circling one number against each statement in case of 'Online Learning'.



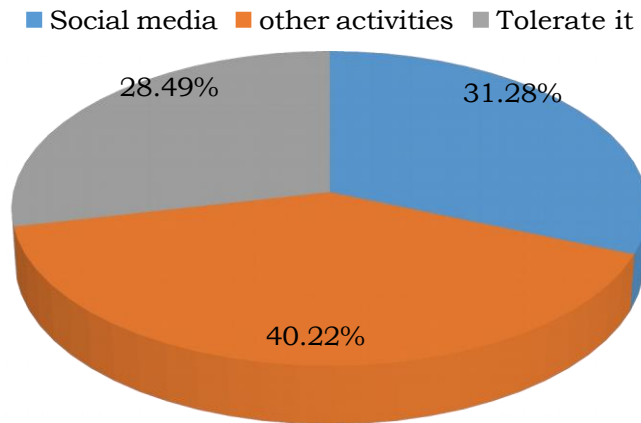
Conclusion: According to the above graph, if we were to summed up, during online lectures, technical issues and being deprived from social activities are inevitable.

ii) If online education implemented properly, its advantages or disadvantages



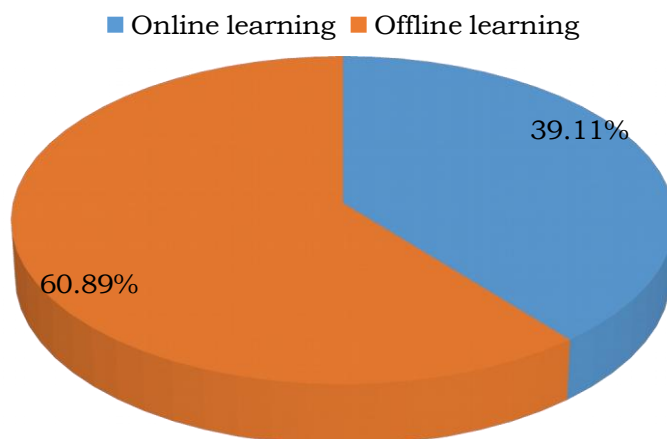
Conclusion: Above graph suggests that, students think Time management, Freedom and Online Examination are perks over traditional education system, if online education implemented properly.

iii) What do you do during weary online lectures?



Conclusions: Above graph suggests that, Students tend to **other activities** and **Social media** when online lectures are boring.

iv) What would you prefer for long learn?



Conclusion: From the above graph we conclude that, still most of the students prefer offline as their main method of education.

3. To study the dependence of method of learning and its mode of education.

Here we use Chi Square test for independence of attributes to test the hypothesis. To check whether different methods of learning and mode of education are independent or not.

H₀: Different methods of learning and mode of education are independent.

vs

H₁: Different methods of learning and mode of education are dependent.

Observation Table:

	Online	Offline	Expected online	Expected offline
Project	211	326	208.5	328.5
Teaching/ Learning	211	326	208.5	328.5
Getting Knowledge	238	299	208.5	328.5
Dispelling Doubts	208	329	208.5	328.5
Social Interaction	194	343	208.5	328.5
Practicals	189	348	208.5	328.5

Under H₀, The test statistics is

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$= 11.5526$$

Tabulated value of χ^2_5 at 5% level of significance = 11.0705

Result: Here, Calculated value of χ^2 is 11.5526 > tabulated value of χ^2_5 is 11.0705, Hence we reject H₀.

Conclusion: Method of learning and mode of education are dependent on each other. And also concludes students find current online education method less effective.

4. To study the various educational issues and their dependence on mode of education.

Here we check whether current online infrastructure is more problematic or not.

H₀: Educational problems are independent of mode of education.

vs

H₁: Educational problems are dependent of mode of education.

Observation table:

Problems	Yes or it is	No or its not	Expected Yes's	Expected No's
Bad effect on mental health	267	270	291.8333	245.1667
Is digital divide an issue	265	272	291.8333	245.1667
Decrease in expenditure	310	227	291.8333	245.1667
Increase in procrastination	298	239	291.8333	245.1667
Bought new device for lectures	301	236	291.8333	245.1667
Participation in online competition	310	227	291.8333	245.1667

Under H₀ Test statistic is,

$$\chi^2 = \sum_{i=1}^R \sum_{j=1}^C \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

$$=15.9028$$

Tabulated value of χ^2_5 at 5% level of significance = 11.0705

Result: Here, Calculated value of χ^2 is 15.9028 > tabulated value of χ^2_5 is 11.0705, Hence we reject H₀.

Conclusion: Hence we conclude that, Educational problems are dependent on mode of education. Issue of digital divide and increase in procrastination or distractions are the biggest problems of online education.

5.To study the preference of junior students that of the senior students regarding to online and offline education.

Here we want to check, what proportion of students and from which phase prefer which education method. This is age wise comparison, junior students (<18) and senior students (>18) according to their preference.

H₀: There is no significant difference between preference of junior students and senior students regarding to online and offline education.

vs

H₁: There significant difference between preference of junior students and senior students regarding to online and offline education.

Observation Table:

	Online	Offline
Junior Students	100	144
Senior Students	83	161

Under H₀ Test statistic is,

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\hat{p}(1 - \hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= 1.589584$$

Tabulated value of Z at 5% level of significance = 1.64485

Result: Here, Calculated value of Z is 1.589584 < tabulated value of Z is 1.6448. Hence we accept H₀.

Conclusion: There is no significant difference between preference of junior students and senior students regarding to online and offline education. All students find offline should be the mode of the education irrespective of their age.

SUMMARY

Overall conclusion can be drawn from this project is, Students have strong bias towards traditional education.

Reasons are as follows:

- i) Not having proper online infrastructure.
- ii) Freedom of time and everything which tend to increase in procrastination.
- iii) No social interfere and lack of interpersonal skill development.
- iv) More distractions.
- v) Online held Examinations and Competitions can't be fair.
- vi) Absentee of students to online held lectures.

Whilst, online education could be game changer if implemented properly, as now it is in its early stages. Teacher dominant environment could solve most of its flaws, and issues related to social interactions might be solved by creating online communities.

Sampled data suggest that, above conclusions are strongly valid to the population. However, the absentee of students leads to increase in dropout rate, also during sampling we found some students don't have device to attend online lectures, which is one of the reason for absentee of the students. The conclusions made from this project further confirm the UDISE+ data released in July 2021.

BIBLIOGRAPHY

1. Basic Statistics, B. L. Agarwal, *New Age International* (2006)
2. Mathematical Statistics, Gupta & Kapoor, *Sultan Chand & Sons*
(2017)
3. www.pib.gov.in.
4. www.udiseplus.gov.in
5. www.mospi.nic.in

QUESTIONNAIRE

Questionnaire for Statistical Study of online learning

A Project Report On

**“ TO STUDY THE RELATION BETWEEN
AGE,WEIGHT AND BLOOD PRESSURE ”**

Submitted to.

(Department of Statistics)

**Padmabhushan Vasantraodada Patil Mahavidyalaya,
Kavathe Mahankal.**

By

- 1) Bhapkar Amruta Shivaji.
- 2) Mujawar Swaleha Yunus.
- 3) Patil Ankita Chavagounda.
- 4) Pawar Swati Pradip.
- 5) Shaikh Simran Shoukat.

Under the Guidance of

Prof.Dr.M.K.Patil. (Head of the department.)

Prof.V.V.Koshti.

Mr.A.P.Patil.

Miss.S.S.Bhosale.

Miss.S.S.Patil.

Through

Prof.Dr.M.K.Patil

2020-2021

Shikshan Prasarak Sanstha's
Padmabhushan Vasantodada Patil Mahavidyalaya
Kavathe Mahankal.

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for the B.Sc. Part III. They are presently submitted their project entitled. This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

“TO STUDY THE RELATION BETWEEN AGE, WEIGHT AND BLOOD PRESSURE”.

1. Miss. Bhapkar Amruta Shivaji.
2. Miss. Mujawar Swaleha Yunus.
3. Miss. Patil Ankita Chavagounda.
4. Miss. Pawar Swati Pradip.
5. Miss. Shaikh Simran Shoukat.

Head

Teacher in charge. Examiner. •Department of Statistics•

Declaration

We hereby declare that the project entitled **“TO STUDY THE RELATION BETWEEN AGE, WEIGHT AND BLOOD PRESSURE ”**. It is the original work carried out by us at the Department of Statistics. This project or part of it has not been submitted previously for any other degree or any other purpose.

Place: Kavathe Mahankal.

Date:

Sr.no	Name	Exam seat No.	Signature
1.	Bhapkar Amruta shivaji		
2.	Mujawar Swaleha Yunus		
3.	Patil Ankita Chavagounda		
4.	Pawar Swati Pradip		
5.	Shaikh Simran Shoukat		

Acknowledgment

We express our hearty thanks and deepest gratitude goes to Shri. Koshti Sir, who made valuable guidance and co-operation during our project work. We are rather infused by the guidance of Mr.Patil sir, Miss. Bhosale madam, Miss. Patil madam, Who put us their encouragement, kind support, and constant inspiration during our project work further we are thankful to all of our college students who gave us the necessary information to carry this project.

Finally, we give sincerest thanks to all our family members For keeping our spirit up and their immense support during every stage of the project work.

Index

Sr. No	Name	Page No.
1.	Introduction	6
2.	Objectives	7
3.	Data collection and sampling method	8
4.	Statistical Tools	9-13
5.	Statistical Analysis	14- 23
6.	Overall conclusion	24
7.	Scope and Limitations	25
8.	Bibliography	26- 27
9.	Questionnaires	28- 32

Introduction

Blood pressure (BP) is the pressure of circulating blood against the walls of blood vessels. Most of this pressure results from the heart pumping blood through the circulatory system. When used without qualification, the term "blood pressure" refers to the pressure in the large arteries. Blood pressure is usually expressed in terms of the systolic pressure (maximum pressure during one heartbeat) over diastolic pressure (minimum pressure between two heartbeats) in the cardiac cycle. It is measured in millimeters of mercury (mmHg) above the surrounding atmospheric pressure.

Blood pressure is one of the vital signs together with respiratory rate, heart rate, oxygen saturation, and body temperature that healthcare professionals use in evaluating a patient's health. Normal resting blood pressure, in an adult is approximately millimeters of mercury (16 kPa) systolic over 80 millimeters of mercury (11 kPa) diastolic, denoted as "120/80 mmHg". Globally, the average blood pressure, age standardized, has remained about the same from 1975 to the present, at approx. 127/79 mmHg in men and 122/77 mmHg in women, although these average data mask quite large divergent regional trends. Blood pressure is influenced by cardiac output and systemic vascular resistance and arterial stiffness and varies depending on the situation.

OBJECTIVES OF STUDY

- 1) To study the information about blood pressure patients.
- 2) To study the symptoms of blood pressure.
- 3) To study at what age blood pressure is more likely to occur.
- 4) To study who has high blood pressure according to gender.
- 5) To study whether blood pressure is more dependent on weight and age.
- 6) To study whether blood pressure can be controlled without medicine.
- 7) To study the side effect of blood pressure.

Source of Data & Sampling Method

As the main aim of our project is “To study the relation between age, weight and blood pressure”. We consider different blood pressure are side effects, symptoms, more likely to age, more depend on age and weight, diet, high blood pressure to gender, etc. So we decide to collect all primary data from the villages Dashing, Hingangaon, Shindevadi, Vithurayachi Vadi, and Kavathe Mahankal, which will provide us the necessary and sufficient data within less time and more accuracy.

So we divide the selected area and from each village, we selected a sample of size 250 by simple random sampling method of the data is collected by filling questionnaires from each selected blood pressure patients by interview method.

Statistical Tools

The following are some statistical tools that lead us towards the appropriate results and findings.

1. Graphical representation:

To represent our data graphically, we use the multiple bar diagram, pie chart method, etc. To construct the appropriate divided bar diagram of given data we used Microsoft Excel.

i) Bar Chart: A bar chart or bar graph is a chart that presents grouped data with rectangular bars with lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

ii) Pie chart: A pie chart displays data, information, and statistics in an easy-to-read 'pie-slice' format with varying slice size telling you how much of one data element exists. The bigger the slice the more of that particular data was gathered.

2. Multiple correlation coefficient:

Multiple linear regression means, “ a numerical index of the degree of relationship between a particular variable and two or more other variables”.

The multiple correlation coefficient of X1 on X2 and X3 is denoted by $R_{x1.x2x3}$ and is given by,

$$R_{1.23}^2 = (r_{12}^2 + r_{13}^2 - 2r_{12}.r_{13}.r_{23}) / (1 - r_{23}^2)$$

$$\text{Similarly, } R_{2.13}^2 = (r_{12}^2 + r_{23}^2 - 2r_{12}.r_{13}.r_{23}) / (1 - r_{13}^2)$$

$$\text{And } R_{3.12}^2 = (r_{13}^2 + r_{23}^2 - 2r_{12}.r_{13}.r_{23}) / (1 - r_{12}^2)$$

The sample correlation coefficient is,

$$r_{12} = \text{Cov}(x_1, x_2) / \sigma_1 \sigma_2$$

$$r_{13} = \text{Cov}(x_1, x_3) / \sigma_1 \sigma_3$$

$$r_{23} = \text{Cov}(x_2, x_3) / \sigma_2 \sigma_3$$

And the partial correlation coefficient is,

$$r_{12.3} = (r_{13} - r_{23} * r_{12}) / \sqrt{(1 - r_{23}^2)(1 - r_{12}^2)}$$

$$r_{13.2} = (r_{12} - r_{23} * r_{13}) / \sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}$$

$$r_{23.1} = (r_{23} - r_{12} * r_{13}) / \sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}$$

Where,

$$\text{Cov}(x_1, x_2) = 1/n * \sum x_1.x_2 - x_{1\bar{}} * x_{2\bar{}}$$

$$\text{Cov}(x_1, x_3) = 1/n * \sum x_1.x_3 - x_{1\bar{}} * x_{3\bar{}}$$

$$\text{Cov}(x_2, x_3) = 1/n * \sum x_2.x_3 - x_{2\bar{}} * x_{3\bar{}}$$

$$\sigma_1^2 = 1/n \sum (x_1 - \bar{x}_1)^2$$

$$\sigma_2^2 = 1/n \sum (x_2 - \bar{x}_2)^2$$

$$\sigma_3^2 = 1/n \sum (x_3 - \bar{x}_3)^2 \text{ and}$$

$$\bar{X}_1 = \sum x_1 / n$$

$$\bar{X}_2 = \sum x_2 / n \text{ and } \bar{X}_3 = \sum x_3 / n$$

Now, the test for sample correlation coefficient is,

$H_0: r_{12} = 0$ (The blood pressure is depend on age)

V/S

$H_1: r_{12} \neq 0$ (The blood pressure is not depend on age)

The test statistic is,

$$t = r / \sqrt{1 - r^2} * \sqrt{n - 1} \sim t_{n-2}$$

Let, α be the 5% level of significance.

Conclusion: If $t > t_{\alpha}$, we reject H_0 at 5% level of significance otherwise accept H_0 .

The test for partial correlation coefficient is,

$H_0: r_{12.3} = 0$ (The blood pressure is depend on age and weight.)

V/S

$H_1: r_{12.3} \neq 0$ (The blood pressure is not depend on age and weight.)

Under H_0 , the test statistic is,

$$t = r / \sqrt{(1-r^2) * \sqrt{(n-k-2)}} \sim t_{n-k-2}$$

Let, α be the 5% level of significance.

Conclusion: If $\text{cal } t > \text{tab } t$, we reject H_0 at 5% level of significance otherwise accept H_0 .

