Project Report
On

A COMPARATIVE STUDY BETWEEN FACE-TO-FACE AND ONLINE LEARNING METHODS

Submitted
in partial fulfilment of the requirements for the degree of
MASTER OF SCIENCE
in
APPLIED STATISTICS AND DATA ANALYTICS
by
BHAVANA RAJAN
(Register No. SM20AS008)
(2020-2022)

Under the Supervision of
Ms. SREELAKSHMI M S

DEPARTMENT OF MATHEMATICS AND STATISTICS
ST. TERESA'S COLLEGE (AUTONOMOUS)
ERNAKULAM, KOCHI - 682011
APRIL 2022

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **A COMPARATIVE STUDY BETWEEN FACE-TO-FACE AND ONLINE LEARNING METHODS is a bonafide record of the work done by Ms. BHAVANA RAJAN under my guidance as partial fulfillment of the award of the degree of Master of Science in Applied Statistics And Data Analytics at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.**

Date:
Place: Ernakulam

Ms. Sreelakshmi M S
Assistant Professor,
Department of Mathematics and Statistics,
St. Teresa's College(Autonomous),
Ernakulam.

Ms. Shanty. B.P
Assistant Professor And HOD,
Department of Mathematics,
St. Teresa's College(Autonomous),
Ernakulam.

External Examiners

1:..............................

2: ....

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of **Ms. SREELAKSHMI M S**, Assistant Professor, Department of Mathematics, St. Teresa's College(Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Ernakulam.	**BHAVANA RAJAN**
Date:	**SM20AS008**

ACKNOWLEDGEMENTS

Primarily I would like to thank the supreme power the almighty god for being able to complete this project with success. It is my privilege to express my profound gratitude to my project guide Ms. Sreelakshmi M S, Assistant professor, Department of Mathematics and Statistics. St.Teresa's College Ernakulam, for the valuable support and guidance throughout my project.

I am extremely thankful for the teaching and non-teaching staff of the Department of Mathematics and Statistics.

I am also very thankful to HOD for the valuable suggestions, and critical examination of work during the progress.

I would like to express my gratitude to each and everyone who helped and encouraged me throughout this project and made this a successful one

Ernakulam. BHAVANA RAJAN
Date: **SM20AS008**

ABSTRACT

Face-to-face learning refers to the traditional , classroom-based method of learning. This style of learning involves in person sessions with an instructor while any type of learning that occurs on the internet could be considered online learning. The educational life worldwide has been shaken by the closure of schools due to the outbreak of the coronavirus pandemic. The ripple effects have been felt in the way both teachers and students have adapted to the constraints imposed by the new online for of education. The study focuses on which teaching method is more effective for students by different statistical methods which compares various factors affecting the teaching methods. Here the data is collected primarily through questionnaires.

Contents

CERTIFICATE ii

DECLARATION iii

ACKNOWLEDGEMENTS iv

ABSTRACT v

CONTENT vi

1 INTRODUCTION 1

1.1 OBJECTIVES 1

1.2 RESEARCH METHODOLOGY 2

1.2.1 Research Approaches 2

1.2.2 Criteria of Good Research 2

1.2.3 Collection of Data through Questionnaires 3

2 Literature Review 6

3 FUNDAMENTALS OF STATISTICS 8

3.1 POPULATION, SAMPLE, AND ESTIMATORS 8

3.1.1 Probability Density Distribution (PDF) 9

3.1.2 Central Limit Theorem and Law of Large Numbers 9

3.1.3 Hypothesis Testing 10

3.1.4 Significance Level and P-Value 11

3.2 SAMPLING DESIGN 12

3.2.1 Criteria of selecting a sampling procedure 12

3.2.2 Characteristics of Good Sample Design 13

3.2.3 How to select a random sample 14

3.3 SAMPLING THEORY 15

3.3.1 Systematic sampling 15

3.3.2 Stratified Sampling 16
3.4 CHARTS AND DIAGRAMS 16
3.4.1 Pie Chart 16
3.4.2 Multiple Bar Diagram 16

4 TESTING PROBLEM **17**
4.1 PROCEDURE OF HYPOTHESIS TESTING 17
4.2 TESTS OF HYPOTHESIS 19
4.3 CHI-SQUARE TEST 19
4.3.1 Chi-square Test of Goodness of Fit 20
4.4 ANOVA 20
4.4.1 One-Way ANOVA 21

5 DIAGRAMS AND ANALYSIS OF DATA **24**
5.1 CHARTS AND DIAGRAMS 24
5.2 DATA ANALYSIS 29

6 SUMMARY AND CONCLUSION **33**
REFERENCES 34

Chapter 1

INTRODUCTION

Learning has been a great part of human history over the years. People have always been looking for ways to advance their knowledge, understand the world better and make their lives easier. The ways of learning have definitely improved but the goals remains the same. Face-to-Face learning is how we have historically been learning. It usually takes place in a classroom and most times is centered around the lecturer. Face-to-Face learning requires you to be at a specific time. It also usually requires interaction or participation at times. A growing number of students are now opting for online classes. They find the traditional classroom learning restrictive, inflexible, and impractical. In this modern age of technological advancement, institutions can provide effective classroom teaching via internet.

1.1 OBJECTIVES

- The aim is to determine which teaching method is more effective and which instructional modality generates better student performance.
- We can examine score variability between mental satisfaction to know whether teaching method have a greater impact.
- In addition to the above objective,we can examine score variability with different educational groups to determine if teaching method have a greater impact.

1.2 RESEARCH METHODOLOGY

1.2.1 Research Approaches

There are two basic research methodologies – quantitative and qualitative approaches. The former entails the collection of quantitative data that may then be submitted to rigorous quantitative analysis in a formal and rigid manner. This approach can be further subdivided into inferential, experimental, and simulation research.

- The goal of the inferential approach to research is to create a database from which to infer population characteristics or relationships. This usually refers to survey research, in which a subset of a population is studied (questioned or observed) to determine its characteristics, and it is then assumed that the entire population has the same characteristics.
- The experimental approach is distinguished by greater control over the research environment, and in this case, some variables are manipulated to observe their effect on other variables.

The qualitative research approach is concerned with the subjective evaluation of attitudes, opinions, and behaviour. In such a case, research is a function of the researcher's insights and impressions. Such a research approach yields results that are either non-quantitative or have not been subjected to rigorous quantitative analysis. In general, focus group interviews, projective techniques, and depth interviews are used.

1.2.2 Criteria of Good Research

Whatever the types of research works and studies are, one thing is certain, they all share a common ground in the scientific method that they employ. Scientific research is expected to meet the following criteria:

1 The research's purpose should be clearly defined, and common concepts should be used.

2 The research procedure should be described in sufficient detail to allow another researcher to repeat the research for further advancement while maintaining the continuity of what has already been accomplished.

3 The research procedure should be carefully planned in order to produce results that are as objective as possible.

4 The researcher should be completely honest about any shortcomings in the procedure design and evaluate their impact on the findings.

5 Data analysis should be sufficient to disclose the relevance of the data, and the methods of analysis utilized should be appropriate. The data's legitimacy and reliability should be double-checked.

6 Conclusions should be limited to those that are supported by the study data and for which the data provide an acceptable foundation.

7 If the researcher is experienced; has a strong research reputation, and is a person of integrity, greater confidence in the research is warranted.

1.2.3 Collection of Data through Questionnaires

This type of data collection is very popular, especially when dealing with large inquiries. Private individuals, researchers, private and public organisations, and even governments are using it. This method involves sending a questionnaire to the people involved (typically via mail) and asking them to answer the questions and return the form. A questionnaire is made up of a group of questions that are printed or typed in a specific order on a form or a set of forms. The questionnaire is distributed to respondents, who are expected to read and comprehend the questions before responding in the space provided on the questionnaire. The respondents are on their own when it comes to answering the questions.

Main features of a questionnaire: A questionnaire is frequently regarded as the heart of a survey operation. As a result, it should be built with great care. The survey will very certainly fail if it is not correctly set up. This necessitates an examination of the most important parts of a questionnaire, including the general structure, question sequencing, and question formulation and language. When it comes to these three primary aspects of a questionnaire, researchers should keep the following in mind:

General Form :

In terms of overall form, a questionnaire can be either organised or unstructured. Structured questionnaires are ones in which the questions are specific, concrete, and pre-determined. All responders are asked the same questions in the same language and in the same order. This type of standardization is used to ensure that all respondents answer the same set of questions. The question can be closed (i.e., 'yes' or 'no') or open (i.e., enabling a free response), but it should be expressed ahead of time and not formed during the interview. Fixed alternative questions may be included in structured questionnaires, limiting the informants' responses to the stated options. As a result, a highly structured questionnaire is one in which all questions and responses are specified, and the number of comments in the respondent's own words is kept to a minimum. When these features are absent from a questionnaire, it is referred to be unstructured or non-structured. More specifically, in an unstructured questionnaire, the interviewer is given a general guide on the type of information to be obtained, but the exact question formulation is largely his responsibility, and the responses are to be taken down in the respondent's own words to the extent possible; tape recorders may be used in some situations to achieve this goal.

Question sequence :

A researcher should pay attention to the question sequence when constructing the questionnaire in order to make it successful and assure the quality of the responses gathered. Individual inquiries are much less likely to be misunderstood when they are asked in the right order. The question sequence must be straightforward and fluid, implying that the relationship between one question and the next should be obvious to the respondent, with the easiest-to-answer questions placed first. The initial few questions are crucial since they are likely to influence the respondent's attitude as well as his willingness to cooperate. The first questions should be designed to pique people's curiosity. In general, the following types of questions should not be used as opening questions in a questionnaire:

1 questions that place an undue strain on the respondent's memory or intellect;

2 questions about a person's personal character;

3 queries about personal riches, etc.

Question formulation and wording :

In terms of this component of the questionnaire, the researcher should keep in mind that each question must be very clear, as any misunderstanding can cause irreversible harm to a survey. In order to avoid painting a skewed view of reality, the question should be neutral. Questions should be written in such a way that they fit within a well-thought-out tabulation strategy. All questions should, in general, meet the following criteria: (a) be simple (i.e., transmit only one notion at a time); (b) be concrete (i.e., conform as much as possible to the respondent's way of thinking); and (c) be concrete (i.e., conform as much as possible to the respondent's way of thinking.

Chapter 2

Literature Review

Kaur et al. (2020) studied that the effectiveness of online learning by internet, communication skills, and knowledge in medical students by taking a cross-sectional survey from a sample size of 983 and analyzing the results using mean and standard deviation. The result of the paper was that online learning is equally effective as compared to offline learning in some parameters. Shenoy et al. (2020) studied that the student's engagement and learning by technology adoption and teaching the results using MS-Excel. The results of the paper were that the class engagement is better online than offline.

Swan (2019) studied that the barriers in online learning occur due to arguing for commonly agreed-upon protocols, and tension, between the social and cognitive presence by taking surveys from a sample size of 270 and analyzing the results using the Col framework. The result of the paper was that some difficulties were found in communication and tension in students' point of view. Yen et al. (2018) studied the course satisfaction in face-to-face learning and online learning by taking the surveys from a sample size of 85 and analyzing the results using Multivariate Analysis of Variance. The results of the paper were that online classes can be just as effective as face-to-face classes in producing satisfactory student outcomes.

Kebritch et al.(2017) studied that there are some issues in the online participation, and it was rather difficult to transit from face-to-face learning to online learning with the help of quantitative, qualitative and mixed methods by taking the sample size of 400 and analyzing the results using Coopers framework. The result of the paper was that there are issues in learners expectations from online learning and participation in online learning.

Anna Ya Ni(2013) studied the effectiveness of online learning on grades by teaching methods and performances and assessments with the help of the surveys on a sample size of 148 and analyzing the results using mean, chi-square value, and p-value. The result of the paper was that on the basis of grade , online learning is less effective in terms of calculative class. Baig(2011) studied that the effectiveness of online learning, face-to-face learning and the grades in school with the help of questionnaires from a sample size of 40 and analysed the results using t-test, mean and SPSS. The result of the paper was that online learning is highly effective but there is a need of more facilities.

Chapter 3

FUNDAMENTALS OF STATISTICS

3.1 POPULATION, SAMPLE, AND ESTIMATORS

The population has an underlying distribution process that we are rarely able to fully comprehend. We can take samples from the population and estimate the population using the samples. But how do you pick the right estimators? A good estimator has three characteristics: it is impartial, consistent, and efficient. The estimator is unbiased; when the expected value of the sample parameter is equal to the population parameter:

$$E[parameter_{sample}] = parameter_{population}$$

The estimator is consistent if the sample parameter's variance reduces as the sample size grows. The estimator with smaller variance is more efficient with the same sample size.

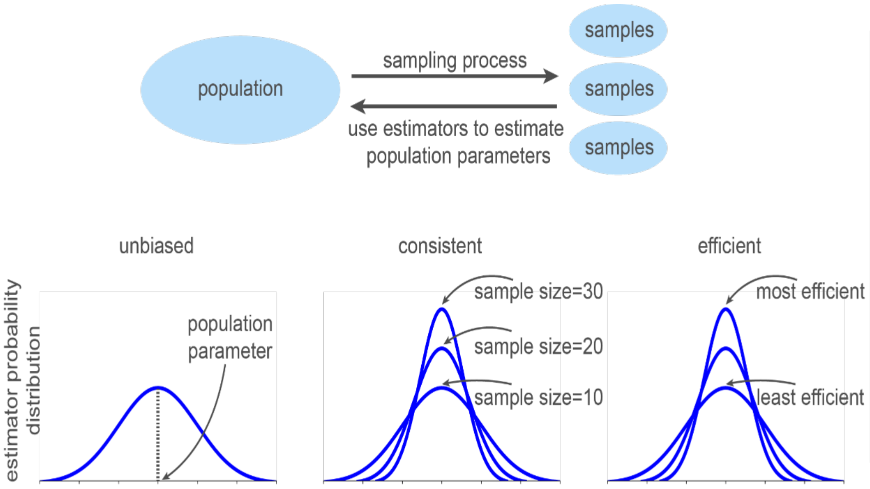


Figure 3.1: A good estimator is unbiased, consistent, and efficient.

3.1.1 Probability Density Distribution (PDF)

The probability density distribution (PDF) is used to specify the probability of the random variable falling within a particular range of values. The probability density at a certain x is denoted as f(x). By applying integral function to f(x) over a range of (x1,x2), the probability of x falling in (x1,x2) can be calculated.