The test for multiple correlation coefficient is,

$H_0: R_{1.23} = 0$ (The blood pressure is depend on age and weight)

V/S

$H_1: R_{1.23} \neq 0$ (The blood pressure is not depend on age and Weight.)

Under H_0 , the test statistic is,

$$F = R^2 / (1-R^2) * (n-k-1) / k \sim F(k, n-k-1)$$

Let, α be the 5% level of significance.

Conclusion: If $\text{cal } F > \text{tab } F$, we reject H_0 at 5% level of significance otherwise accept H_0 .

3. Test for one sample population proportion:

we have to test,

$H_0 : P = P_0 = 0.5$ (The population proportion is equal to 0.5)

V/S

$H_1 : P > P_0 = 0.5$ (The population proportion is greater than 0.5)

Under H_0 , the test statistic is,

$$Z = (p - P_0) / \sqrt{(P_0 \cdot Q_0 / n)}$$

Where,

$P_0 =$ Population proportion under $H_0 = 0.5$

$P =$ Sample proportion

$Q_0 = 1 - P_0 = 1 - 0.5 = 0.5$

$n =$ Sample size

Let, α be the 5% level of significance.

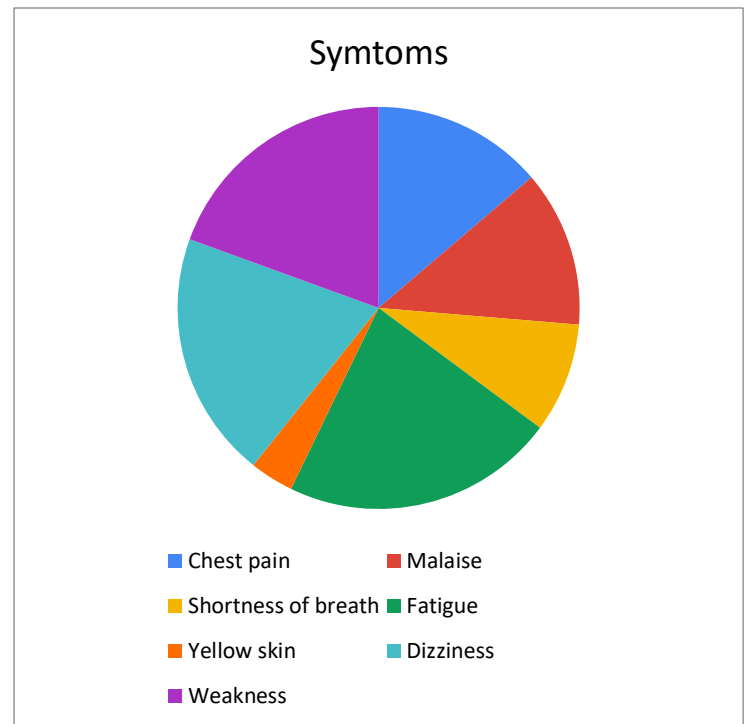
Conclusion:

If cal. $Z \geq 1.645$, then we reject H_0 otherwise accept H_0 at 5% level of significance .

Statistical analysis

1) To study the symptoms of blood pressure.

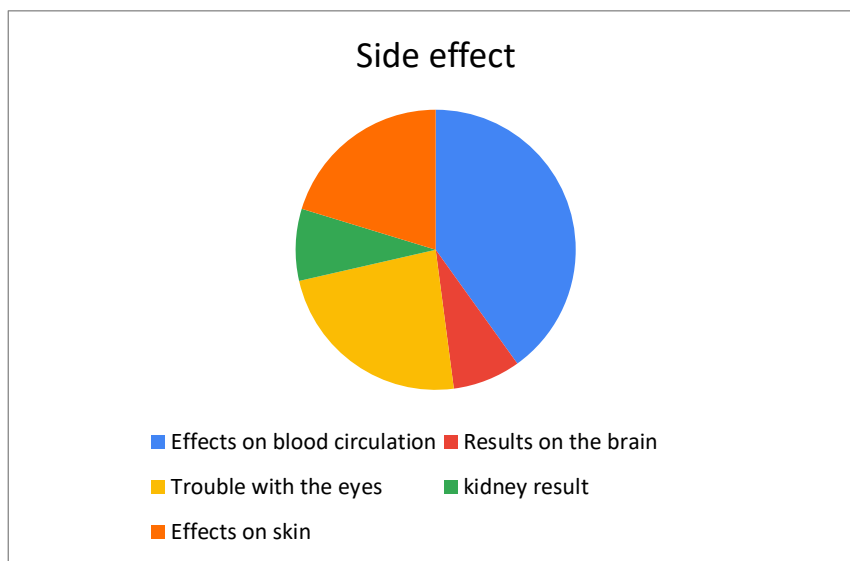
Symptoms	Total
Chest pain	23
Malaise	20
Shortness of breath	9
Fatigue	99
Yellow skin	6
Dizziness	50
Weakness	43



Comment: From the above diagram we conclude that , the most common symptom of blood pressure is fatigue.

2) To study the side effect of blood pressure.

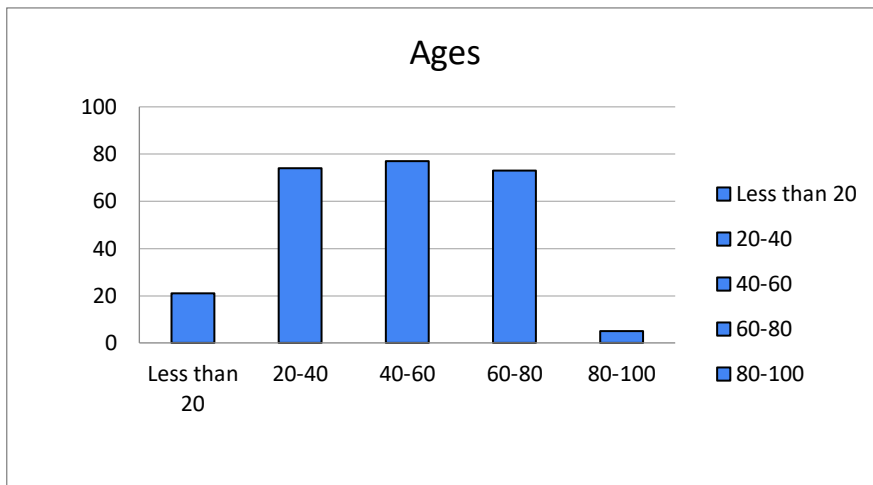
Side effect	Total
Effects on blood circulation	89
Results on brain	56
Trouble with the eyes	60
Kidney result	20
Effects on skin	25



Comment: From the above diagram we conclude that , the most common side effect of blood pressure is blood circulation.

3) To study at what age blood pressure is more likely occur.

Ages	Total
Less than 20	21
20-40	74
40-60	77
60-80	73
80-100	5



Comment: From the above diagram, The highest BP is in the age group of 40-60.

4)To study whether bp is more dependent on weight and age.

We have to test,

$H_0: R_{1.23} = 0$ (The blood pressure is depend on age and weight.)

V/S

$H_1: R_{1.23} \neq 0$ (The blood pressure is not depend on age and weight.)

Let, α be the 5% level of significance.

Observation Table:

X1	X2	X3	X1 ²	X2 ²	X3 ²	X1X2	X1X3	X2X3
140	40	75	19600	1600	5625	5600	10500	3000
140	49	70	19600	2401	4900	6860	9800	3430
142	45	66	20164	2025	4356	6390	9372	2970
145	60	45	21025	3600	2025	8700	6525	2700
157	50	75	24649	2500	5625	7850	11775	3750
135	54	73	18225	2916	5329	7290	9855	3942
139	59	67	19321	3481	4489	8201	9313	3953
137	43	70	18769	1849	4900	5891	9590	3010
147	62	75	21609	3844	5625	9114	11025	4650
142	69	65	20164	4761	4225	9798	9230	4485
140	68	80	19600	4624	6400	9520	11200	5440
128	60	62	16384	3600	3844	7680	7936	3720

135	75	45	18225	5625	2025	10125	6075	3375
140	56	45	19600	3136	2025	7840	6300	2520
155	72	75	24025	5184	5625	11160	11625	5400
135	64	65	18225	4096	4225	8640	8775	4160
120	59	65	14400	3481	4225	7080	7800	3835
76	48	69	5776	2304	4761	3648	5244	3312
190	85	50	36100	7225	2500	16150	9500	4250
75	46	73	5625	2116	5329	3450	5475	3358
170	65	52	28900	4225	2704	11050	8840	3380
160	75	87	25600	5625	7569	12000	13920	6525
136	19	50	18496	361	2500	2584	6800	950
124	36	56	15376	1296	3136	4464	6944	2016
147	40	54	21609	1600	2916	5880	7938	2160

Totals	$\sum X_1=3455$	$\sum X_2=1399$	$\sum X_3=1609$	$\sum X_1^2=491067$
--------	-----------------	-----------------	-----------------	---------------------

$\sum X_1X_2=196965$	$\sum X_1X_3=221357$	$\sum X_2X_3=90291$
----------------------	----------------------	---------------------

Calculation: Here,

$\bar{X}_1=138.2$, $\bar{X}_2=55.96$ and $\bar{X}_3=64.36$

$\sigma_1 = 23.31179959$, $\sigma_2=14.40411052$ and $\sigma_3=11.53734805$

$n_1=25$. and $n_2=25$

$Cov(X_1,X_2)= 144.928$

$Cov(X_1,X_3)= -40.272$ and

$Cov(X_2,X_3)=10.0544$

Sample correlation coefficient is,

$$r_{12}=0.431608564$$

$$r_{13}=-0.14973434$$

$$r_{23}=0.060501163$$

Multiple correlation coefficient between X1 on X2 and X3 is

$$R_{1.23}=0.461373932$$

Similarly,

$$R_{2.13}=0.444910847 \quad \text{and}$$

$$R_{3.12}=0.188920573$$

Partial correlation coefficient is,

$$r_{12.3}=0.44487583$$

$$r_{13.2}=-0.19458221 \quad \text{and}$$

$$r_{23.1}=0.137149403$$

Testing of the values:

Sample correlation coefficient is,

$$t_{cal}=2.2946$$

$$t_{cal}=0.7261$$

$$t_{cal}=0.2906$$

Partial correlation coefficient is,

$$t_{cal}=2.2758$$

$$t_{cal}=0.9086$$

$$t_{cal}=0.6342$$

Multiple correlation coefficient is,

F cal=2.9735

F cal=2.7146

F cal= 15.28

The tabulated value as,

Sample correlation=t tab=2.06865761

Partial correlation=t tab=2.079613845

Multiple correlation=F tab=3.443356779

Result: Here, α be the 5% level of significance.

Sample correlation coefficient

- 1) If cal t > tab t, we reject Ho
- 2) If cal t < tab t, we accept Ho
- 3) If cal t < tab t, we accept Ho

Partial correlation coefficient

- 1) If cal t > tab t, we reject Ho
- 2) If cal t < tab t, we accept Ho
- 3) If cal t < tab t, we accept Ho

Multiple correlation coefficient

- 1) If cal F < tab F, we accept Ho
- 2) If cal F < tab F, we accept Ho
- 3) If cal F > tab F, we reject Ho

Conclusion: We conclude that,

For the sample correlation coefficient,

- 1) The blood pressure is depend on age.

2) The blood pressure is depend on weight.

For the partial correlation coefficient,

1)The blood pressure is depend on age.

2) The blood pressure is depend on weight.

For the multiple correlation coefficient,

1)The age is depend on blood pressure.

2)The weight is does not depend on blood pressure.

5) To study who has high BP according to Gender.

we have to test,

$H_0: P = P_0 = 0.5$ (The male and female are equally affected by blood pressure.)

V/S

$H_1: P < P_0$ ($P < 0.5$) (The females patients are less affected by blood pressure as compared to male blood pressure patients.)

Let, $\alpha = 5\%$ be the level of significance .

Calculation: Here,

$n =$ Total number of samples = 250

$X =$ Number of female = 121

$P =$ Sample proportion = $X/n = 121/250 = 0.48$

Under H_0 , the test statistics is,

$$Z = (0.484 - 0.5) / \sqrt{(0.25/250)}$$

$$Z = -0.5059$$

$$\text{Cal } Z = -0.5059$$

Let, α be the 5% level of significance.

$$\text{Tab. } Z = -1.645$$

Result: Here, the calculated value of Z is greater than the tabulated value of Z . Hence, we reject H_0 at a 5% level of significance.

Conclusion: We conclude that, the female blood pressure patients are less affected by blood pressure as compared to male blood pressure patients.

Overall all Conclusion

- 1) The most common symptom of blood pressure is fatigue.
- 2) The most common side effect of blood pressure is blood circulation.
- 3) The highest age group is the in blood pressure is 40-60.
- 4)
 - i) The blood pressure is depend on age.
 - ii) The blood pressure is depend on weight.
 - iii) The blood pressure is depend on age.
 - iv) The blood pressure is depend on weight.
 - v) The age is depend on blood pressure.
 - vi) The weight is does not depend on blood pressure.
- 5) The female blood pressure patients are less affected by blood pressure as compared to male blood pressure patients.

LIMITATIONS

- 1) We have to limit the scope of this research according to the result we obtain from the questionnaires.
- 2) We consider the whole population is normal but in reality, it may be or may not be normal.
- 3) The time constraint was one of the major limitations.

SCOPE

- 1) A well-planned Study can be taken including some characteristics.
- 2) It can be extended to a greater region for better estimates.

BIBLIOGRAPHY

This bibliography is intended as a resource for the design and implementation of blood pressure programs, and is useful for blood pressure treatment protocol development and adaptation. The aim is to provide a brief overview of the scientific basis of key aspects of blood pressure control for a comprehensive approach to blood pressure management. This list is not representative of all aspects of blood pressure management and is largely drawn from the scientific literature published in English.

The bibliography is organized in the following manner: first, a review of the burden of to underscore its importance as a non-communicable disease; second, a review of the evidence for protocol-based treatment; third, a review of the evidence on task sharing; fourth, a summary of the literature on medical supplies and patient-centered services; and lastly, a synopsis on information systems.

REFERRED BOOKS:

Sr. No	Book name	Author
1.	Control the blood pressure	Robert Rowan M.D
2.	Blood pressure solutions	Jessica Robbins
3.	The complete guide to lowering blood pressure naturally	Deborah Mitchell



Questionnaires

1) तुमचे वय किती आहे?

a) 20 पेक्षा कमी

b) 20-40

c) 40-60

d) 60-80

e) 80-100

2) तुमचे वजन किती आहे ?

a) 30-40

b) 40-50

c) 50-60

d) 60-70

e) 70-80

f) 80-90

g) 90-100

3) तुम्हाला कोणत्या प्रकारचे बीपी आहे?

a) जास्त

b) कमी

4) तुमच्या बीपी ची रेंज काय आहे?

a) 60-70

b) 70-80

c) 120-140

d) 140-160

e) 160-180

f) 180-200

g) 200 पेक्षा अधिक

5) तुम्ही कोणत्या प्रकारचे आहार घेता?

a) फळे

b) पालेभाज्या

c) मांसाहार

d) गोड पदार्थ

6) तुम्ही कोणत्या प्रकारचे आहार घेत नाही?

a) मीठ

b) आंबट पदार्थ

c) दुग्धजन्य पदार्थ

d) चहा कॉफी

7) तुम्ही व्यायाम करता का?

a) हो b) नाही

8) जर व्यायाम करत असतील तर कोणते ?

a) चालणे

b) पळणे

c) सायकल चालवणे

- d)** प्राणायाम
- e)** योगासने
- f)** नृत्य करणे
- 9)** तुम्ही कोणत्या प्रकारचे औषध घेता ?
- a)** ऍलोपॅथिक
- b)** होमिओपॅथिक
- c)** इतर
- 10)** तुम्हाला कोणत्या प्रकारची लक्षणे आढळून आलीत?
- a)** छातीत दुखणे
- b)** अस्वस्थता
- c)** श्वसनास त्रास
- d)** थकवा
- e)** पिवळी त्वचा
- f)** चक्कर मारणे
- g)** अशक्तपणा
- 11)** वजन वाढल्यास अथवा कमी झाल्यास तुमच्या बीपी वर कोणता परिणाम होतो का?
- a)** हो
- b)** नाही
- 12)** बीपी कंट्रोल मध्ये आणण्यासाठी तुम्ही कोणते प्रयत्न करता?
- a)** निरोगी वजन ठेवणे
- b)** अधिक व्यायाम करणे

c) मिठाचा वापर कमी करणे

d) फळे- पालेभाज्या आहारात वाढवणे

13) अचानक बीपी वाढल्यास तुम्ही काय करता?

a) घरगुती उपाय

b) दवाखान्यात जाणे

14) घरगुती उपाय कोणते?

a) लिंबू पाणी

b) तणाव कमी करणे

c) नारळ पाणी

d) इतर

15) बीपी मुळे तुमच्या शरीरावर परिणाम होतात का?

a) हो

b) नाही

16) बीपी मुळे तुमच्या शरीरावर कोणते परिणाम होतात?

a) रक्ताभिसरणवर परिणाम

b) मेंदूवर परिणाम

c) डोळ्यांचा त्रास

d) किडनी वर परिणाम

e) त्वचेवर परिणाम

17) बिपीमुळे मानसिक परिणाम होतात का?

a) भीती वाटणे

b) चिडचिड होणे

c) काळजी वाटणे

18) कोणत्या कारणामुळे तुम्हाला बिपीचा त्रास सुरू झाला?

a) तणाव

b) अधिक वजन

c) मद्यपान

d) अधिक वय

e) हृदय विकार

f) किडनी प्रॉब्लेम

g) प्रेग्नेसी

h) रक्ताची कमतरता

i) जीवनसत्त्वाची कमतरता

☺ Thank You ☺

A

Project Report

On

**“TO STUDY THE ROLES OF FERTILIZERS ON GRAPES
PRODUCTION”**

Submitted to

(Department of Statistics)

Padmabhushan Vasantrodada Patil Mahavidyalay Kavathe Mahankal

By

Miss. Kadam Komal Sukhdev

Miss.Kadam Gauri Suresh

Miss. Bhosale Seema Ajay

Miss.More Aishwarya Arvind

Under the Guidance of

Prof.Koshti Sir

Prof.Patil sir

Prof. Bhosale madam

Prof.Patil madam

Through

The Principal,

Dr.Patil.M.K

Padmabhushan Vasantrodada Patil Mahavidyalaya, Kavathe Mahankal

For the academic year 2020-2021

DEPARTMENT OF STATISTIC

Padmabhushan Vasatraodada Patil Mahavidyalaya, Kavathe Mahankal

CERTIFICATE

This is to certify that the following students are working under my guidance. They have satisfactorily carried out and completed the project work for the B. Sc. Part III. They is presently submitted their project entitled.

This work is genuine, original and the project or the part of it has not been submitted for the award of any other degree.

“TO STUDY THE ROLE OF FERTILIZERS ON GRAPES PRODUCTION”.

- 1. Miss. Kadam Komal Sukhdev.**
- 2. Miss.Kadam Gauri Suresh**
- 3. Miss. Bhosale Seema Ajay**
- 4. Miss. More Aishwarya Arvind**

Teacher in charge

Examiner

Head

•Department of Statistic•

DECLARATION

We hereby declare that the project entitled “**TO STUDY THE ROLE OF FERTILIZERS ON GRAPES PRODUCTION**”. It is the original work carried out by us at the Department of Statistics.

This project or part of it has not been submitted previously for any other degree of any other purpose.

Place: Kavathe Mahankal

Date:

Sr. No.	Roll No .	Name	Exam seat No.	Signature
1		Kadam.K.S		
2		Kadam.G.S		
3		Bhosale.S.A		
4		More.A. A		

AKNOWLEDGMENT

We express our hearty thanks and deepest gratitude goes to Prof.Koshti Sir, who made valuable guidance and co-operation during our project work. We are rather infused by the guidance of Prof.Patil sir, Prof. Bhosale madam , Prof. Patil madam, Who put us their encouragement, kind support, and constant inspiration during our project work further we are thankful to all of our college students who gave us necessary information to carry this project.

Finally we give sincerest thanks to all our family members For keeping our spirit up and their immense support during every stage of the project work.

INDEX

SR. NO.	NAME	PAGE NO
1	Introduction	6
2	Objectives	7
3	Source of data & sampling method	8
4	Statistical tools used:	
	a) Graphical representation	9
	b) Chi-Square test for independence of attributes	9-10
	c) Paired t-test for difference of two population means.	10
	d) Estimation of parameters by confidence interval.	11
	e) Correlation	11-12
5	Statistical analysis	13-23
6	Overall Conclusion	24
7	Limitations & Scope	25
8	Bibliography	26
9	Questionnaires	27-29

INTRODUCTION

The title of the project is “To study the role of fertilizer on Grapes production. “By using this data and some statistical tools for analysing that data. In this project we reported here some results related to the objectives of the survey. The main purpose of the project is to know the role of most important factor for production of grapes and are Fertilizers . India is agriculture country . 13.7% economy is depend on agriculture .In Maharashtra state Sangli and Nashik district are famous for grapes production .In Sangli district most production of grapes in Kavathe Mahankal Tehsil. In Kavathe Mahankal Tehsil 60% crops of grapes and occupation of most farmers are grapes .In grapes production the most affected factor are Fertilizers. Therefore we decided To Study The Role Of Fertilizers On Grapes Production. We have collected data in Agran Dhulgaon, landgewadi, Shelkevadi, Kokale in Kavathe Mahankal Tehsil. i.e. information of grapes production in three different villages .The information is age, education, grapes type, soil type, Fertilizers type, income per acre and many. We collected data by direct personal investigation and Google form.

Today Fertilizers has become essential to modern agriculture to feed the growing population . Chemical fertilizers are used extensively in modern agriculture , in order to improve crop yield. Urea is most popular and widely used dry N Fertilizers . Increased crop production largely relies on the type of Fertilizers used to supplement essential nutrients for plants. Now a day, there is increased emphasis on the impact on soil environment due to continuous use of chemical fertilizers.

OBJECTIVES

- ❖ To know the role of organic and inorganic Fertilizers in grapes production.
- ❖ To compare the proportion of farmers which use organic and inorganic fertilizers.
- ❖ To study independence among soil type and variety of grapes.
- ❖ To study difference between production of two years.
- ❖ To find percentage of farmers are satisfied with the rate for grapes.
- ❖ To study which variety give more income by using Fertilizers.
- ❖ To study the correlation between expenses on Fertilizers and income of grapes.

SOURCE OF DATA & SAMPLING METHOD

Sampling is part of our day to day life which we use advertently or inadvertently. For instance housewife takes one or two grains of rice from the cooking pan and decides whether it is cooked or not, a pathologist takes few drops of blood and test for any change in blood of the whole body. In all these situations, sampling is inevitable and gives satisfactory results.

As the main aim of our project “ To Study the role of Fertilizers in grapes production.” We consider different varieties of grapes ,types of soil,types of Fertilizers etc.So we have selected some villages Agran Dhulgaon, landgewadi,Shelkewadi, Kokale in Kavathe Mahankal Tehsil by convinient sampling method, which provides us necessary and sufficient data within less time and more accuracy. Thus the data is collected by filling questionnaire from the farmers by the interview method.

STATISTICAL TOOLS

The following are some statistical tools leads us towards the appropriate results and finding.

1) Graphical representation:

- a) Histogram: A histogram is a graphical representation of the distribution of numerical data . It is an estimate of the probability distribution of a continuous variables (quantitative variable) and was first introduced by Karl Pearson.
- b) Pie Diagram: A pie Diagram or pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart the arc length of each slice is proportional to the quantity is represent.
- c) Multiple Bar Diagram: In multiple bar diagram two or more sets of inter-related data are represented .The technique of making simple bar chart is used to draw this diagram but the difference is that we use different shades ,colors,or dots to distinguish between different phenomenon.

2) Chi-Square test for independence of attributes:

The hypothesis that, the one factor independent of other or not, can be tested by the Chi-Square test. The hypothesis that,

H₀: Two attributes are independent of each other.

H₁: Two attributes are dependent.

Under H₀ , the test statistic is,

$$X^2 = \sum \sum (O_{ij}^2 / E_{ij}) - N \sim \chi^2(m-1) (n-1) d.f$$

Statistic X^2 has $(m-1)(n-1)$ d.f in formula (a), E_{ij} is the expected frequency corresponding to (i,j) th cell of observed frequency O_{ij} .

Under the null hypothesis, the expected frequency is,

$E_{ij} = (\text{i th row total} * \text{j th column total}) / \text{sample size}$

$$= (R_i * C_j) / n$$

Once the expected frequency are calculated, the value of Chi-Square is obtained with the help of formula (a). To make a design about H_0 , calculated value of Chi-Square is compared with the table attributes are not independent. Otherwise accept H_0 .

3) Paired t-Test for difference of two population means:

Let us consider a particular situation where,

a) The sample size are equal i.e. $n_1 = n_2 = n$ (say)

b) The sample observation (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) are not completely independent but they are dependent in pairs.

Let $d_i = x_i - y_i$, $i = 1, 2, 3, \dots, n$ denote the difference in the observations for the i th unit.

Here, we want to test,

$H_0 : \mu_d = 0$ i.e. the difference between means is insignificant.

V/S

$H_1 : \mu_d \neq 0$ i.e. the difference between means is significant.

Under H_0 , the test statistic is,

$$t = (\bar{d}) / (s / \sqrt{n}) \sim t_{n-1}$$

Where, $d = Y_i - X_i$, here, $Y_i > X_i$, $\bar{d} = \sum d_i / n$ &

$$S^2 = 1 / (n-1) \sum (d_i - \bar{d})^2$$

Conclusion:

If calculated value of $|t| > t_{\alpha/2}(n-1)$, we reject the null hypothesis at $\alpha\%$ l.o.s and conclude that the difference between means is significant, otherwise we accept H_0 at $\alpha\%$ l.o.s.

4) Estimating parameter values by using confidence interval:

The assumption is justifiable only if the size of sample n is large and neither p nor q is very small. If n is large the value of either p or q is not very small, then the difference between them is not much and the distribution would be very much like a normal one. We have the value of pq is can never exceed $1/2 * 1/2$ or $1/4$. As such, if we do not know the value of p in the universe and if the value of p in the sample is very small or very large (so that the value of q is very large or very small) then instead taking the sample values of p and q . The value of pq may be taken as $1/4$ which is maximum value.

Then the standard error is,

$$S.E = \sqrt{(pq/n)}$$

And the confidence interval at $\alpha\%$ l.o.s is given by $p \pm q * Z \alpha$.

5) Correlation:

Correlation means , “ The study of existence and the magnitude and direction of variation between two or more variable”.

Karl Pearson’s coefficient of correlation: The most commonly used Method which gives a mathematical expression for correlation is the one suggested by Karl Pearson.

Here , $\sigma^2x = 1/N \sum (xi - x \text{ bar})^2$ gives a measure of variation in x .

& $\sigma^2y = 1/N \sum (yi - y \text{ bar})^2$ gives a measure of variation in y .

Also, $1/N \sum (xi - x \text{ bar}) \sum (yi - y \text{ bar})$ gives a measure of simultaneous variation in x & y . Then the correlation coefficient is,

$$r = (\sum (xi - x \text{ bar})(yi - y \text{ bar})) / (\sqrt{\sum (xi - x \text{ bar})^2 * \sum (yi - y \text{ bar})^2})$$

The probable error and coefficient of correlation:

The reliability of the coefficient of correlation is determined by calculating the probable error. If n is the size of the sample the probable error is given by,

$$\text{Probable error} = 0.6745 * (1 - r^2)/n$$

Where , r is correlation coefficient.

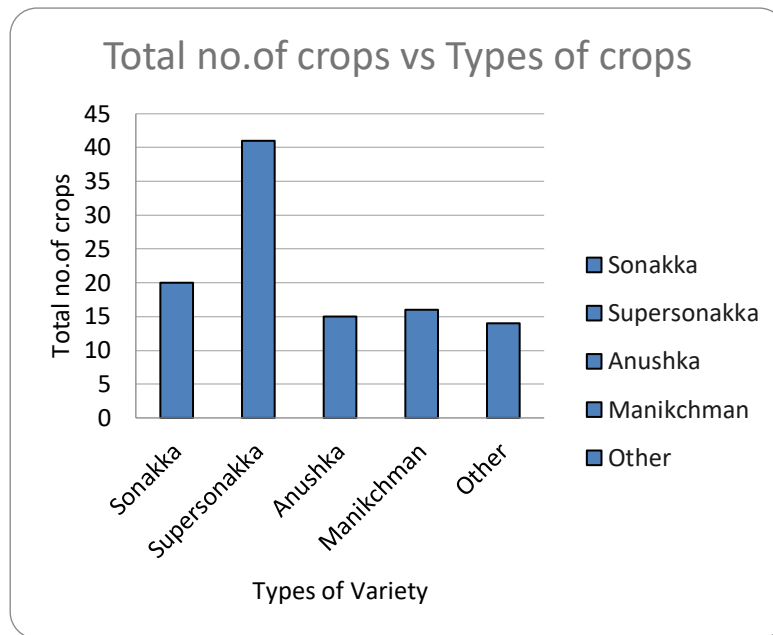
Conclusion:

If the value of r is less than probable error then the correlation is not significant. If the value of r is greater than six times the probable error then the correlation is significant.

STATISTICAL ANALYSIS

1) To study the distribution of variety of grapes:

Variety	Total
Sonakka	20
Supersonaka	41
Anushka	15
Manikchman	16
Other	14

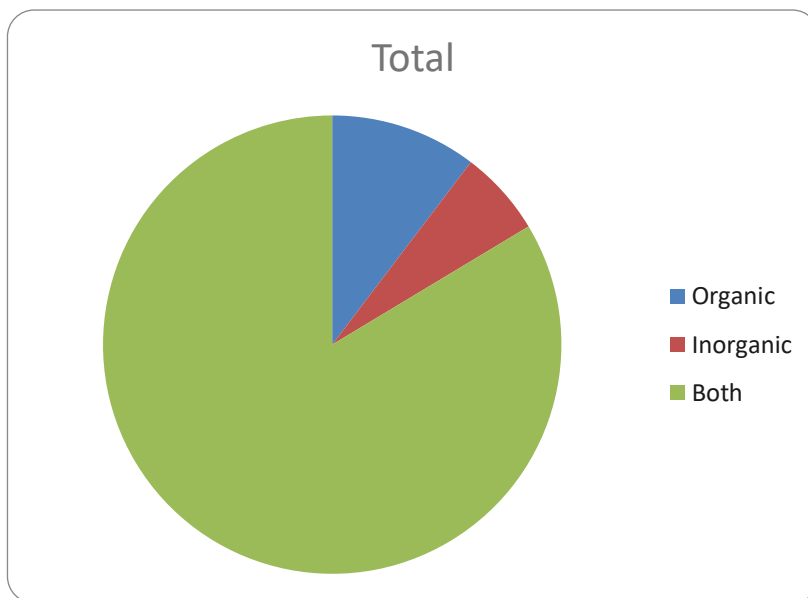


Conclusion:

From the above graph diagram we conclude that most farmers use Supersonaka variety of grapes as compare with other varieties.