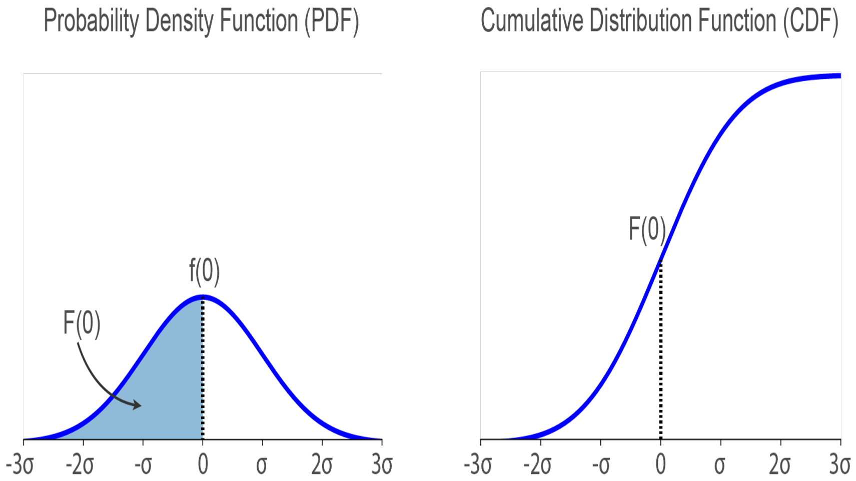
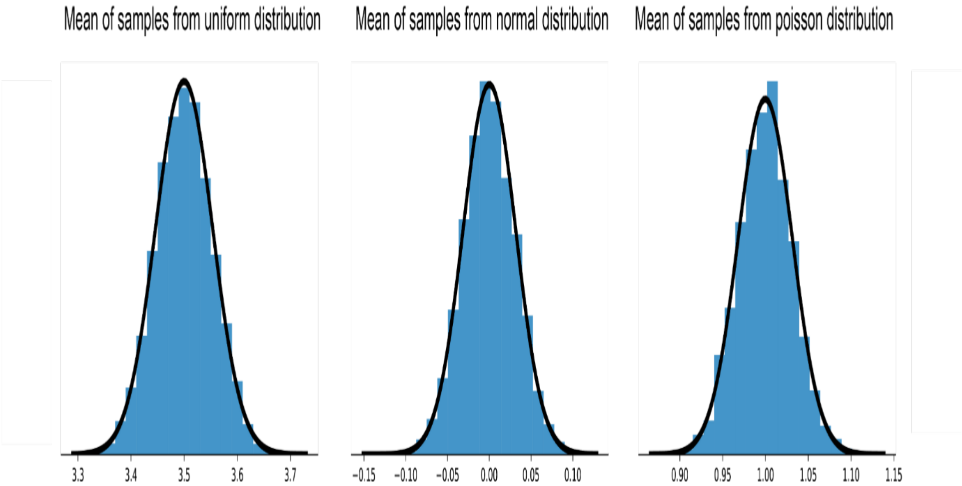


Figure 3.2: PDF and CDF of normal Distribution.

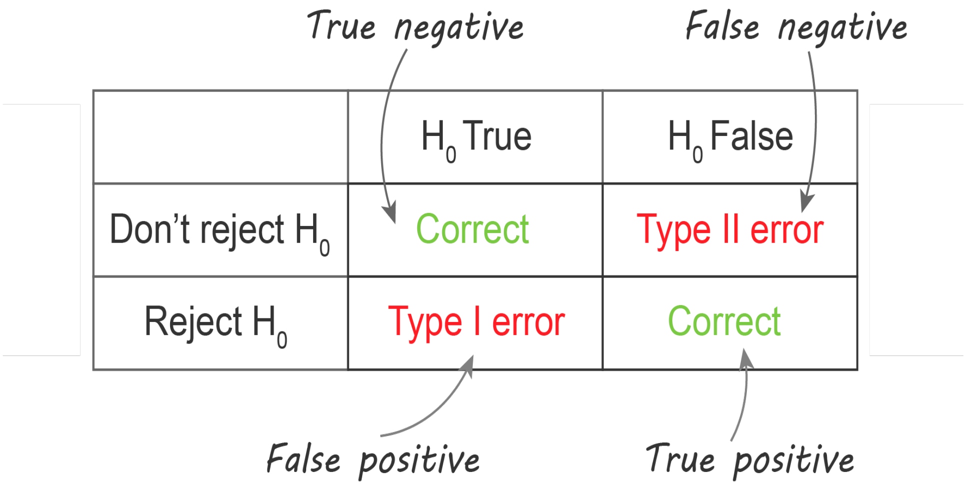
3.1.2 Central Limit Theorem and Law of Large Numbers

Central Limit Theorem states that when the sample size is large, the sample mean of the independent random variable follows normal distribution. Typically, when sample size is larger than 30, the requirement of large sample size is considered fulfilled. The independent random variables can follow any distribution, while the sample mean of these independent random variables follows normal distribution. The law of large numbers states that as the number of trials increases, the estimator's average approaches the theoretical value. If we only repeat the trials 10 times for the aforementioned experiment, the distribution will be considerably different from the plots.



3.1.3 Hypothesis Testing

Given that only sample parameters can be determined, we must use hypothesis testing to infer population parameters. A set of complementary hypotheses, consisting of a null hypothesis and an alternative hypothesis, is proposed in hypothesis testing. We choose to assume the null hypothesis is correct while doing hypothesis testing. We do not reject the null hypothesis if the observed value is likely to occur under the assumption that the null hypothesis is true. If the observed value is improbable to occur, the null hypothesis is rejected and the alternative hypothesis is accepted.



	H_0 True	H_0 False
Don't reject H_0	Correct	Type II error
Reject H_0	Type I error	Correct

True negative

False negative

False positive

True positive

Figure 3.3: Hypothesis Matrix

3.1.4 Significance Level and P-Value

We must first define a significance level before doing hypothesis testing. The level of significance dictates how much we want to trust in the null hypothesis. If the significance level is set to 0.05, the null hypothesis is not rejected as long as the likelihood of the observation is greater than 5%. We reject the null hypothesis and accept the alternative hypothesis if the likelihood of the observation is less than 5%. Between Type I and Type II errors, there is a trade off. In other words, a greater significance level makes rejecting the null hypothesis easier. Although a greater significance level reduces the Type II error, it also increases the Type I error. The only way to reduce both Type I and Type II error is by increasing the sample size.

The p-value is the probability of the observed value. A low p-value indicates that the observation is unlikely to happen if the null hypothesis is correct. We reject the null hypothesis when the p-value is less than the significance level. One thing to keep in mind is that the p-value is binary: it is only larger or smaller than the significance level.

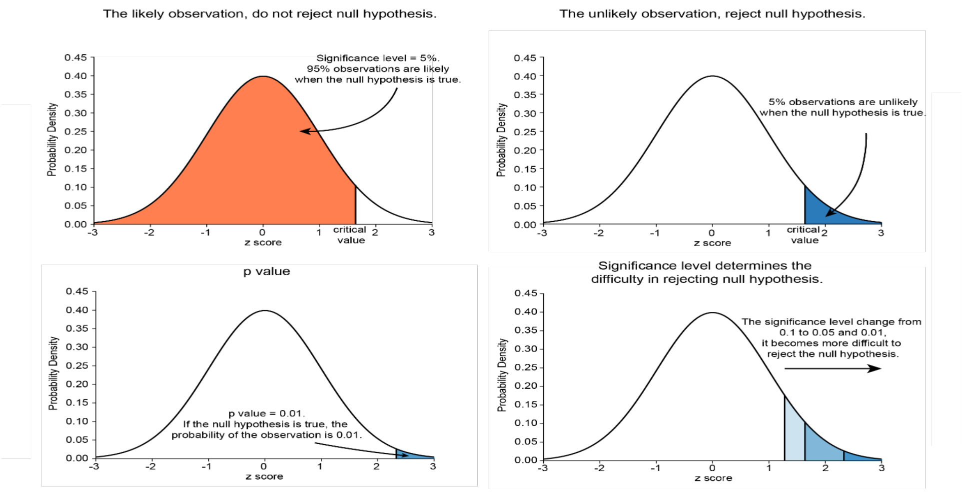


Figure 3.4: How to interpret significance level and p value

3.2 SAMPLING DESIGN

3.2.1 Criteria of selecting a sampling procedure

In this context, it's important to note that a sampling study has two costs: the cost of collecting the data and the cost of making an inaccurate conclusion based on the data. The two causes of inaccurate inferences, systematic bias and sampling error, must be kept in mind by the researcher. Errors in the sampling procedures cause systematic bias, which cannot be minimised or avoided by increasing the sample size. The reasons of these inaccuracies can, at best, be identified and corrected. A systematic bias is usually caused by one or more of the causes listed below:

1 Inappropriate sampling frame: A systematic bias will emerge if the sample frame is unsuitable, i.e., a skewed depiction of the universe.