2) To Study Distribution Of Fertilizers :

Fertilizers	Total
Organic	12
Inorganic	7
Both	97



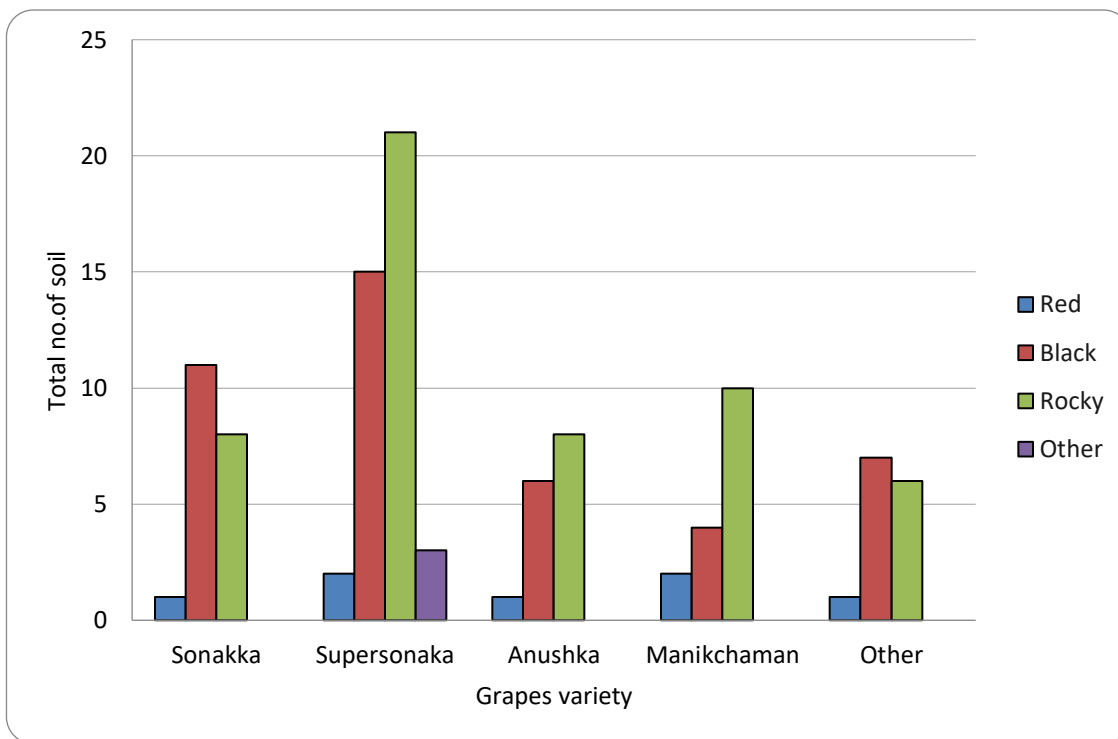
Conclusion:

From the above Pie Diagram , it is clear that most farmers uses the both fertilizers for more production of grapes. i.e. Organic as well as inorganic fertilizers play important role in more production of grapes.

Most of the farmers used both the fertilizers but use of organic fertilizers is greater than the inorganic fertilizers.

3) To study Distribution of grapes variety with soil type :

	Red	Black	Rocky	Other
Sonakka	1	11	8	0
Supersonaka	2	15	21	3
Anushka	1	6	8	0
Manikchaman	2	4	10	0
Other	1	7	6	0

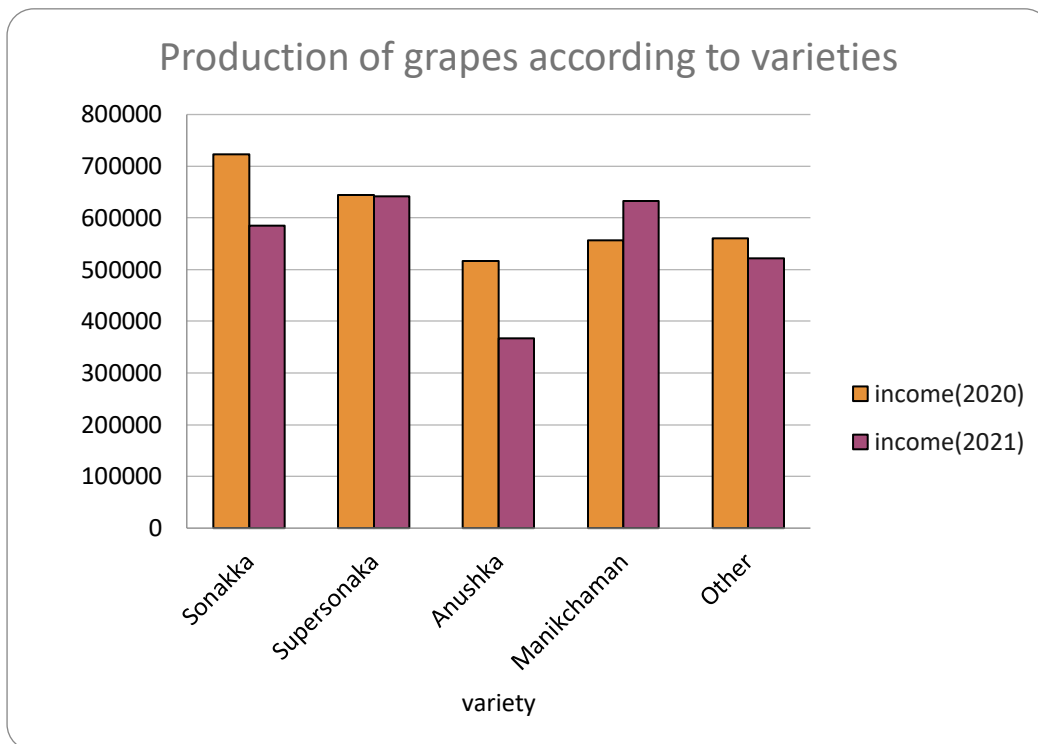


Conclusion:

From the above graph we conclude that cultivation of Supersonaka is greater in Rocky Soil compare with other soil.

4) To Study which variety of grapes give more income:

variety	Average income(2020)	Average income(2021)
Sonakka	723500	584750
Supersonaka	644268.2927	641268.2927
Anushka	516666.6667	366666.6667
Manikchaman	556250	632812.5
Other	560714.2857	521428.5714



Conclusion:

From the graph , we see that variety of Sonakka as well as Supersonaka give more income to the farmers. But Supersonaka gave most income in both 2020 & 2021 years.

5)To study independence between soil type and variety of grapes:

Here we check the soil type for the different varieties of grapes are Independent or not.

Attributes:A-variety of grapes &B-soil type

Here, A1-Sonakks, A2-Super sonakka, A3-Anushka, A4-Manikchaman

A5-Other

Here, we want to test the hypothesis that,

H0:Soil type and variety of grapes are independent

H1:Soil type and variety of grapes are not independent

Observation Table 1:

	Red	Black	Rocky	Other	Total
Sonakka	0	12	8	0	20
Super sonakka	2	15	21	3	41
Anushka	4	5	6	0	15
Manikchaman	2	4	10	0	16
Other	0	7	6	0	13
Total	8	43	51	3	105

Observation Table 2:

Sr.No	Oij	Eij	Oij ² /Eij
1	0	1.523809524	0
2	12	8.19047619	17.58139535
3	8	9.714285714	6.588235294
4	0	0.5714285714	0
5	2	3.123809524	1.280487805
6	15	16.79047619	13.40045377
7	21	19.91428571	22.14490674
8	3	1.171428571	7.682926829
9	4	1.142857143	14
10	5	6.142857143	4.069767442
11	6	7.285714286	4.941176471
12	0	0.4285714286	0
13	2	1.219047619	3.28125
14	4	6.552380952	2.441860465
15	10	7.771428571	12.86764706
16	0	0.4571428571	0
17	0	0.9904761905	0
18	7	5.323809524	9.203935599
19	6	6.314285714	5.701357466
20	0	0.3714285714	0
		Total	125.1854003

Now, Calculated X^2 is given by,

$$X^2 = \sum \sum (O_{ij}^2 / E_{ij}) - N = \underline{20.1854003}$$

The table value of X^2 with $(m-1)(n-1)$ d.f. at 5% l.o.s. is,

$$X^2(12, 0.05) = 21.026$$

$$P - \text{value} = 0.10 > \alpha = 0.05$$

Result :

Here,

Tabulated value of chi-square is $21.026 >$ Calculated value chi-square

Hence we Accept H_0 .

Conclusion :

Hence we conclude that soil type and variety of grapes are independent.

6) To study the difference between production of two years:

In order to check the difference between production of two years, we use Paired t-test.

Let X_1, X_2, \dots, X_n be a random sample of size 105 from production Of 2020 of farmers.

Let Y_1, Y_2, \dots, Y_n be a random sample of size 105 from production Of 2021 of farmers.

Here, we want to test the hypothesis that,

H_0 : There is no significant difference between production of two years of Farmers .i.e. $\mu_1 = \mu_2$

V/s

H_1 : The production of farmers in 2020 is greater than production of 2021

i.e. $\mu_1 > \mu_2$

Under H_0 . The test statistic is,

$$T = \frac{\bar{d}}{(s/\sqrt{n})} \sim t_{n-1}$$

$$\text{Where, } \bar{d} = \frac{\sum d_i}{n} \text{ \& } S^2 = \frac{1}{n-1} \left(\sum (d_i - \bar{d})^2 \right)$$

$$\text{Here, } n=105, \quad S^2 = 7964039179, \quad S = 282206.2901$$

$$\text{\& } \bar{d} = 21361.90476, \quad \text{Then,}$$

$$T(\text{cal}) = 0.775653818$$

$$T(\text{tab}) = 1.664$$

Result: Here, $t(\text{cal}) < t(\text{tab})$,

Hence, we accept H_0 .

We conclude that, production of farmers in 2020 is greater than production Of 2021.

7) To compare the proportion of farmers using organic and inorganic fertilizers:

In a data, 12 farmers out of 105 farmers are found to be users of organic fertilizers.

Now, we have to discuss whether this information supports the view that majority Of farmers are users inorganic fertilizers.

$$P=0.1142857143$$

$$Q=1-p=0.8857142857$$

$$S.E = \sqrt{pq/n} = 0.03104903799$$

Hence, the confidence limits for p, the proportion of farmers used organic fertilizers

At 95% confidence level is,

$$0.1142857143 \pm 0.0310493799 * 1.96 = 0.0534289297 \text{ to } 0.1751454989$$

Conclusion:

The view that majority of farmers used inorganic fertilizers is equivalent to the View that majority of farmers used organic fertilizers which amounts to situation That the proportion p of farmers used organic fertilizers should always be less than 0.5 since, the confidence limits for p include values more than 0.5 in this case , the The given information does not support the view that majority of farmers used Inorganic fertilizers at 95% confidence level.

8) To study the correlation between expense of fertilizers and income of grapes:

We have to study the correlation between expense of fertilizers and income of Grapes. For this we use Karl Pearson's coefficient of correlation.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Here, we have

\bar{x}	112742.8571
\bar{y}	1168447.619
$\sum(x_i - \bar{x})$	-112743
$\sum(y_i - \bar{y})$	-1168447.614
$\sum(x_i - \bar{x})^2$	260521008970
$\sum(y_i - \bar{y})^2$	26865395800252
$\sum(x_i - \bar{x})(y_i - \bar{y})$	$1.317342894 \times 10^{11}$

Hence, $r = 0.04979443803$

Probable error:

reliability of coefficient of correlation is determined by calculating the

Probable error & is given by,

$$\text{Probable error} = 0.6745 \cdot (1 - r^2) / n$$

Here we have, $r = 0.04978443803$

Hence,

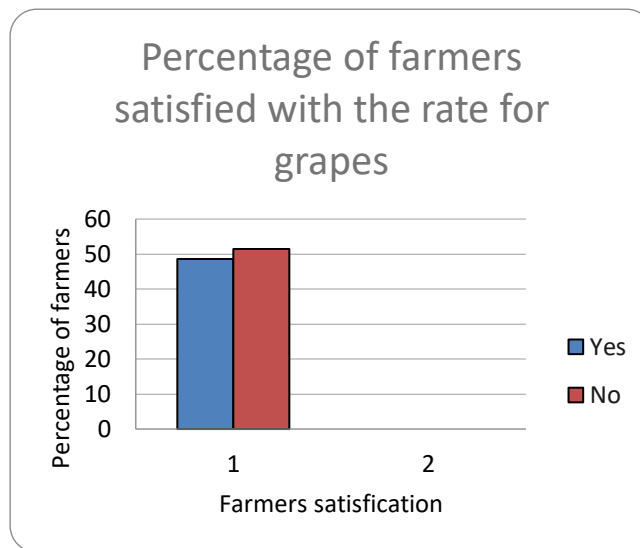
$$P. E = 0.006407881778$$

Conclusion:

Coefficient of correlation is greater than probable error hence the Correlation is significant.

9) To find percentage of farmers are satisfied with the rate for grapes:

Farmers satisfied	Yes	No	Total
percentage	48.54	51.46	100



Conclusion:

From the above graph we see that 48.54% farmers are satisfied with the rate getting for grapes and 51.46% farmers are not satisfied with the rate getting for grapes.

Overall Conclusions

1. Both the Organic as well as inorganic fertilizers play important role in production of grapes as most of the farmers used both the Fertilizers.
2. Types of soil and variety of grapes are independent.
3. Production of farmers in 2020 is greater than the production of 2021.
4. Supersonaka is the variety of grapes which give more income for the farmers.
5. There is a positive correlation between expenses on Fertilizers and income of grapes & it is significant.
6. Most of the farmers are not satisfied with the rate getting for grapes and 48% farmers are satisfied with the rate getting for grapes.
7. The most of the farmers used variety of grapes is Super Sonakka.

LIMITATIONS

- 1) We have to limit the scope of this research according to the result we obtain from the questionnaires.
- 2) This survey is limited to P.V.P Mahavidyalaya Kavathe Mahankal.
- 3) The time Constraints was one of the major limitations.

SCOPE

- 1) A well planned Study can be taken including some relevant characteristics.
- 2) It can be extended for greater region for better estimates.

BIBLIOGRAPHY:

REFERRED BOOKS:

Sr. No.	Book Name	Authors
1	Basic Statistics	B. L. Agarwal
2.	Fundamentals of statistics.	A.C.Gupta & V.K.Kapoor

Website and research engines:

1. Wikipedia
2. Google

QUESTIONNAIRES

- 1) द्राक्ष लागवडी खालील एकूण किती क्षेत्र आहे.
- 2) तुम्ही कोणत्या पद्धतीने शेती करता?
 - a) पारंपरिक
 - b) आधुनिक
- 3) तुमच्या बागेसाठी लागणारी जमीन कोणत्या प्रकारची आहे?
 - a) काळी b) खडकाळ C) तांबडी d) इतर
- 4) बागेला पाणीपुरवठा कसा करता?
 - a) ठिबक सिंचन b) भुपाठ
- 5) तुम्ही कोणत्या जातीचे पीक घेता?
 - a) सोनाक्का b) सुपरसोनक्क
 - c) अनुष्का D) मनिकचमन e) इतर
- 6) बागेवरती कधी रोगाचा प्रभाव झाला आहे का?
 - a) हो b) नाही
- 7) असेल तर कोणत्या रोगाचा प्रभाव झाला आहे?
 - a) थिप्स b) बुरशी c) डाऊनी d) इतर
- 8) बागेसाठी कोणत्या प्रकारची खते वापरता?
 - a) सेंद्रीय खत b) रासायनिक खत c) दोन्ही

- 9) तुम्ही कोणत्या पद्धतीचे सेंद्रिय खत वापरता?
a) गांदुळखत b) शेणखत c) कॉबडीखत
d) स्लरी e) इतर
- 10) तुम्ही कोणत्या पद्धतीचे रासायनिक खत वापरता?
a) SOP b) NPK c) Urea
d) Ammonium sulphate e) इतर
- 11) खतांच्या वापरासाठी कोणाचं सल्ला घेता?
a) अनुभवी शेतकरी b) कृषितज्ञ
c) इंटरनेट/सोशलमीडिया d) इतर
- 12) बागेवरती कोणत्या खतांचा चांगला परिणाम जाणवतो?
a) सेंद्रिय b) रासायनिक c) दोन्ही
- 13) द्राक्ष बागेवरती हवामानाचा परिणाम जाणवतो का?
a) होय b) नाही
- 14) सेंद्रिय खतांचा वार्षिक सरासरी खर्च किती येतो?
- 15) रासायनिक खतांचा वार्षिक सरासरी खर्च किती येतो?
- 16) प्रतीएकर किती उत्पन्न मिळते?
a) 2020-
b) 2021-
- 17) खर्चच्या तुलनेत अपेक्षेप्रमाणे उत्पन्न मिळते का?

a) होय b) नाही

18) जर वर्षी असणारा द्राक्षेचा दर तुम्हाला योग्य वाटतो का?

a) होय b) नाही

19) जर वर्षी असणारा द्राक्षेच्या दराबद्दल शासनाला काय सांगू इच्छित का?

a) होय b) नाही

20) द्राक्ष शेती फायदेशीर वाटते का?

a) होय b) नाही

A Project Report on
“STATISTICAL ANALYSIS OF EFFECT OF FERTILIZERS AND
SOIL TYPE ON SUGARCANE”

Submitted to

Padmabhushan Vasantrodada Patil Mahavidyalaya, Kavathe mahankal.

Department of Statistics

By

Miss. Babar Aishwarya Sanjay

Miss. Patil Bhagyashri Harishchandra

Miss. Chavan Sneha Shrikant

Miss. Patil Vaishnavi Vilas

Under the Guidance of

Prof.(Dr.) M.K.Patil (Head of department)

Assoc.Prof.V.V.Koshti

Mr.A.P.Patil

Miss.S.S.Bhosale

Miss.S.S.Patil

Through

The Principal,

Dr.M.K.Patil

Padmabhushan Vasantrodada Patil Mahavidyalaya, Kavathe mahankal

For The Academic Year 2020-2021
DEPARTMENT OF STATISTICS
PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA,
KAVATHE MAHANKAL

CERTIFICATE

This is to certify that the following students are working under
My guidance. They have satisfactorily carried out and completed the project work
For the B.Sc. Part III. They have presently submitted their project entitled.

This work is genuine and the project or the part of it has not been
Submitted for the award of any other degree.

**“STATISTICAL ANALYSIS OF EFFECT OF FERTILIZERS AND SOIL
TYPE ON SUGARCANE”**

- 1. Miss. Babar Aishwarya Sanjay**
- 2. Miss. Patil Bhagyashri Harishchandra**
- 3. Miss. Chavan Sneha Shrikant**
- 4. Miss. Patil Vaishnavi Vilas**

Teacher in charge

Examiner

Head of Department of
statistics

DECLARATION

We hereby declare that the project entitled **“STATISTICAL ANALYSIS OF EFFECT OF FERTILIZERS AND SOIL TYPE ON SUGARCANE”** is the original work carried out by us at the Department of statistics. This project or part of it has not been submitted previously for any other degree of any other purpose.

Place: Kavathe mahankal

Date:

Sr. No.	Roll No.	Names	Exam seat no.	Sign
1		Babar Aishwarya Sanjay		
2		Patil Bhagyashri Harishchandra		
3		Chavan Sneha Shrikant		
4		Patil Vaishnavi Vilas		

ACKNOWLEDGEMENT

We express our hearty thanks and deepest gratitude goes to Dr. M.K.Patil Sir and Mr.V.V.Koshti Sir, who provided valuable guidance and co-operation during our project work. We are rather infused by the king's guidance of Mr.A.P.Patil , Miss.S.S.Bhosale, Miss. S.S.Patil, Who put us through their encouragement, kind support, and constant inspiration during our project work. Further we are thankful to all our college students and farmers who gave us necessary information to carry this project.

Finally we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.

INDEX

SR. NO.	NAME	PAGE NO.
1	Introduction	6
2	Objectives	7
3	Source of data collection and sampling methods	8
4	Statistical tools used :	9-11
	a) Graphical representation	
	b) Estimation of parameters by confidence interval	
	c) Chi-square test for independence of attributes	
	d) Correlation	
5	Statistical analysis	12-20
6	Overall conclusion	22
7	Limitations and scope	23
8	Bibliography	24-25
9	Questionnaires	26-28

INTRODUCTION :

The title of the project is “Statistical analysis of the effect of fertilizers and soil type on sugarcane yield”. By this data and using some statistical tools for analyzing that data. In this project we reported here some results related to the objectives of the survey. The main purpose of the project is to know the role of the most important factor for production of sugarcane yield and fertilizers. India is an agricultural country. 16.1% of the economy is dependent on agriculture (by the reference of the internet). In Maharashtra and Karnataka state Sangli and Belgaum district are famous for sugarcane production. In Sangli district most production of sugarcane is in Kavathe Mahankal Tehsil. In Kavathe Mahankal Tehsil 60% yield of Sugarcane and occupation of most farmers are sugarcane. In Sugarcane production the most affected factor are fertilizers. Therefore we decided to study the role of fertilizers in the yield of Sugarcane production. We have collected data in Jayagavhan, Kuchi, Dhulagaon, Irali in Kavathe Mahankal Tehsil and Khilegaon-Pandegaon in Athani Tahsil. (We choose those villages because most of the people in that village grow sugarcane) In these four villages data was collected i.e. information of yield of Sugarcane production in four different villages. The information is age, education, types of Sugarcane, soil type, fertilizers types, income per acre, loss per acre and many others. We collected data by direct personal investigation or interview/(Google forms). We proceed to study how the observations are classified and a frequency distribution is formed.

Nowadays fertilizer has become essential to modern agriculture to feed the growing population. Chemical fertilizers are used extensively in modern agriculture, in order to improve crop yield. Urea is the most popular and widely used dry N fertilizer. Increased crop production largely relies on the type of fertilizers used to supplement essential nutrients for plants. Fertilizer application is required to replace crop land nutrients that have been consumed by previous plant growth with the ultimate goal of maximizing productivity and economic returns. Nowadays, there is increased emphasis on the impact on the soil environment due to continuous use of chemical fertilizers.

OBJECTIVES

- ❖ To know the best season for sugarcane planting.
- ❖ To know the role of different types of fertilizers for Sugarcane production.
- ❖ To find average quantity of which fertilizer used per acre by the farmers for effective production of Sugarcane.
- ❖ To know which diseases are most prevalent in sugarcane crops.
- ❖ To compare the proportion of farmers which use organic and inorganic fertilizers on Sugarcane production.
- ❖ To study independence among soil types and varieties used for cultivation.
- ❖ To study the correlation between expense on fertilizers and income of sugarcane.

SOURCE OF DATA COLLECTION AND SAMPLING METHOD

Sampling is a part of our day to day life which we use advertently or inadvertently. For instance, a housewife takes one or two grains of rice from the cooking pan and decides whether it is cooked or not, a pathologist takes a few drops of blood and tests for any change in blood of the whole body. In all these situations, sampling is inevitable and gives satisfactory results.

As the main aim of our project **“STATISTICAL ANALYSIS OF EFFECT OF FERTILIZERS AND SOIL TYPE ON SUGARCANE”** we consider different varieties of Sugarcanes, types of soil, types of fertilizers etc. So we have selected some villages as Jaygavhan, Kuchi, Dhulagaon, Irali in Kavathe Mahankal Tehsil and Khilegaon-Pandegaon in Athani Tahsil, by convenient sampling method, which provide us necessary and sufficient data within less time and more accuracy. Thus the data is collected by filling questionnaires from the farmers by the interview method / Google forms (because of these Corona situation)

Statistical Tools

1] GRAPHICAL REPRESENTATION:

a) **HISTOGRAM:** A histogram is a graphical representation of the distribution of numerical data. It is an estimate of the probability distribution of a continuous variable (quantitative variable) and was first introduced by Karl Pearson.

b) **PIE DIAGRAM:** A pie diagram or pie chart is a circular statistical graphic, which is divided into slices to illustrate numerical proportion. In a pie chart the arc length of each slice is proportional to the quantity it represents.

c) **MULTIPLE BAR DIAGRAM:** In multiple bar diagrams two or more sets of interrelated data are represented. The technique of making a simple bar chart is used to draw this diagram but the difference is that we use different shades, colors, or dots to distinguish between different phenomena.

d) **LINE CHART:** A line graph is a graphical display of information that changes continuously over time. Within a line graph, there are various data points connected together by straight lines that represent values represented by the data points. A line chart is used to represent data over a continuous time span. It is generally used to show the trend of a measure (or a variable) over a time. Data values are plotted as points that are connected using line segments.

e) **BAR-CHART :** A bar chart or bar graph is a chart or a graph that present categorical data with rectangular bars with heights or length of proportional to the values that it represents. The bars can be plotted vertically or horizontally. A vertical bar chart is sometimes called column chart

2] ESTIMATING PARAMETER VALUES BY USING CONFIDENCE INTERVAL:

The assumption is justifiable only if the size of sample n is large and neither p nor q is very small. If n is large the value of either p or q is not very small, then the difference between them is not much and the distribution would be very much like a normal one. We have the value of pq is can never exceed $\frac{1}{2} * \frac{1}{2}$ or $\frac{1}{4}$. As such, if we do not know the value of p in the universe and if the value of p in the sample is very small or very large (so that the value of q is very large or very small), then, instead taking the sample values of p and q , the values of pq may be taken as $\frac{1}{4}$ which is the maximum value.

Then the standard error is,

$$\text{S.E.} = \sqrt{\frac{pq}{n}}$$

And the confidence interval at $\alpha\%$ l.o.s. is given as, $p \pm q * Z_{\alpha}$

3] CHI-SQUARE TEST FOR INDEPENDENCE OF ATTRIBUTES:

The hypothesis that one factor independent of another or not, can be tested by the chi-square test. The hypothesis that,

H_0 : Two attributes are independent of each other.

H_1 : Two attributes are dependent.

Under H_0 , the test statistics is,

$$X^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{O_{ij}}{E_{ij}} - N \sim X^2_{(m-1)(n-1)}$$

Statistic X^2 has $(m-1)(n-1)$ d.f. in formula (a), E_{ij} is the expected frequency corresponding to $(i, j)^{\text{th}}$ cell of observed frequency O_{ij} .

Under the null hypothesis, the expected frequency is,

$$E_{ij} = \frac{\text{ith row total} * \text{jth column total}}{\text{sample size}} = \frac{R_i * C_j}{n}$$

Once the expected frequencies are calculated, the value of chi - square is obtained with the help of formula (a). To make a design about H_0 , the calculated value of chi - square compared with the table attributes are not independent. Otherwise accept H_0 .

4] Correlation:

Correlation means, “The study of existence and the magnitude and direction of variation between two or more variables”.

Karl Pearson’s coefficient of correlation:

The most commonly used method which gives a mathematical expression for correlation is the one suggested by Karl Pearson.

Here, $\sigma_x^2 = \frac{1}{N} \sum (xi - \bar{x})^2$ gives a measure of variation in x

& $\sigma_y^2 = \frac{1}{N} \sum (yi - \bar{y})^2$ gives a measure of variation in y

Also, $\frac{1}{N} \sum (xi - \bar{x}) \sum (yi - \bar{y})$ gives a measure of simultaneous variation in x & y. Then the correlation coefficient is,

$$r = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{\sqrt{\sum (xi - \bar{x})^2 \sum (yi - \bar{y})^2}}$$

The probable error and the coefficient of correlation:

The reliability of the coefficient of correlation is determined by calculating the probable error. If n is the size of the sample the probable error is given by,

$$\text{Probable error} = 0.6745 * \frac{(1-r^2)}{n}$$

Where, r is the correlation coefficient.

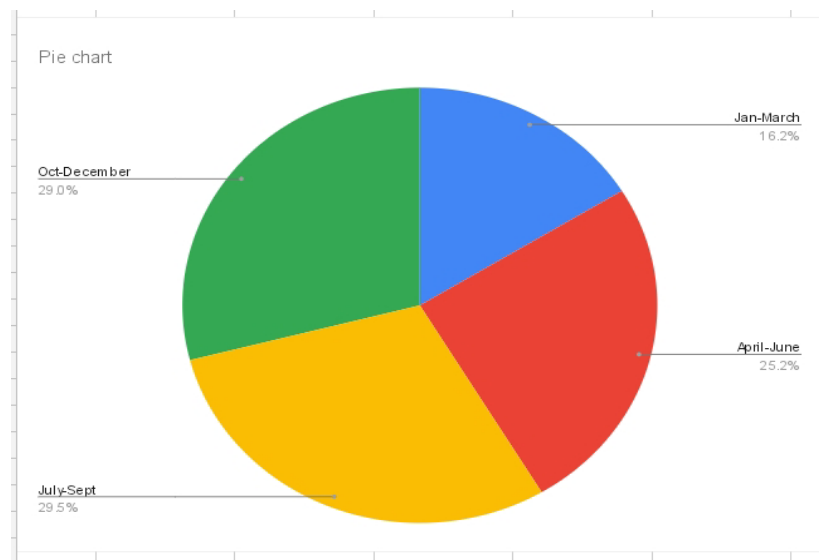
Result If the value of r is less than probable error then the correlation is not significant. If the value of r is greater than six times the probable error then the correlation is significant.

STATISTICAL ANALYSIS

1] To know the best season for sugarcane planting :(Depending on the season in which most farmers harvest)

Season	Toral(Farmers)
Jan-March	34
April-June	53
July-Sept	62
Oct-Decembe	61

Percentage of best season for planting of Sugarcane

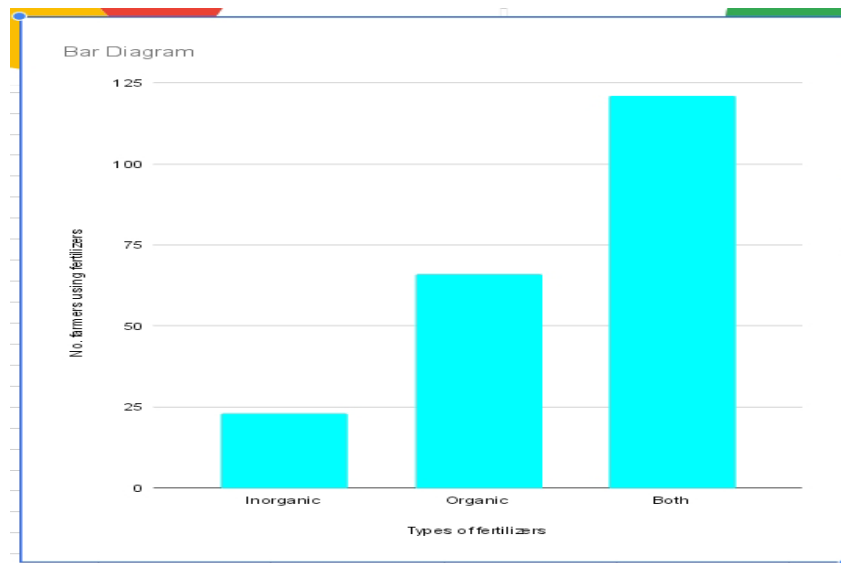


Conclusion: From the above Pie-diagram we observed that , there is higher percentage in July-sept season i.e.29.5%(depending on the season in which most farmers harvest)

2] To know the role of fertilizers in Sugarcane production:

Fertilizers Types	Total(fertilizer type used by farmers)
Inorganic	23
Organic	66
Both	121

Average quantity of fertilizer type used by farmers



Conclusion:

From the above Bar-diagram it is clear that, about 121 farmers use both i.e. organic & inorganic fertilizers for more production of Sugarcane yield .