2 Defective measuring device: Systematic bias will arise if the measuring apparatus is consistently in error. If the questionnaire or the interviewer is biased, systemic bias can occur in survey work. Similarly, if the physical measuring device is defective, the data collected through that device will be systematically biased.

3 Non-respondents: There may be a systematic bias if we are unable to sample all of the people who were initially included in the sample. The rationale behind this is that in such a setting, the possibility of making contact with

or receiving a response from an individual is frequently associated with the estimated value.

4 Indeterminacy principle: When people are kept under surveillance, we can notice that they behave differently than when they are kept in non-observed conditions. For example, workers who are aware that they are being observed in the course of a work study on the basis of which the average length of time to complete a task will be determined and, as a result, the quota for piece work will be set, tend to work more slowly than workers who are not being observed. As a result, the indeterminacy principle could be a source of systematic bias.

5 Natural bias in the reporting of data: Natural bias in data reporting by respondents is frequently the source of systematic bias in many studies. The income data collected by the government revenue department has a downward bias, whereas the income data obtained by some social organisations has an upward bias. When asked about their income for tax purposes, people tend to understate it, but they inflate it when asked about their social position or prosperity. People tend to give what they believe is the 'right' answer rather than sharing their genuine feelings in psychological questionnaires.

3.2.2 Characteristics of Good Sample Design

We can list the features of a good sample design as follows based on what has been stated above:

(a) The sample design must result in a sample that is actually representative.

(b) The sample design must be such that the sampling error is modest.

(c) The sample design must be feasible in the context of the research study's budget.

(d) The sample design must be such that systematic bias can be better controlled.

(e) The sample size should be such that the sample study's findings can be applied to the entire universe with a fair level of confidence.

3.2.3 How to select a random sample

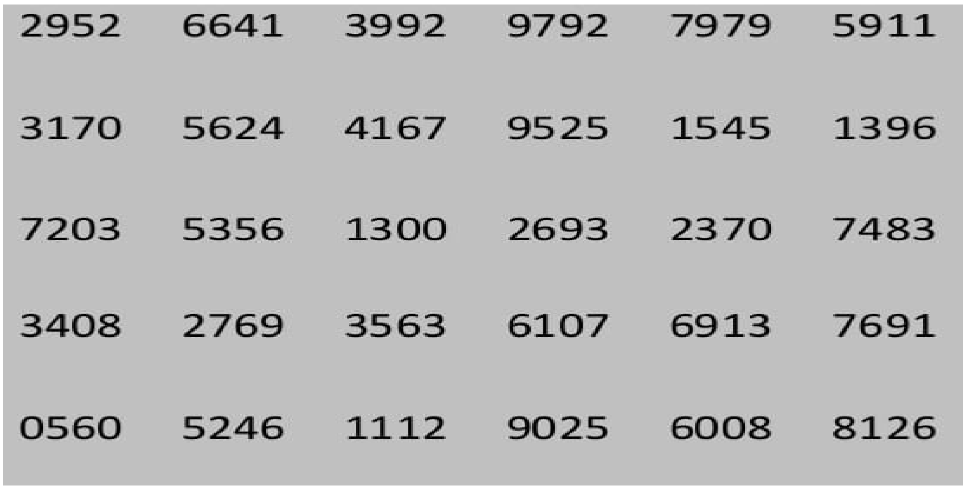
In simple scenarios like the one above, we may write each of the possible samples on a slip of paper, thoroughly mix these slips in a container, and then draw as a lottery either blindfolded or by rotating a drum or by any other similar device. In complex sampling issues, such an approach is clearly unworkable, if not impossible. In actuality, the method's practical value is severely limited.

Fortunately, we can obtain a random sample in a much simpler manner than enumerating all potential samples on paper slips, as mentioned previously. Instead, we can write the name of each element of a finite population on a slip of paper, place the prepared slips of paper in a box or bag, thoroughly mix them, and then draw (without seeing) the appropriate number of slips for the sample one by one without replacement. In doing so, we must ensure that each of the remaining parts of the population has the same chance of being selected in subsequent drawings. Each conceivable sample will have the same probability as a result of this method. Using the preceding example, we can verify this. Because we have a finite population of 6 elements and want to select a sample of size 3, the probability of drawing any one element for our sample in the first draw is 3/6, the probability of drawing one more element in the second draw is 2/5 (the first element drawn is not replaced), and so on. Because these draws are random, the combined probability of the three parts that make up our sample is the product of their separate probabilities, which equals 3/6x2/5x 1/4 = 1/20.

This verifies our earlier calculation.

The usage of random number tables can even simplify this rather simple way of producing a random sample in practise. Various statisticians, like Tippett, Yates, and Fisher, have created random number tables that can be used to select a random sample. Tippet's random number tables are commonly utilised for this purpose. Tippett provided a total of 10400 four-digit numbers. He took 41600 digits from census reports and divided them into fours to get random numbers that could be used to generate a random sample.

An example might be used to demonstrate the technique. First, we reproduce Tippet's first thirty sets of number. Suppose we are interested in taking a sample



2952	6641	3992	9792	7979	5911
3170	5624	4167	9525	1545	1396
7203	5356	1300	2693	2370	7483
3408	2769	3563	6107	6913	7691
0560	5246	1112	9025	6008	8126

Figure 3.5: How to interpret significance level and p value

of 10 units from a population of 5000 units, bearing numbers from 3001 to 8000. We shall select 10 such figures from the above random numbers which are not less than 3001 and not greater than 8000. If we randomly decide to read the table numbers from left to right, starting from the first row itself, we obtain the following numbers: 6641, 3992, 7979, 5911, 3170, 5624, 4167, 7203, 5356, and 7483.

The units bearing the above serial numbers would then constitute our required random sample.

It's worth noting that using random number tables to take random samples from finite populations is only possible when lists are available and items are easily numbered. However, in some cases, proceeding in the manner described above is impracticable. It would not be possible to number the trees and choose random numbers to select a random sample if we wanted to estimate the mean height of trees in a forest, for example. In such cases, we should choose some trees for the sample carelessly, with no particular goal in mind, and regard the sample as a random sample for research purposes.

3.3 SAMPLING THEORY

3.3.1 Systematic sampling

Systematic sampling is a sampling strategy in which just the first unit is selected using random numbers and the rest is selected automatically according to a predetermined pattern. Suppose N units of the population are numbered from 1 to N. Let N = nk where n is the size of the sample to be selected and k is an integer.