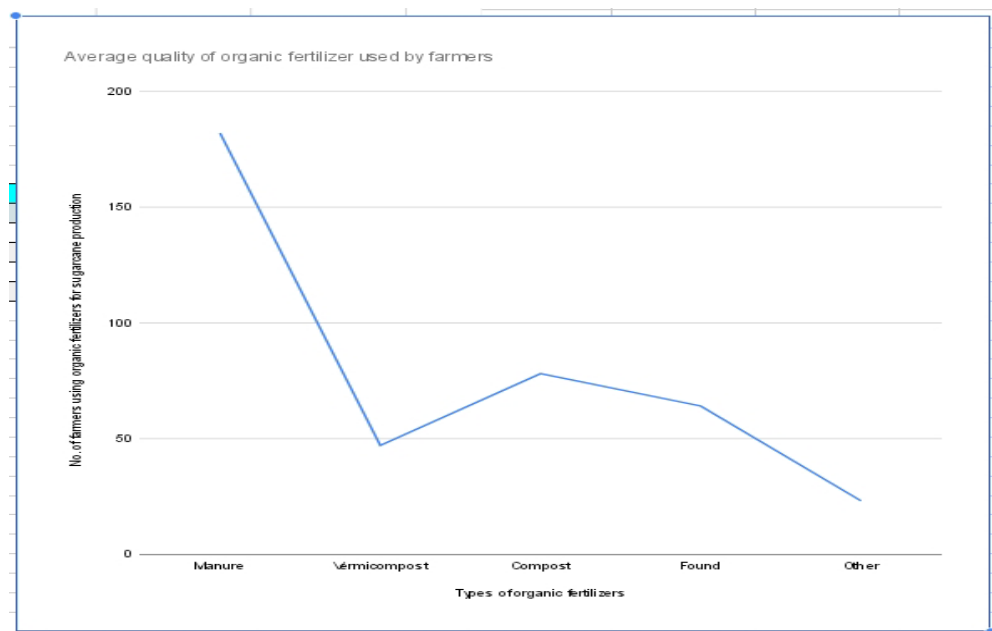
Most of the farmers used both fertilizers but use of organic fertilizers is greater than inorganic fertilizers.

3] To find the average quantity of fertilizers used per acre by the farmers for effective production of Sugarcanes yield:(ton per acre)

i) For organic fertilizers:

Organic fertilizer	No.of farmers
Manure	182
Vermicompost	47
Compost	78
Found	64
Other	23

Average quantity of organic fertilizers used by farmers(ton per acre

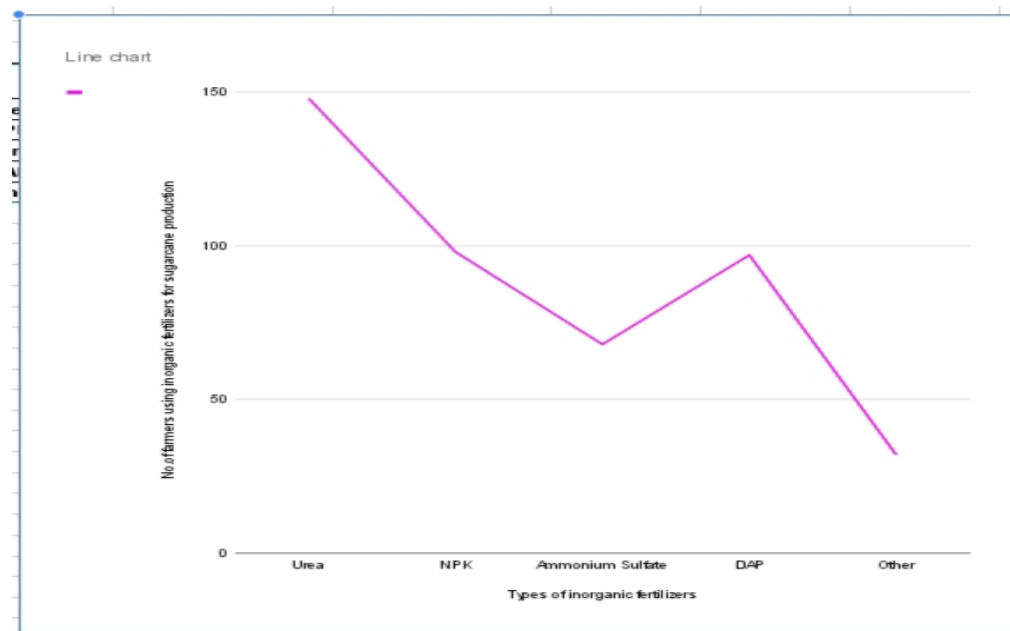


Conclusion: From the above line-chart, we see that above the 150 farmers used manure fertilizer, above 60 farmers used compost and found , less than 60 farmers used vermicompost and other.

ii)For inorganic fertilizer:(Kg per acre)

Inorganic fertilizer	No.of farmers
Urea	148
NPK	98
Ammonium sulfate	68
DAP	97
Other	32

Average quantity of inorganic fertilizer used by farmers (kg per acre)



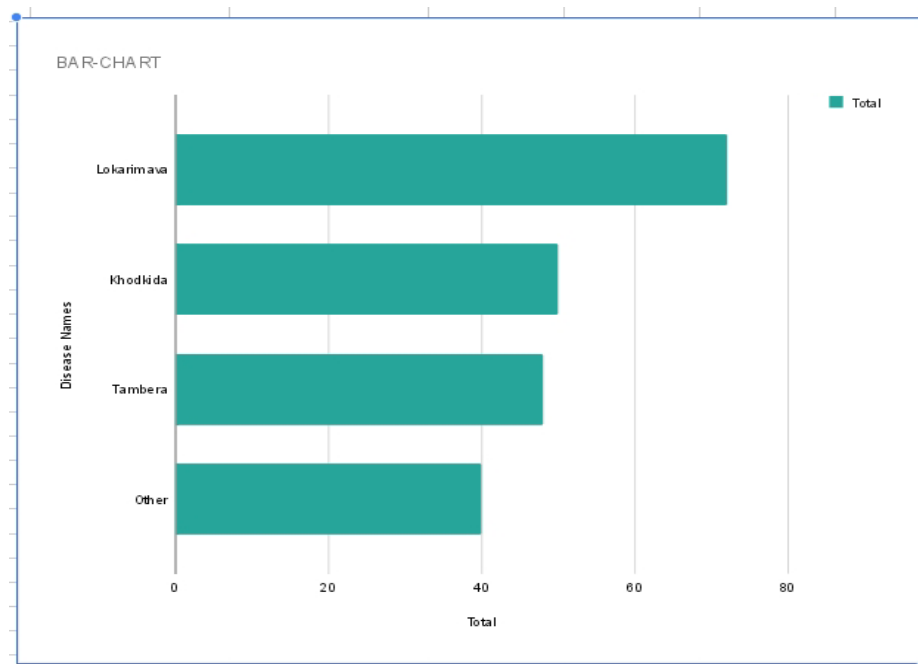
Conclusion:

From the above line chart , we see that above 100 farmers are using urea, above 50 and below 100 farmers are using NPK, Ammonium sulfate, DAP and below 50 farmers used other inorganic fertilizer.

4]To know which diseases are most prevalent in sugarcane crops:

Disease Names	Total
Lokarimava	72
Khodkida.	50
Tambera	48
Other	40

For finding most prevalent disease in Sugarcane crop



Conclusion: From the above Bar- chart we conclude that disease of Lokari Mava is more Prevalent in Sugarcane Crop.

5] To compare the proportion of farmers using organic and inorganic fertilizers:

In the data, 66 farmers out of 210 farmers are found to be users of organic fertilizers. Now, we have to discuss whether this information supports the view that the majority of farmers are users of inorganic fertilizers.

$$p = 0.3142$$

$$q = 1 - 0.3142 = 0.6858$$

$$\text{S.E.} = \sqrt{\frac{pq}{n}} = 0.03203$$

Hence, the confidence limits for p , the proportion of farmers used organic fertilizers at 95% confidence level is,

$$0.31428571 \pm 0.032031 * 1.96 = 0.25150 \text{ to } 0.3770$$

Result:

The view that majority of farmers used inorganic fertilizers is equivalent to the view that majority of farmers used organic fertilizers which amounts to situation that the proportion p of farmers used organic fertilizers should always be less than 0.5 since, the confidence limits for p include values more than 0.5 in this case, the given information does not support the view that majority of farmers used inorganic fertilizers at 95% confidence level.

Conclusion:

There is a majority in use of organic fertilizers as compared with inorganic fertilizer.

6] To study independence between soil type & variety of Sugarcane yield :

Here we check the type of soil for the different varieties of Sugarcanes are independent or not.

Attributes: - A-variety of Sugarcanes & B- soil type

Here, A1-Nira86032, A2-265, A3-671, A4-8014

B1-Rocky, B2-Red clay, B3-Alluvial, B4-Black, B5-other

Here, we want to test the hypothesis that,

H_0 : Soil type and variety of sugarcanes are independent.

H_1 : Soil type and variety of Sugarcanes are not independent.

Observation Table 1]:

	Rocky	Red clay	Alluvial	Black	Other	Total
Nira86032	15	16	6	77	0	114
265	9	11	2	18	5	45
671	7	5	4	16	1	33
8014	4	1	5	8	0	18
Total	35	33	17	119	6	210

Observation Table 2]:

Sr.no.	Oij	Eij	Oij ² /Eij
1	15	19	11.8421
2	16	17.9142	14.2903
3	6	9.2285	3.9009
4	77	64.6	91.7801
5	0	3.2571	0
6	9	7.5	10.8
7	11	7.0714	17.1111
8	2	3.6428	1.0980
9	18	25.5	12.7058
10	5	1.2857	19.4446
11	7	5.5	8.9090

12	5	5.1857	4.8209
13	4	2.6714	5.9893
14	16	18.7	13.6898
15	1	0.9428	1.06067
16	4	3	5.3333
17	1	2.8285	0.3535
18	5	1.4571	17.1573
19	8	10.2	6.2745
20	0	0.5142	0
			=276.7406

Now, calculated X^2 is given by,

Here, $m=5$ & $n=4$

$$X^2 = \sum_{i=1}^m \sum_{j=1}^n \frac{O_{ij}}{E_{ij}} - N = 276.7406 - 210 = \underline{66.7406}$$

The table value of X^2 with $(m-1)(n-1)$ d.f. at 5% l.o.s. is,

$$X^2_{(12, 0.05)} = 21.026$$

Result:

Here, calculated $X^2 >$ table X^2 .

Therefore we reject H_0 , we conclude that type of soil and the variety of Sugarcane crops are not independent.

Conclusion: There is a relationship between soil type and varieties of Sugarcane for the production of Sugarcane yield.

7] To study the correlation between expense on fertilizers and income of Sugarcane:

We have to study the correlation between expense on fertilizers and income of Sugarcane . For this we use Karl Pearson's coefficient of correlation,

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Here, we have,

$$\bar{X} = 24525.97613, \quad \bar{Y} = 10094.33160, \quad \sum(X_i - \bar{X}) = 5150454.738$$

$$\sum(Y_i - \bar{Y}) = 2119809.524 \quad \sum(X_i - \bar{X})^2 = 2.6527 \times 10^{13}, \quad \sum(Y_i - \bar{Y})^2 = 4.4936 \times 10^{12}$$

$$\sum(X_i - \bar{X})(Y_i - \bar{Y}) = 1.0918 \times 10^{13}$$

Hence, $\underline{r = 0.97846}$

Probable error:

The reliability of coefficient of correlation is determined by calculating the probable error & it is given by,

$$\text{Probable error} = 0.6745 * \frac{(1-r^2)}{n}$$

Here we have, $r = 0.0690018$

Hence,
 $\underline{\text{P.E.} = 0.0001368}$

Result:

Here, we see that there is a positive correlation between expense on fertilizers and income of Sugarcane yield. There is a linear relationship between expense and income of Sugarcane's yield.

Also, the probable error is greater than coefficient of correlation hence the correlation is significant

Conclusion: As we use more fertilizers, we will have more income on Sugarcane's yield.

OVERALL CONCLUSIONS

- 1] Both the Organic as well as inorganic fertilizers play important role in production of Sugarcane as most of the farmers used both the fertilizers.
- 2] The given data supports the view that the majority of farmers used organic fertilizers.
- 3] Different types of soil and varieties of Sugarcane crops are dependent.
- 4] From this analysis we conclude that the best season for planting Sugarcane crop is in July-Sept and Oct-December (according to farmers)
- 5] The Lokari Mava disease is more prevalent in Sugarcane.
- 6] Most of the farmers used manure fertilizer in organic fertilizer and Urea in inorganic fertilizer.
- 7] There is a positive correlation between expense on fertilizers and income of grapes & and it is significant.

Limitations

1. We have limits on the scope of this research according to the result we obtain from the questionnaires.
2. This survey is limited to P.V.P.college kavathe mahankal.
3. The time constraints was one of the major limitations.
- 4.Our population is not too large , therefore there is no restrictions over resources in terms of money or time

Scope

1. A well planned study can be taken up including some relevant characteristics.
2. It can be extended to a greater region for better estimates.

Bibliography

Referred books:

Sr. No.	Book Name	Authors
1.	Fundamentals of applied statistics	A. C. Gupta & V. K. Kapoor
2.	Basic Statistics	B. L. Agarwal
3.	Bsc.part II STATISTICS (paper VII)	Dr.K.G.Potdar,Dr.M.K.Patil, Dr.A.A.Kalagonda,Dr.A.V.Dorugade

From 1.Google

2.Wikipedia



(a) Tiller red rot



(b) Lamina red rot



(c) Midrib red rot



(d) Stem red rot



Ring Spot of Sugarcane | Pests ...
plantix.net



Eyespot of Sugarcane | Pests &...
plantix.net



Minor disease and ring spot disease

Shikshan Prasarak Sanstha's

Padmabhushan Vasantrodada Patil Mahavidyalaya, Kavathe mahankal.

Department of Statistics

Project name: "Statistical analysis of effect of fertilizers and soil type on sugarcane yield".

Questionnaire:

गाव : **वय :** **शिक्षण:**

१).कोणत्या प्रकारच्या मातीत ऊसाची लागवड तुम्ही करता?

a) खडकाळ b) तांबडी c) गाळाची d) काळी. e) इतर

२).ऊसाची लागवडीखालील एकूण किती जमीन आहे? (एकर)

३).कोणत्या प्रकारचे खत वापरून उसाचे चांगले उत्पन्न मिळते असे तुम्हाला वाटते?

a) रासायनिक. b) सेंद्रिय. c) दोन्ही

4). रासायनिक खताच्या लागवडीखालील मातीवर काही दुष्परिणाम होतो का? तुमचे मत सांगा.

a)होय b) नाही

5).सेंद्रिय खतांमध्ये यापैकी कोणत्या खताचा वापर तुम्ही करता?

a) शेणखत b) गांडूळ c)कंपोस्ट

a) मळी e) इतर

6).रासायनिक खतांमध्ये यापैकी कोणत्या खतांचा वापर तुम्ही करता?

a) Urea. b) NPK. c) Ammonium sulfate

b) DAP. . e) othar

c) **7). तुम्ही यापैकी कोणत्या जातीच्या ऊसाची लागवड केली आहे?**

a) नीरा 86032 b) 265 फुले. c) 671

d) 8014. d) other

8).उसाच्या दोन सरींमधे किती अंतर ठेवता?

a) 3 फूट b) 3.5 फूट. c) 4 फूट d)other

9).तुमच्या ऊसाचे हे कितवे तोड आहे?

a) लागण(पाहिले पीक) b) खोडवा(दुसरे पीक)

c). निडवा(तिसरे पीक). d) इतर

10). प्रति एकर सरासरी कीती टन ऊस जातो?