Draw a random number i ¡= k and select the unit with the corresponding serial number and every k th unit in the population thereafter. ‘i’ is called random start and ‘k’ is called sampling interval. The sample will contain n units with serial number i, i+k, i+2k,. . . , i+(n-1)k. Such a sample is known as systematic sample and such a procedure is termed as linear systematic sampling. This technique will generate ‘k’ possible systematic sample with equal probability.

3.3.2 Stratified Sampling

Simple random sampling will produce satisfactory results when the population is more or less homogeneous in terms of the attribute under consideration. However, if the data is highly heterogeneous, simple random sampling may produce estimates that are significantly different from the genuine values. One way to reduce the variability of estimates is to use stratified sampling. We divide the diverse population into more or less homogeneous strata or sub-populations using stratified sampling. The sample is made up of basic random samples selected from several strata. Assume there are N units in the population and that we need a sample size of n. We choose simple random samples of sizes n1 , n2 , . . . , nk from the k strata and they together form our sample of size n.

3.4 CHARTS AND DIAGRAMS

3.4.1 Pie Chart

The percentage values are shown as slices of a pie in a pie chart. A pie chart (sometimes known as a circle chart) is a circular statistical graphic divided into slices to depict numerical proportions. Each slice’s arc length (and thus its centre angle and area) in a pie chart is proportionate to the quantity it represents.

3.4.2 Multiple Bar Diagram

A multiple bar graph depicts the relationship between various data variables. A column in the graph represents each data value. With the addition of columns, a multiple bar graph displays many data points for each category of data.

Chapter 4

TESTING PROBLEM

4.1 PROCEDURE OF HYPOTHESIS TESTING

To test a hypothesis, the researcher must determine whether the hypothesis appears to be valid based on the evidence gathered. The basic question in hypothesis testing is whether the null hypothesis should be accepted or rejected. The processes we take to choose between rejecting and accepting a null hypothesis are referred to as the procedure for hypothesis testing. The following are the steps involved in hypothesis testing:

- Making a formal statement: The phase entails formally stating the null hypothesis (H0) as well as the alternative hypothesis (H1) (H1). This means that, given the nature of the study challenge, hypotheses should be articulated explicitly. For example, if Mr. Mohan of the Civil Engineering Department wants to test an ancient bridge's load bearing capacity, which must be greater than 10 tonnes, he can formulate his hypothesis as follows: Null hypothesis $H0 : \mu = 80$ Alternative Hypothesis $H1 : \mu \neq 80$ The formulation of hypotheses is an important step which must be accomplished with due care in accordance with the object and nature of the problem under consideration. It also indicates whether we should use a one-tailed test or a two-tailed test. If H1 is of the type greater than (or of the type lesser than), we use a one-tailed test, but when H1 is of the type "whether greater or smaller" then we use a two-tailed test.

- Selecting a significance level: The hypotheses are tested at a predetermined level of significance, which should be mentioned as well. In practise, either a

5% or a 1% threshold is used for this reason. The following factors influence the level of significance:

(a) The size of the difference between the sample means.

(b) The size of the samples

(c) The measurement variability within samples

(d) whether or not the hypothesis is directed (A directional hypothesis is one which predicts the direction of the difference between, say, means). In summary, the level of significance must be appropriate for the goal and type of the investigation.

- Deciding the distribution to use: The next stage in hypothesis testing is to identify the proper sampling distribution after settling on the level of significance. The normal distribution and the t-distribution are the most common options. The rules for choosing the correct distribution are similar to those we discussed earlier in relation to estimation.

- Selecting a random sample and computing an appropriate value: Another step is to choose a random sample(s) and use the applicable distribution to compute an acceptable value for the test statistic from the sample data. To put it another way, draw a sample to provide empirical data.

- Calculation of the probability: The likelihood that the sample result would diverge as greatly from expectations as it has if the null hypothesis were true must then be calculated.

- Comparing the probability : Another step is to compare the probability that has been calculated with the significance level value that has been specified. If the estimated probability is equal to or less than the α value in a one-tailed test (and $\alpha/2$ in a two-tailed test), reject the null hypothesis (accept the alternative hypothesis), but accept the null hypothesis if the calculated probability is larger. If we reject H0, we run the danger of committing a Type I error (at most the level of significance), but if we accept H0, we run the risk of committing a Type II error (the amount of which cannot be determined as long as H0 is ambiguous rather than precise).

4.2 TESTS OF HYPOTHESIS

Hypothesis testing is a technique for determining whether a population hypothesis is likely to be true or false based on a sample of facts. For the aim of testing hypotheses, statisticians have devised numerous tests of hypotheses (also known as significance tests) that can be classed as follows: Parametric tests, also known as standard tests of hypotheses, and non-parametric tests, also known as distribution-free testing of hypotheses, are two types of tests. The parent population from which we collect samples is frequently assumed to have specific traits using parametric testing. Before parametric tests can be performed, assumptions such observations come from a normal population, sample size is high, and assumptions about population parameters like mean, variance, and so on, must hold true. However, there are times when the researcher is unable or unwilling to make such assumptions. In such cases, we employ non-parametric tests, which are statistical procedures for evaluating hypotheses that do not rely on any assumptions about the parameters of the parent population. Furthermore, most non-parametric tests assume just nominal or ordinal data, whereas parametric tests require at least an interval scale of measurement. As a result, in order to attain the same amount of Type I and Type II errors, nonparametric tests require more observations than parametric tests.

4.3 CHI-SQUARE TEST

Among the many tests of significance devised by statisticians, the chi-square test is an important one. Chi-square is a statistical metric for comparing a variance to a theoretical variance in the context of sampling analysis. It can be used to determine if categorical data shows dependency or the two classifications are independent" as a nonparametric test. When categories are included, it can also be used to compare theoretical populations to actual data. As a result, the chi-square test can be used to solve a wide range of problems. In fact, the test is a tool that allows all researchers to

1 test the goodness of fit

2 test the significance of the relationship between two variables.

3 test the homogeneity or the significance of population variance

4.3.1 Chi-square Test of Goodness of Fit

Prof. Karl Pearson introduced the "Chi-square test of goodness of fit" in 1900, which is a very powerful tool for determining the importance of the disagreement between theory and experiment. It allows us to determine if the experiment's deviation from theory is due to chance or to the theory's inability to fit the observed facts. If Oi,(i = 1,2,..., n) is a set of observed (experimental) frequencies and Ei,(i = 1,2,..., n) is a set of expected (theoretical or hypothetical) frequencies, then Karl Pearson's chi-square is

$$x^2 = \Sigma_{i=1}^{n}[\frac{(o_i - E_i)^2}{E_i}]$$

follows Chi-square distribution with (n-1) degrees of freedom.

4.4 ANOVA

ANOVA is a procedure for determining whether or not distinct sets of data are homogeneous. "The core of ANOVA is that it divides the entire amount of variation in a set of data into two categories: that which can be attributed to chance and that which can be assigned to specific causes." There could be differences between samples and within sample products. The goal of ANOVA is to partition the variance for analytical purposes. As a result, it is a means of dissecting the variance that a response is exposed to into its numerous components, each of which corresponds to a different source of variation. This technique can be used to explain if different types of seeds, fertilisers, or soils differ significantly so that a policy decision can be made based on that information in the context of agricultural research. Similarly, using the ANOVA technique, differences in numerous types of feed created for a given class of animal or various types of medications manufactured to treat a specific condition can be analysed and assessed to be significant or not. Similarly, a manager of a large corporation can examine the performance of numerous salespeople in his company to see if their results differ dramatically. As a result, using the ANOVA technique, one can analyse any number of postulated or claimed factors that influence the dependent variable. It's also worth looking at the distinctions between other categories within each of these criteria, which could have a wide range of possible values. When we explore the differences between the many categories of a single factor with many possible values, we use one-way ANOVA, and when we investigate two factors at the same time, we use two-way ANOVA. The interaction (i.e., inter-relation between two independent variables/factors), if any, between two independent variables affecting a dependent variable can also be investigated for better decisions in a two or more way ANOVA.

4.4.1 One-Way ANOVA

We evaluate only one component in the one-way ANOVA and then remark that the reason for that factor's importance is that it can contain numerous different sorts of samples. After that, we look to see whether there are any variances within that factor. The procedure entails the following steps:
(i)Obtain the mean of each sample i.e, obtain

$$\overline{X}_1, \overline{X}_2, ..., \overline{X}_k$$

When there are k samples.
(ii)Work out the mean of the sample means as follows:

$$\overline{X} = \overline{x}_1 + \overline{x} + ... + \overline{x}_k / Noofsamples(k)$$

(iii) Calculate the square of the deviations of the sample means from the mean of the sample means, which can then be multiplied by the number of items in the relevant sample to get the total. The sum of squares for variance between samples (or SS between) is what this is called. This can be written symbolically as:
SS between $= n_1(\overline{x}_1 - \overline{\overline{x}})^2 + n_2(\overline{x}_2 - \overline{\overline{x}})^2 + ... + n_k(\overline{x}_k - \overline{\overline{x}})^2$

(iv) To calculate variance or mean square (MS) between samples, divide the result of the (iii) step by the degrees of freedom between the samples. This can be written symbolically as: MS between = SS between/(k-1) Where (k-1) represents degrees of freedom (d.f.) between samples.

(v) Calculate the squares of the deviations of the values of the sample items for all the samples from the corresponding means of the samples, and then add them together to get the total. The sum of squares for variance within samples is what this total is called (or SS within). This can be written symbolically as:

$$SS\ within = \Sigma(x_{1i} - \overline{x}_1)^2 + \Sigma(x_{1i} - \overline{x}_1)^2 + ... + \Sigma(x_{ki} - \overline{x}_k)^2$$

(vi) To get the variance or mean square (MS) within samples, divide the result of the (v) step by the degrees of freedom within samples. This can be written symbolically as:
MS within = SS within/(n-k)
Where (n-k) represents degrees of freedom within samples
n = total number of items in all the samples i.e.,
k= number of samples.

(vii) When the deviations for individual items in all the samples have been subtracted from the mean of the sample means, the sum of squares of deviations for total variance may be calculated by summing the squares of deviations. This can be written symbolically as:

$$SS\ for\ total\ variance = \Sigma(x_{ij} - \overline{\overline{x}})^2 \quad i, j = 1, 2, 3..$$

SS for total variance The total should be equal to the total of the result of the (iii) and (v) steps explained above
ie; SS for total variance = SS between + SS within The degrees of freedom for total variance will be equal to the number of items in all samples minus one
ie. (n - 1). The degrees of freedom for between and within must add up to the degrees of freedom for total variance
i.e, (n - 1) = (k - 1) + (n - k)
This fact explains the additive property of the ANOVA technique.

(viii) Finally, F-ratio may be worked out as under:

$$F - ratio = MSbetween/MSwithin$$

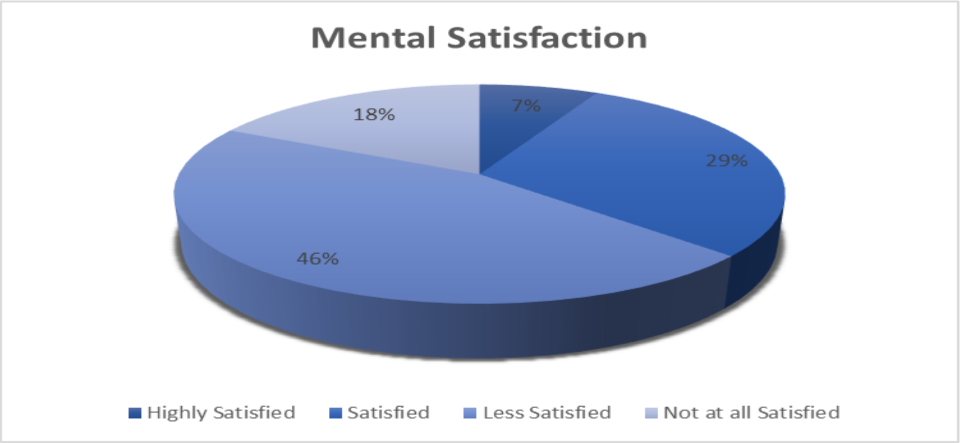
This ratio is used to determine whether the difference between the means of numerous samples is substantial or merely due to sampling variability. For this, we consult the table, which lists the values of F for various degrees of freedom at various levels of significance. If the worked-out value of F is less than the table value of F, the difference is considered unimportant, i.e., attributable to chance, and the null hypothesis of no difference between sample means is maintained. In case the calculated value of F happens to be either equal or more than its table value, the difference is considered as significant (which means the samples could not have come from the same universe) and accordingly the conclusion may be drawn. The higher the calculated value of F is above the table value, the more definite and sure one can be about his conclusions.