11). प्रति एकर जमिनीत एका उसाच्या पिकाखाली खताचा कीती वापर करता?(kg)

12). प्रति एकर जमिनीत तुम्हाला खतांचा वार्षिक खर्च कीती येतो?

13). वारंवार एकाच प्रकारच्या मातीत पिक घेतल्याने कोणता दुष्परिणाम जाणवला आहे का?

a) होय. b) नाही

14). ऊसाच्या पिकावर रोगाचा कधी प्रादुर्भाव झाला आहे का?

a) होय. b) नाही

15).असेल तर यापैकी कोणता?

a) लोकरीमावा. b) खोडकीडा.

b) तांबेरा रोग. d) इतर

16) .ऊसाची लागवड कोणत्या सिझन मधे घेता?

a) जाने- मार्च. b) एप्रिल- जून

b) जुलै- सप्टेंबर d)ऑक्टोबर- डीसेंबर

_____ Thank you.....

A
Project Report

On

**“TO STUDY THE TIME SPENT BY STUDENTS ON SOCIAL
MEDIA”**

Submitted to

(Department of Statistics)

Shikshan Prasarak Sanstha's

**PADMABHUSHAN VASANTRAODADA PATIL MAHAVIDYALAYA,
KAVATHE MAHANKAL.**

By

Miss. Shinde Jyoti Subhash.

Miss. Mujawar Simaran Nisar.

Miss. Mali Pooja Sahadev.

Miss. Mali Sonali Vithoba.

Miss. Kore Pratiksha Jayant.

Under the Guidance of

Prof. Dr. Patil M.K. (Head of the dept.)

Asso. Prof. Koshti V.V.

Miss. Bhosale S.S.

Miss. Patil S.S.

Mr. Patil A.P.

Through

The principal

Dr. Shri. Patil M.K.

Padmabhushan Vasantrodada Patil Mahavidyalaya Kavathe Mahankal.

For the academic year 2020-2021

Shikshan Prasarak Sanstha's
**PADMABHUSHAN VASANTRAODADA PATIL
MAHAVIDYALAYA KAVATHE MAHANKAL.**

(Department of Statistics)

CERTIFICATE

This is to certify that the following students are working under my Guidance. They have satisfactorily carried out and completed the project Work for the B. Sc. Part III. They are presently submitted their project entitled.

This work is genuine, original, and the project or the part of it has not has submitted for the award of any other degree.

“TO STUDY THE TIME SPENT BY STUDENTS ON SOCIAL MEDIA”.

Miss. Shinde Jyoti Subhash.

Miss. Mujawar Simaran Nisar.

Miss. Mali Pooja Sahadev.

Miss. Mali Sonali Vithoba.

Miss. Kore Pratiksha Jayant.

Teacher in Charge

Examiner

Head Department of statistic

DECLARATION

We hereby declare that the project entitled **“TO STUDY THE TIME SPENT BY STUDENT ON SOCIAL MEDIA”**.

It is the original work carried out by us at the department of statistics.

This project or part of it has not been submitted previously for any other degree or any other purpose.

Sr. No.	Roll No.	Name	Exam seat No	Signature
1.		Shinde J. S		
2.		Mujawar S. N		
3.		Mali P. S		
4.		Mali S. V		
5.		Kore P. J		

Place: Kavathe Mahankal.

Date:

ACKNOWLEDGEMENT

We express our heartfelt thanks and deepest gratitude goes to Prof Koshti V. V Sir, who made valuable guidance and co-operation during our project work. We are rather inspired by the kind guidance of Prof. Mr.Koshti V. V, Prof. Mr. Patil A.P, Prof. Miss Bhosale S. S, Prof. Miss Patil S.S.

Who put us their encouragement, kind support, and constant inspiration during our project work further we are thankful to all of our college students who gave us the necessary information to carry this project.

Finally, we give sincerest thanks to all our family members for keeping our spirit up and their immense support during every stage of the project work.

INDEX

Sr.No.	Name of topic	Page No.
1.	Introduction.	6
2.	Method of data collection.	7
3.	Objectives.	8
4.	Statistical tools used : A)Graphical representation. B)Chi-square test for independence of two attributes. C)Z test for equality of two population proportion.	9-11
5.	Analysis of data.	12-22
6.	Overall conclusion.	23
7.	Limitations and scope.	24
8.	Bibliography.	25
9.	Questionnaire.	26-27

INTRODUCTION

Social media is most important in human life and also helps to study. We study how much can social media is important in student's life and we also study the good and bad effects of social media.

According to statistics, it is estimated that, in 2020, around 3.6 billion people are using social media worldwide; up from 3.4 billion in 2019. This number is expected to increase to 4.41 billion in 2025.

We study how many students are using social media and how much time is spent, students on social media, and also check which app is more used by students. Students use social media to share information and make connections. On a personal level, social media allows due to communicating with friends and family learn new things. social media is neither inherently bad nor good it all depends on how it is being used.

The power of social media is the ability to connect and share information with anyone on earth or with many people simultaneously.

Social media is organized as a way to interact with family and friends. Social media is an interesting based form of communication. There are many forms of social media. Some of the most popular social media websites with over a hundred million registered users include whatsapp, facebook, telegram, instagram, youtube, hotstar.

METHOD OF DATA COLLECTION

We surveyed Kavathe Mahankal & Jath area and collect the data. We select the convenient sample by the random sampling method. We collect the data through the online questionnaire. 399 samples are the optimum sample size. We collect data from the online method. The data of our project is primary because we collect the data by taking online methods from each student male and female.

OBJECTIVES

Social media are being used in large quantities in the world today, and so we plan to project on this topic.

1. To find out the amount of time spent by students on social media.
2. To study which App more used by students.
3. To check most active times on social media by students.
4. To study age-wise use of social media by students.
5. To study the effects of social networking sites on the education of students.
6. To check whether social networking sites are popular in the male and female gender.

STATISTICAL TOOLS

The following are some statistical tools leads us towards the appropriate results and finding.

A) GRAPHICAL REPRESENTATION :

To represent our data graphically, we use the multiple bar diagram, pie chart method, etc. To construct the appropriable divided bar diagram of given data we used Microsoft Excel.

1.Bar Chart: A bar chart or bar graph is a chart that presents grouped data with rectangular bars with a lengths proportional to the values that they represent. The bars can be plotted vertically or horizontally. A vertical bar is sometimes called a column bar chart.

A bar graph is a chart that uses either horizontal or vertical bars to show comparisons among categories being compared, and the other axis represents a discrete value.

2.Pie Chart: A pie chart displays data, information, and statistics in an easy-to-read pie-slice format with varying slice sizes telling you how much of one data element exists. The bigger the slice the more of that particular data was gathered.

3.Line Chart: A line chart or line plot or line graph or curve chart is a type of chart that displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart common in many fields. It is similar to a scatter plot (except that the measurement points are ordered typically by their x-axis value) and joined with straight line segments. A line chart is often used to visualize a trend in data over intervals of time a time series thus the line is often drawn chronologically. In these cases, they are known as run charts.

B) TEST OF INDEPENDENCE OF ATTRIBUTES:

In 2*2 Contingency Table

The hypothesis, that one factor is independent of each other or not, can be tested by the Chi-square area. The hypothesis,

Ho: Two attributes are independent of each other.

V/s

H1: Two attributes are dependent,

Attribute	B1	B2	Total
A1	a	b	a+b
A2	c	d	c+d
Total	a+c	b+d	N

The above hypothesis can be tested by the statistic,

$$\chi^2 = \frac{N[ad-bc]^2}{((a+b)*(a+c)*(c+d)*(b+d))} = \chi^2_{1 d.f}$$

Conclusion :

If $\chi^2_{(cal)} > \chi^2_{(tab)}$ with 1 d.f , we reject Ho .

This means that the two attributes are not independent. Two attributes are independent of each other.

C) TEST FOR EQUALITY OF TWO POPULATION PROPORTIONS ($P_1=P_2$) :

Let P_1 and P_2 be the proportions of the unit of possessing the given attribute in population. Let a random sample of n_1 units be drawn from a population with the population proportion P_1 , in which the number of units possessing the given attribute is X_1 and a random sample of n_2 units be drawn from a population P_2 , in which the number of units possessing the given attribute is X_2 . the corresponding standard normal variate is,

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\left(\frac{P_1 Q_1}{n_1} + \frac{P_2 Q_2}{n_2}\right)}} \sim N(0,1)$$

Here we want to test the null hypothesis, $H_0: P_1 = P_2$

i.e., there is no significant difference between population proportions.

V/s. $H_1: P_1 \neq P_2$

Under the null hypothesis,

$$Z = \frac{(p_{11} - p_{21})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Let α be the level of significance and Z_α being the critical value.

Conclusion :

I. Suppose $H_1: P_1 \neq P_2$ if compute $|Z| > \text{tab. } Z_{\alpha/2}$ we reject H_0 at $\alpha\%$ l.o.s and that there is a significant difference between population proportions.

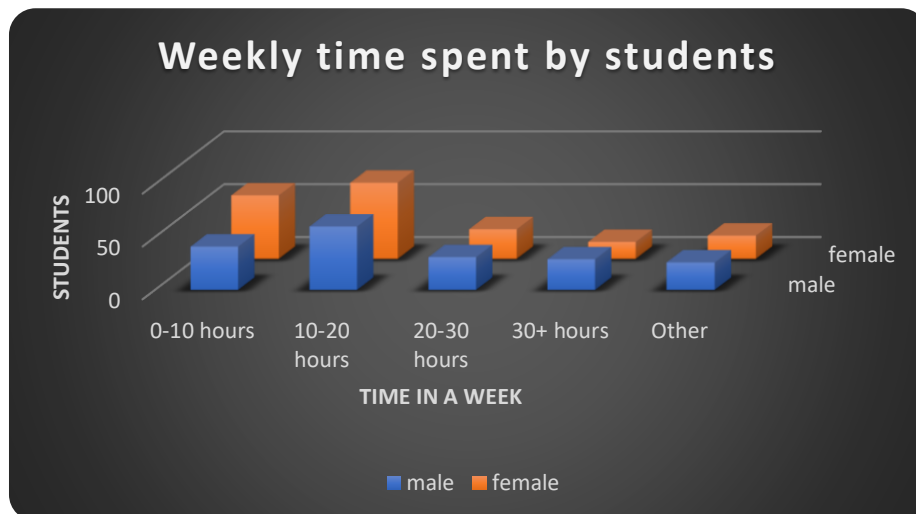
II. Suppose $H_1: P_1 > P_2$ if computed $Z > \text{tab. } Z_\alpha$ we reject H_0 at $\alpha\%$ l.o.s. otherwise, accept.

III. Suppose $H_1: P_1 < P_2$, if computed $Z < \text{tab. } -Z_\alpha$ we reject H_0 at $\alpha\%$ l.o.s. otherwise, accept.

STATISTICAL ANALYSIS

1. To find out the amount of time spent by students on social media.

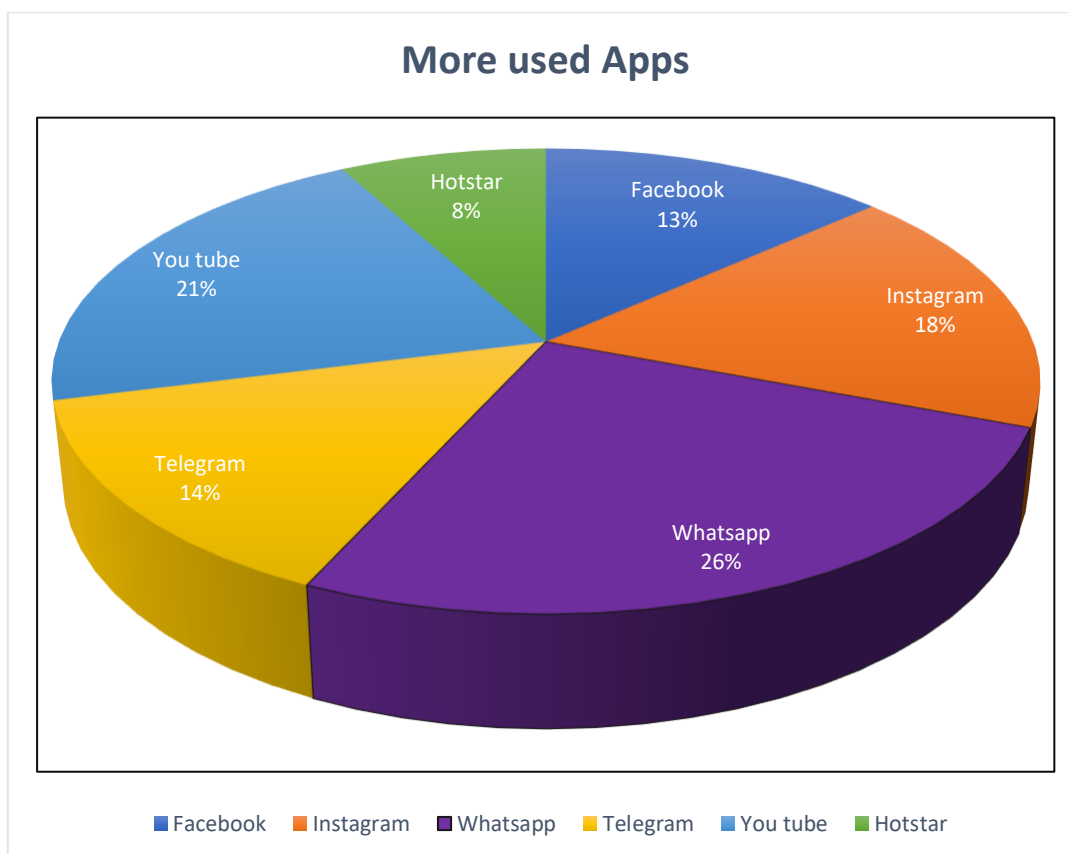
Hours in a week	Male	Female
0-10 hours	41	60
10-20 hours	60	72
20-30 hours	31	28
30+ hours	29	16
Other	26	22



Conclusion: From the above graph, we observe that the female & male students same time spent 20-30 hours a week on social media.

2. To study which App more used by students.

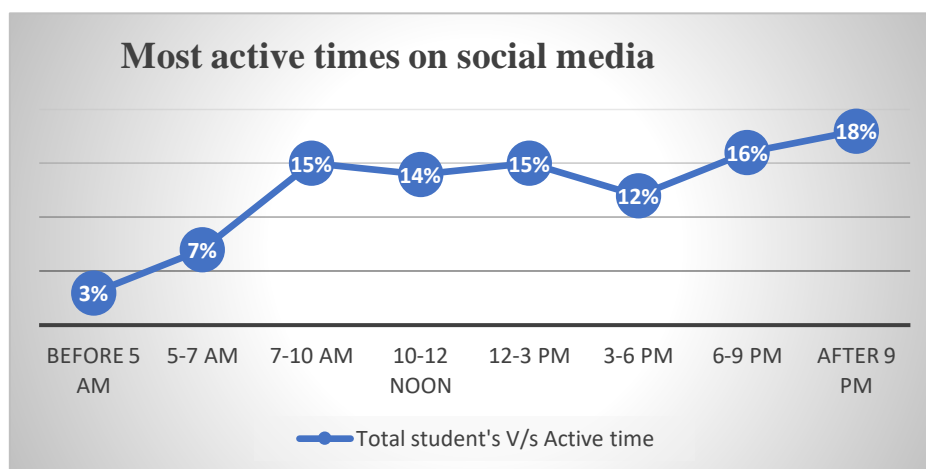
Social networking sites	Total no of students	Total no. of students in %
Facebook	183	13%
Instagram	238	18%
Whatsapp	355	26%
Telegram	194	14%
Youtube	287	21%
Hotstar	109	8%



Conclusion: From the above graph, we observed that most of the 26.0% of students use Whatsapp.