Chapter 5

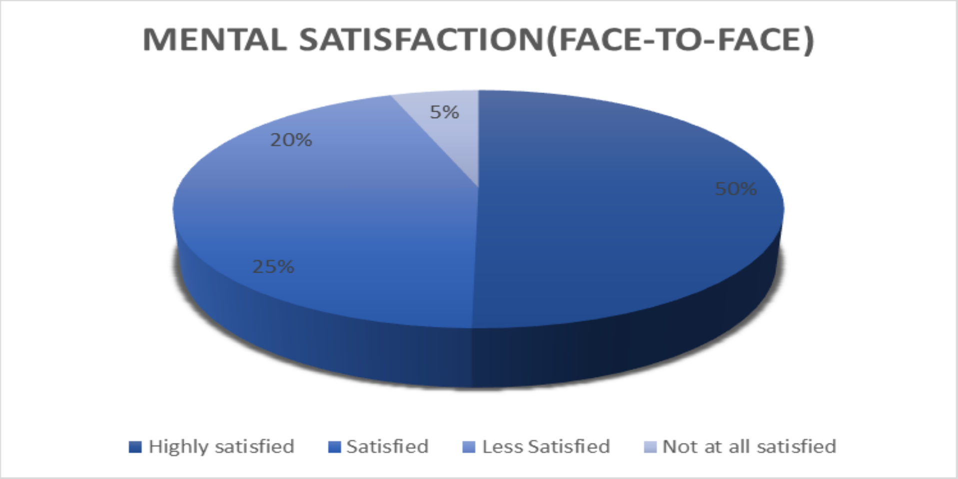
DIAGRAMS AND ANALYSIS OF DATA

5.1 CHARTS AND DIAGRAMS

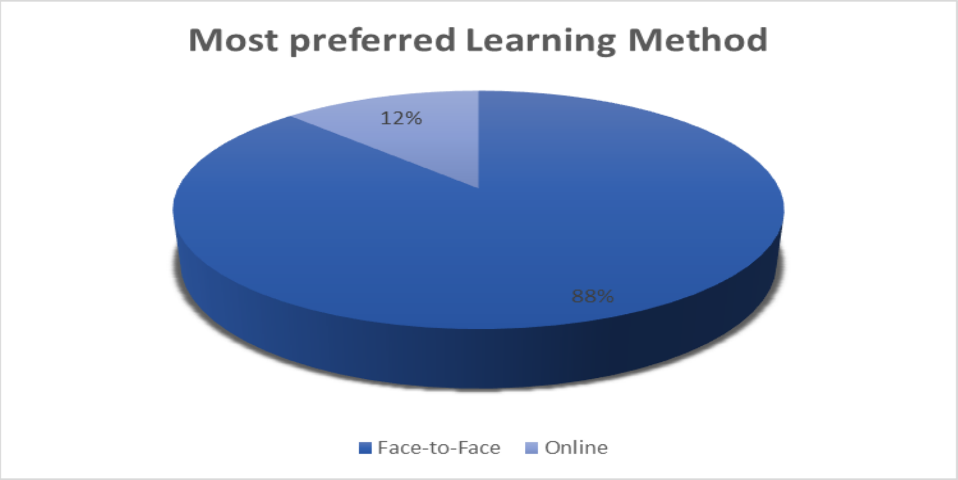
1. Pie chart for Mental Satisfaction of Students during online class.



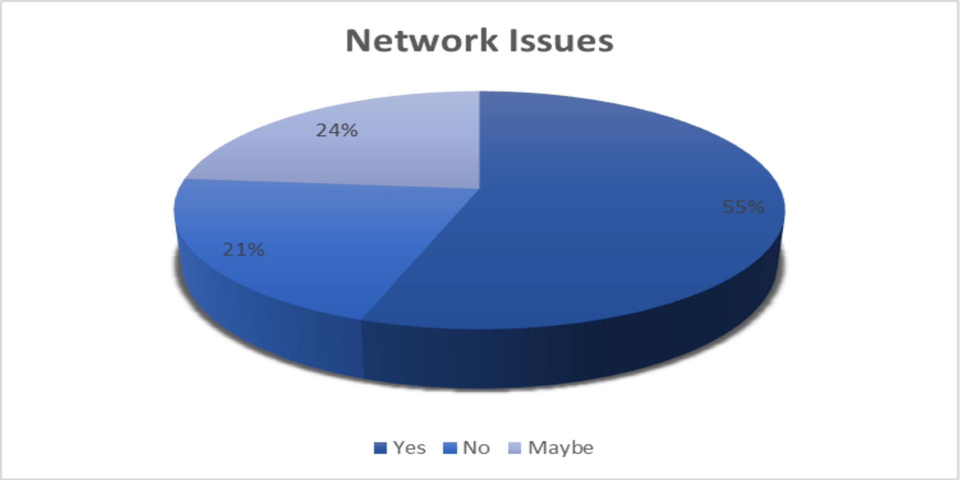
2. Pie chart of mental satisfaction of students during face-to-face class.



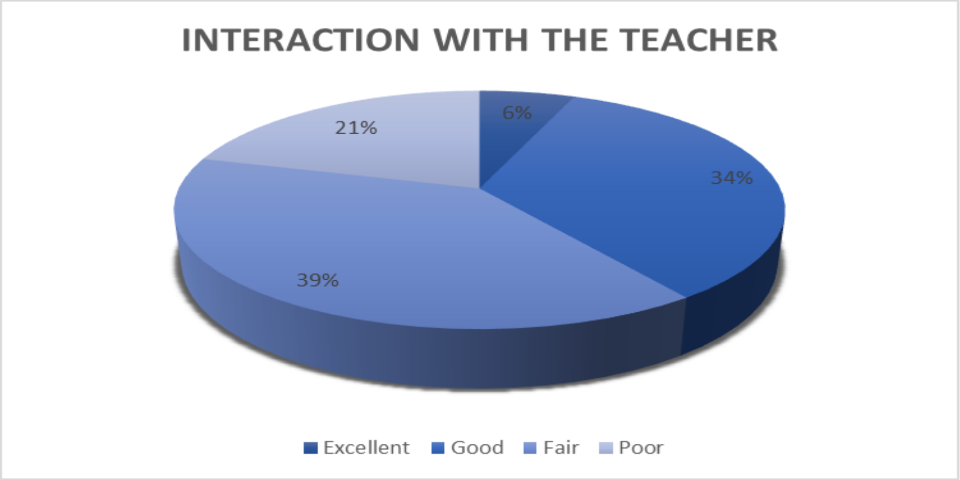
3. Pie chart for most preferred Learning Method



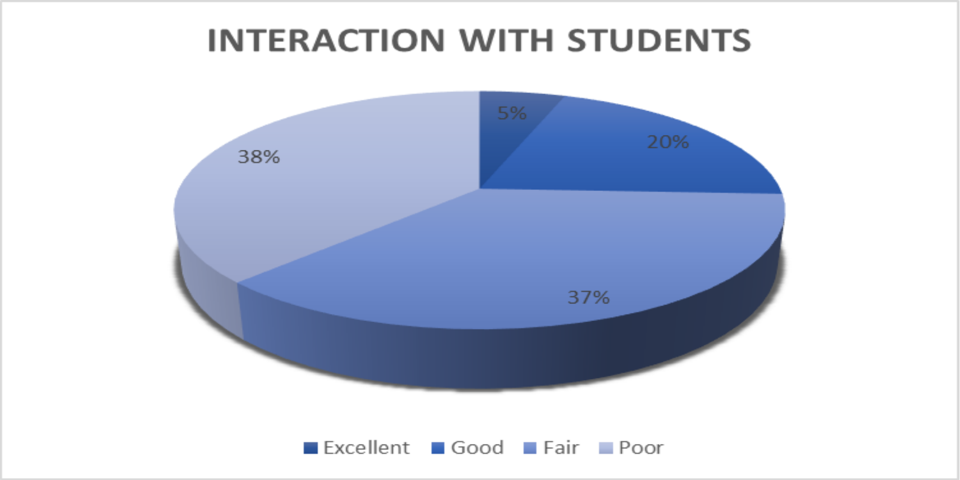
4. Pie chart of Network issues faced by Students



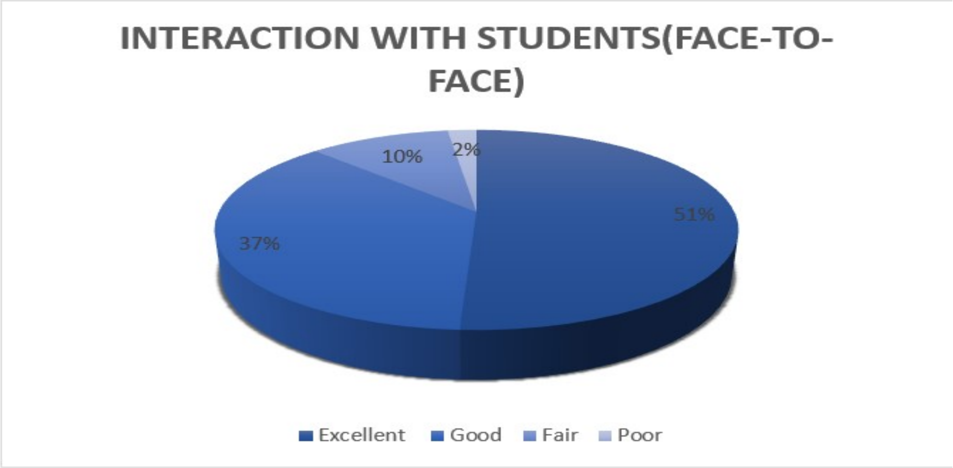
5. Pie chart of interaction with teacher during online class.