3.To check most active times on social media by students.

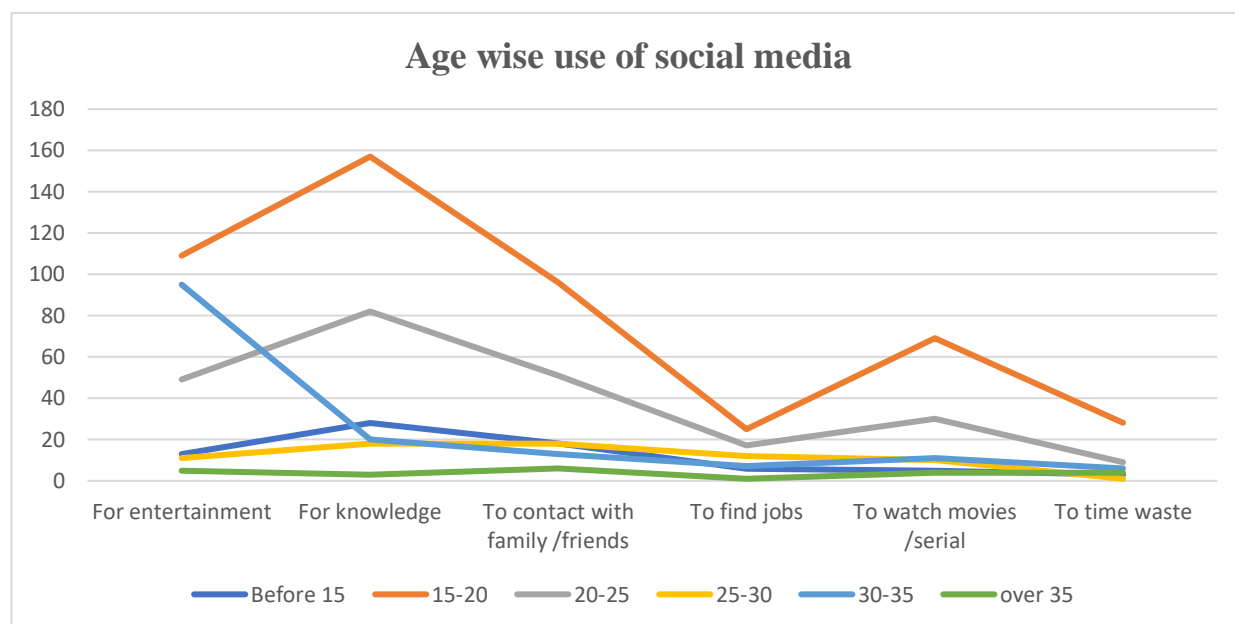
Active time	Total students	Total students in %
Before 5 am	23	3%
5-7 am	51	7%
7-10 am	101	15%
10-12 am	98	14%
12-3 am	101	15%
3-6 am	80	12%
6-9 am	110	16%
After 9 pm	126	18%



Conclusion: From the above graph, we observe that 18.3% of students are active on social media after 9 pm.

4.To study age-wise use of social media by students.

Use of social media	For entertainment	For knowledge	To contact with family and friends	To find jobs	To watch movies/serial	To time waste
Before 15	13	28	18	6	5	3
15-20	109	157	96	25	69	28
20-25	49	82	51	17	30	9
25-30	11	18	18	12	10	1
30-35	95	20	13	7	11	6
Over 35	5	3	6	1	4	4



Conclusion: From the above graph, we observed that the 15-20 age group students most use social media.

5. To study the effects of social networking sites on the education of students.

Attribute :

A: Effect of social networking sites B: Student

Here we denoted, A : Yes α : No

B: Male β : Female

Ho: A and B are independent of each other.

H1: A and B are dependent on each other.

Statistical Table :

Attribute	A	α	Total
B	95	99	194
β	99	106	205
Total	194	205	399

The test statistic: $\chi^2 = \frac{N[ad-bc]^2}{((a+b)*(a+c)*(c+d)*(b+d))} = \chi^2_{1 d.f}$

$$\chi^2_{(cal)} = 0.0182 \quad \chi^2_{(tab)} = 3.8414$$

$$\chi^2_{(cal)} = (0.0182) < \chi^2_{(tab)} = (3.8414)$$

Result: Here, Cal $\chi^2 < \text{tab } \chi^2_{(1,0.05)}$ hence, accept H0 i.e., reject H1.

i.e., A & B are independent to each other.

Conclusion: we concluded that the social networking sites have an effect on students' education.

6.To check whether social networking sites are popular in the male and female gender.

1) Z test for Facebook :

Hypothesis :

Ho: There is no significant difference between Facebook use in males and Facebook use in females. i.e. Ho: $P_{11} = P_{21}$ V/s

H1: There is a significant difference between Facebook use in males and Facebook use in females. i.e. H1: $P_{11} \neq P_{21}$

P11: Proportion of Facebook use in females.

P21: Proportion of Facebook use in males.

Observation Table :

n1	n2	X11(users) in n1	X21(users) in n2	p11	p21
205	194	61	126	0.2975	0.6494

Calculation:

$$P = \frac{n_1 p_{11} + n_2 p_{21}}{n_1 + n_2} = 0.4685$$

$$Q = 1 - P = 0.5314$$

Then calculated Z,

$$\text{Cal } |Z| = \frac{(p_{11} - p_{21})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 7.0404$$

The table value of Z at 5% l.o.s., Tab |Z| = 1.96

Result:

$Z_1 = \text{Cal } |Z| > \text{Tab } |Z|$, hence we reject Ho at 5% l.o.s.

Conclusion: We conclude that the proportion of Facebook uses in males and females is not same.

2) Z test for Instagram:

Hypothesis :

Ho: There is no significant difference between Instagram use in males and Instagram use in females.i.e. Ho: $P_{12} = P_{22}$

V/s

H1: There is a significant difference between Instagram use in males and Instagram uses in females. i.e. $H_1: P_{12} \neq P_{22}$

P12: Proportion of Instagram use in females.

P22: Proportion of Instagram use in males.

Observation Table :

n1	n2	X12(users) in n1	X22(users)in n2	p 12	p 22
205	194	121	112	0.5902	0.5773

Calculation:

$$P = \frac{n_1 p_{12} + n_2 p_{22}}{n_1 + n_2} = 0.5839$$

$$Q = 1 - P = 0.4160$$

Then calculated Z,

$$\text{Cal } |Z| = \frac{(p_{12} - p_{22})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 0.2612$$

The table value of Z at 5% l.o.s., Tab $|Z| = 1.96$

Result:

$Z_2 = \text{Cal } |Z| < \text{Tab } |Z|$, here we accept Ho at 5% l.o.s.

Conclusion: We conclude that the proportion of Instagram uses in males and females is same.

3) Z test for Whatsapp :

Hypothesis :

Ho: There is no significant difference between Whatsapp use in males and Whatsapp use in females.i.e. Ho: $P_{13} = P_{23}$

V/s

H1: There is a significant difference between Whatsapp use in males and Whatsapp use in females i.e. H1: $P_{13} \neq P_{23}$

P13: Proportion of WhatsApp use in females.

P23: Proportion of WhatsApp use in males.

Observation Table :

n 1	n 2	X 13(users) in n1	X 23(users) in n2	p 13	p 23
205	194	182	177	0.8878	0.9123

Calculation:

$$P = \frac{n_1 p_{13} + n_2 p_{23}}{n_1 + n_2} = 0.8997$$

$$Q = 1 - P = 0.1002$$

Then calculated Z,

$$\text{Cal } |Z| = \frac{(p_{13} - p_{23})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 0.8142$$

The table value of Z at 5% l.o.s., Tab |Z| = 1.96

Result:

$Z_3 = \text{Cal } |Z| < \text{Tab } |Z|$, here we accept Ho at 5% l.o.s.

Conclusion: We conclude that the proportion of Whatsapp uses in males and females is same.

4) Z test for Telegram :

Hypothesis :

Ho: There is no significant difference between Telegram use in males and Telegram use in females.i.e. Ho: P14 = P24

V/s

H1: There is a significant difference between Telegram use in males and Telegram use in females. i.e. H1: P14 \neq P24

P14: Proportion of telegram use in females.

P24: Proportion of telegram use in males.

Observation Table :

n1	n2	X14(users) in n1	X24(users) in n2	p 14	p 24
205	194	94	99	0.4585	0.5103

Calculation :

$$P = \frac{n_1 p_{14} + n_2 p_{24}}{n_1 + n_2} = 0.0125$$

$$Q = 1 - P = 0.9874$$

Then calculated Z,

$$\text{Cal } |Z_4| = \frac{(p_{14} - p_{24})}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 1.4693$$

The table value of Z at 5% l.o.s., Tab |Z| = 1.96

Result: $Z_4 = \text{Cal } |Z| < \text{Tab } |Z|$, here we accept Ho at 5% l.o.s.

Conclusion: We conclude that the proportion of Telegram uses in males and females is same.

5) Z test for YouTube :

Hypothesis :

Ho: There is no significant difference between YouTube use in males and YouTube use in females. i.e. Ho: $P_{15} = P_{25}$

V/s

H1: There is a significant difference between YouTube use in males and YouTube use in females. i.e. H1: $P_{15} \neq P_{25}$

P15: Proportion of youtube use in females.

P25: Proportion of youtube use in males.

Observation Table :

n1	n2	X15(users) in n1	X25(users) in n2	p 15	p 25
205	194	156	138	0.7609	0.7113

Calculation :

$$P = \frac{n_1 p_{15} + n_2 p_{25}}{n_1 + n_2} = 0.7367$$

$$Q = 1 - P = 0.2632$$

Then calculated Z,

$$\text{Cal } |Z_5| = \frac{(p_{15} - p_{25})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 1.1243$$

The table value of Z at 5% l.o.s., Tab |Z| = 1.96

Result:

$Z_5 = \text{Cal } |Z| < \text{Tab } |Z|$, here we accept Ho at 5% l.o.s.

Conclusion:

We conclude that the proportion of Youtube use in males and females is same.

6) Z test for Hotstar :

Hypothesis :

Ho: There is no significant difference between Hotstar use in males and Hotstar use in females.i.e. Ho: $P_{16} = P_{26}$

V/s

H1: There is a significant difference between Hotstar use in males and Hotstar use in females. i.e. H1: $P_{16} \neq P_{26}$

P16: Proportion of Hotstar use in females.

P26: Proportion of Hotstar use in males.

Observation Table :

n1	n2	X16(users) in n1	X26(users) in n2	p 16	p 26
205	194	48	52	0.2341	0.2680

Calculation :

$$P = \frac{n_1 p_{15} + n_2 p_{25}}{n_1 + n_2} = 0.2505$$

$$Q = 1 - P = 0.7494$$

Then calculated Z,

$$\text{Cal } |Z_6| = \frac{(p_{12} - p_{22})}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 0.7809$$

The table value of Z at 5% l.o.s., Tab |Z| = 1.96

Result:

$Z_5 = \text{Cal } |Z| < \text{Tab } |Z|$, here we accept Ho at 5% l.o.s.

Conclusion:

We conclude that the proportion of Hotstar uses in males & females is same.

OVERALL CONCLUSION

- Both male & female students same time spent 20-30 hours a week on social media.
- Most of the 26% of students use Whatsapp.
- 18.3% of students are active on social media after 9 pm.
- The 15-20 age group students most use social media.
- The social networking sites have an effect on students' education.
- Female students use Facebook more than male students.
- The proportion of Instagram use in female & male students is equal.
- WhatsApp use of female & male students is the same.
- The proportion of Telegram use in female & male students is equal.
- YouTube use of female & male students is the same.
- The proportion of Hotstar use in female & male students is equal.

LIMITATIONS

- 1) Sample size is very small to represent the whole city.
- 2) We have to limit the scope of this research according to the result we obtain from the questionnaires.
- 3) This survey is limited to Kavthe Mahankal and Jath in Maharashtra.
- 4) We consider the whole population is normal but in real reality is or may not be normal.
- 5) There is a time limit for collecting data which is not sufficient.

SCOPE

- 1) This project can be extended to the international level.
- 2) We will-planned Study can be taken including some relevant characteristics.
- 3) It can be extended to the greater region for better estimates

BIBLIOGRAPHY

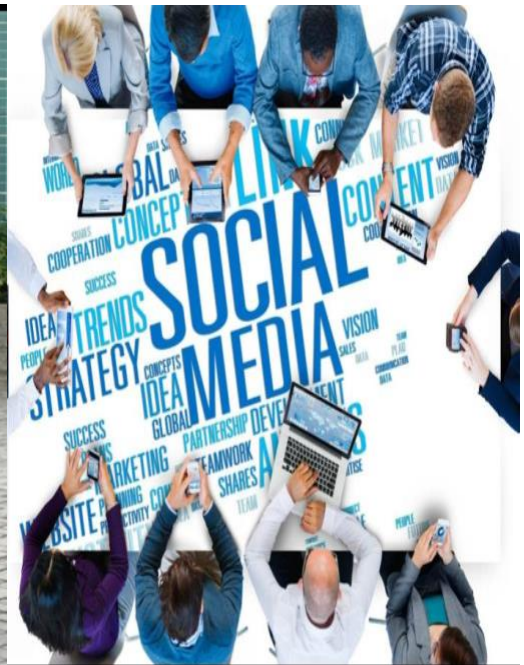
Basic Statistics. B. L. Agarwal

Mathematical Statistics. Kapoor & Gupta

Websites and search engines:-

1. <https://www.google.com>

2. <https://www.wikipedia.com>



QUESTIONNAIRE

1) Gender

Female

Male

2) Age

Before 15

16-20

21-25

26-30

31-35

Over 35

3) How many types of social networking accounts do you maintain?

1

2

3

4

5

6

4) What social networking sites do you have accounts to maintain?

Facebook

Instagram

Whatsapp

Telegram

YouTube

Hotstar

5) How many times a day do you see social networking sites?

Once a day

2-5 times a day

5-10 times a day

10+ times a day

6) How many hours do you spend on social networking sites in a day?

Minimum 2 hour

2-5 hours

5-10 hours

10+ hours

Other

7) How many hours do you spend on social networking sites in a week?

0-10 hours

10-20 hours

20-30 hours

30+ hours

Other

8)What time are you most active on social media sites?

Before 5 am
5-7 am
7-10 am
10-12 noon
12-3 pm
3-6 pm
6-9 pm
After 9 pm

9) When do you use social networking sites?

When during free time
While at school/college work
Before/after mealtime
Before/after sleeping time

10) Do social networking social affect your studies?

Yes
No

11) What are the side effects of using social networking sites?

Feeling of tension
Incidence of headache
Difficulty in sleeping
Difficulty in concentration
May lead to other mental health issues

12) What are the benefits of using social networking sites?

To Help with studies
To learn new skills
Digital representations
Entertainment
Other

13) Why do you use social media?

For entertainment
For knowledge
To contact family /friends
To watch movies /serial
To time waste
Other