6. Pie chart of the interaction of students with classmates during Online class.

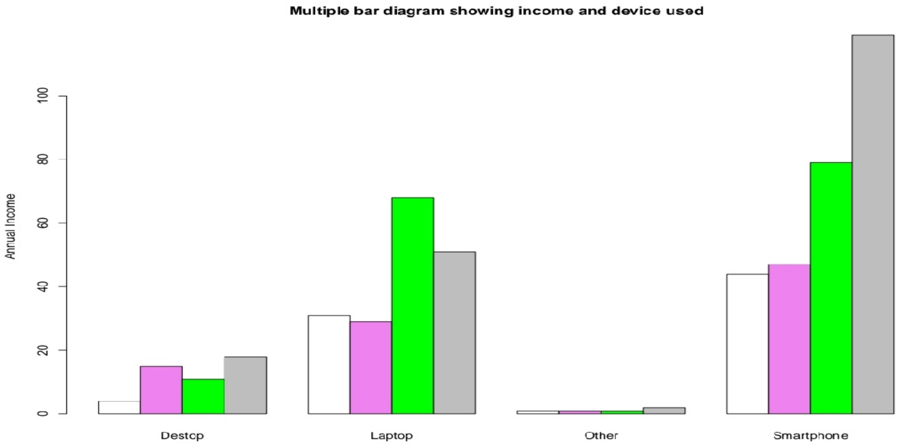


7. Pie chart and of the interaction of students with classmates during face-to-face class.



8. Multiple bar diagram of mental satisfaction during online class and face-to-face class.

9. Multiple bar diagram of annual income and device used.



5.2 DATA ANALYSIS

1. Test the independence of interaction with classmates during online and face-to-face classes.
H0 : Interaction with classmates during online and face-to-face class are independent.
Pearson's Chi-square Test

x-squared = 522.18, df =9, p-value = 0.00225
Since the p value is 0.00225, which is less than 0.05, we reject the hypothesis at 5% level of significance. Hence interaction with classmates during online and face-to-face classes are dependent.

2. Test the independence of mental satisfaction during online class and Class/Course.
H0 : Mental satisfaction during online class and Class/Course are independent.
Pearson's Chi-square Test

x-squared = 25.728, df =12, p-value = 0.01173
Since the p value is 0.01173 , which is less than 0.05, we reject the hypothesis at 5% level of significance. Hence mental satisfaction during online class/course are dependent.

3. Test the independence of mental satisfaction during face-to-face class and Class/Course .
H0 : Mental satisfaction during face-to-face class and Class/Course are independent.
Pearson's Chi-square Test

x-squared = 25.837, df =12, p-value = 0.01132
Since the p value is 0.01132 , which is less than 0.05, we reject the hypothesis at 5% level of significance. Hence mental satisfaction during face-to-face and class/course are dependent.

4. Test the independence of mental satisfaction during online class and Gender score.
H0 : Mental satisfaction during online class and gender score are independent.
Pearson's Chi-square Test

x-squared = 12.733, df = 6, p-value = 0.04748
Since the p value is 0.04748 , which is less than 0.05, we reject the hypothesis at 5% level of significance. Hence mental satisfaction during online class and gender score are dependent.

5. Test the independence of mental satisfaction during face-to-face class and Gender score.
H0 : Mental satisfaction during face-to-face class and gender score are independent.
Pearson's Chi-square Test

x-squared = 6.1975, df = 6, p-value = 0.4014
Since the p value is 0.4014 , which is greater than 0.05, we accept the hypothesis at 5% level of significance. Hence mental satisfaction during face-to-face class and gender score are independent.

6. Test the independence of interaction with teacher during face-to-face and online class.
H0 : Interaction with teacher during face-to-face and online class are independent.
Pearson's Chi-square Test

x-squared = 467.91, df = 9, p-value = 0.00225
Since the p value is 0.00225 , which is less than 0.05, we reject the hypothesis at 5% level of significance. Hence interaction with teacher during face-to-face and online class are dependent.

7. To test the mental satisfaction scores during online class and face-to-face are same.

H0 : Mental satisfaction scores during online class and face-to-face are same.
Pearson's product moment correlation

t = -18.297 , df = 519 , p-value = 0.00225
alternative hypothesis : true correlation is not equal to 0
95 percentage confidence interval : -0.6757458 -0.5710039
Sample estimates : cor -0.626192

8. To test the interaction with teacher during online class and face-to-face class are same.

H0 : Interaction with teacher scores during online and face-to-face class are same.
Pearson's product moment correlation

t = -17.715 , df = 519 ,p-value = 0.00225
alternative hypothesis : true correlation is not equal to 0
95 percentage confidence interval : -0.6647056 -0.5573396
Sample estimates : cor -0.6138534

9. To test the interaction with classmates during online class and face-to-face class are same.

H0 : Interaction with classmates during online and face-to-face class are same.
Pearson's product moment correlation

t = -13.572 , df = 519 ,p-value = 0.00225
alternative hypothesis : true correlation is not equal to 0
95 percentage confidence interval : -0.5725351 -0.4454860
Sample estimates : cor -0.5118035

10. One-way analysis of means for Face-to-Face

H01 : All four educational groups have the same average mean satisfaction during face-to-face class.

Source	Degrees of freedom	Sum of squares	MSS	F calculated	p-value	Inference
b/w group	1	5.4	5.35	4.303	0.0385	Reject H_{01}
Within group	519	645.3	1.243			

Since p-value = 0.0385, which is less than 0.05, we reject the null hypothesis at 5% level of significance. That is all four educational groups have different average mean satisfaction score during Face-to-Face class.

11. One-way analysis of means for online

H02 : All four educational groups have the same average mean satisfaction score during Online Class.

Source	Degrees of freedom	Sum of squares	MSS	F calculated	p-value	Inference
b/w group	1	10.1	10.21	8.2	0.00436	Reject H_{02}
Within group	519	640.5	1.234			

Since p-value = 0.00436 , which is less than 0.05 , we reject the null hypothesis at 5% level of significance. That is all four educational groups have different average mean satisfaction score during online class.

Chapter 6

SUMMARY AND CONCLUSION

Interaction of students with classmates during face-to-face are dependent. Mental satisfaction during online class and class/course is dependent. Mental satisfaction during face-to-face and class/course is also dependent. Mental satisfaction scores during online classes and gender scores are dependent whereas mental satisfaction during face-to-face classes and gender scores are independent. Interaction with the teacher during face-to-face and online class are dependent. All four educational groups have the same average mean satisfaction score during face-to-face class. All four educational groups have different mental satisfaction score during online class.

REFERENCES

[1] Das M.N and Giri N.C (1984). Design and Analysis of Experiments, Wiley, New York.

[2] V.K Rogatgi (1984). An introduction to probability and mathematical statistics, Wiley, Eastern limited, New Delhi.

[3] .R Kothari. Research Methodology (Methods and Techniques), New Age International publishers

[4] S.C Gupta and V.K Kapoor. Fundamentals of Mathematical Statistics, 10th edition, Published by Sultan Chand and Sons, New Delhi

[5] Allen, M., Bourhis, J., Burrell, N. & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education. The American Journal of Distance Education, 16(2). 83-97.

[6] Aragon, S., Shaik, N., Palma-Rivas, N. & Johnson, S. (1999). Comparative Analysis of Online vs. Face-to-Face Instruction. In Proceedings of WebNet World Conference on the WWW and Internet 1999 (pp. 581-586). Chesapeake, VA: AACE.