IOSR Journal of Dental and Medical Sciences (IOSR-JDMS)
e-ISSN: 2279-0853, p-ISSN: 2279-0861.Volume 22, Issue 3 Ser.15 (March. 2023), PP 14-21
www.iosrjournals.org

Importance of Statistics in Dental Research

Dr. Sehvish Qureshi, *PG student MDS, Department of prosthodontics, The Oxford Dental College*
Dr. Malathi Dayalan, *Head of Department of Prosthodontics, The Oxford Dental College*
Dr. Mohammed Sadiq Kazi, *PG student MDS, Department of prosthodontics, The Oxford Dental College*

ABSTRACT: *Statistical education is not based solely on formulas and figures, but also interpretation of data so that decisions can be applied correctly. The sound dental research, as well as the ability to understand published results of research, increasingly depends on a clear understanding of the fundamental of statistical design and analysis. As the dental researchers become more sophisticated in their statistical ability and knowledge to analyse data and choose most appropriate and powerful techniques from the vast array of modern biostatistical procedures the dental practitioners should improve their quantitative skills, although didactically unrelated to their dental training, to understand and interpret the increasingly complex statistical procedures that will be employed to analyse raw data from rapidly changing patterns of oral and concomitant research production and publication in order to keep abreast in their fields. The article covers some fundamentals of biostatistics and their application in dental research.*

KEYWORDS: *Statistics, hypothesis, normal distribution, parametric tests, non-parametric tests.*

Date of Submission: 11-03-2023 Date of Acceptance: 25-03-2023

I. INTRODUCTION

Statistics is a branch of science that deals with the collection, organization, analysis of data and drawing of inferences from the samples to the whole population. According to Kim and Dailey, Statistics is defined as the field of mathematical sciences that deals with data. Besides, "Biostatistics is a branch of statistics that emphasizes the statistical applications in the biomedical and health sciences (including Dentistry). It is concerned with making decisions under uncertainties that occur when the data are subjected to variation".[11] This requires a proper design of the study, an appropriate selection of the study sample and choice of a suitable statistical test. An adequate knowledge of statistics is necessary for proper designing of an epidemiological study or a clinical trial. Improper statistical methods may result in erroneous conclusions which may lead to unethical practice. Statistical procedure corresponds to these steps: Data collection, Statistical data analysis and its presentation.[6]

In daily practice, professionals know that their work is not based on the use of statistics, mainly because they are accustomed to working with people in a one-on-one setting, as opposed to with a collection of individuals. This circumstance risks limiting these professionals from developing research in their area. It is important to highlight that the researchers, having some notion of statistical procedures, can understand and manage data variability which is often produced by "uncertainty" and "uncertainties" which are often health problem consequences. Statistics are important in part because they support academic research at different levels and subsequently generate rigorous knowledge for the benefit of society.[7] Statistics are important in part because they support academic research at different levels and subsequently generate rigorous knowledge for the benefit of society. Currently, there is a growing interest in statistical knowledge because many authors are convinced that statistics must be understood through inductive and deductive reasoning. Statistical education is not based solely on formulas and figures, but also interpretation of data so that decisions can be applied correctly.[8] The sound dental research, as well as the ability to understand published results of research, increasingly depends on a clear understanding of the fundamental of statistical design and analysis.

DESCRIPTIVE ANALYSIS: Statistical analysis determines, mathematically and with high probability, whether an observed outcome occurred because of a real factor (an intervention or an exposure to risk) or simply by chance. A variable is any condition that can vary or change in quantity or quality. Before choosing any statistical analysis, it is necessary to know that data obtained during the experiment can be better understood when the variables are classified correctly, and that the type of variable being studied primarily determines the appropriate analytical method to be carried out. It can be independent variable or dependent, discrete or continuous, alternative (binary) or non-alternative variable. The independent variable, or treatment, is under the control of and administered by the experimenter. The behaviour that is potentially affected by the treatment and that we measure is called the

dependent variable. It is referred to as dependent because a change in it depends on the effect of the independent variable.

Discrete variable has limited or countable number of values and the basic unit of measurement cannot be meaningfully subdivided eg; number of dental students in a class, number of daily admissions of patients in hospital etc. whereas continuous variable has an infinite number of possible values and its basic unit of measurement can be meaningfully subdivided.eg: height, weight, skull circumference, meter is a unit of length; it can be subdivided into centimetres, millimetres etc. Alternative (Dichotomous or binary) data represent measurable categories in that outcome can take only one of two values: yes or no, "improved/not improved" and "completed task/failed to complete task." Non-Alternative data represent measurable categories in that outcome can take many values. Example is severity of disease level (mild, moderate, severe)

Data and their types

Data can be broadly classified into-

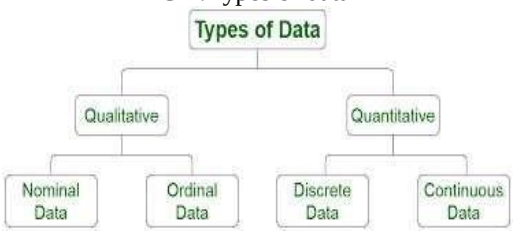
Qualitative Data: • Measuring a characteristic for which there is no natural numeric scale (can be subdivided into nominal and ordinal data) E.g: Gender, Eye colour, a child may or may not show evidence of dental caries at a particular moment in time. In this case the observation describes the presence or absence of a characteristic, and it is therefore qualitative rather than quantitative.

Quantitative Data: • There is a natural numeric scale (can be subdivided into interval and ratio data) E.g: height, weight age etc.

Scales of Measurement: The scale of measurement has implications for the way information is displayed and summarized and determines the statistical methods for analysing the data. There are 3 scales of measurements in statistics:

- Nominal (classificatory)
- Ordinal (ranking)
- Numerical (Interval and Ratio scales)

FIG 1:Types of data



Populations and samples: When conducting a study, it is necessary to define the target population (all patients who may be eligible for the study). However, it is often not possible to observe the whole target population. A study population is then defined as a subset of the target population (e.g., patients with denture stomatitis within a specified period).

Sample Surveys

Some medical data are routinely collected, such as numbers of births and deaths, and for these measures the behaviour of the whole population is known to a fair degree of accuracy. However, when the data are not routinely collected, it is usually both difficult and costly to study the whole population, so members are selected to provide a sample. The sample needs to be chosen in such a way that it is representative of the population. The primary techniques are:

- face- to- face interview of participants (patients or dentists) at the dental practice/hospital,
- telephone interviews
- and postal or Internet- based surveys.

Depending on the nature of the topic, opportunistic face- to- face patient interviews can achieve close to 100% response (Porter 2006), postal surveys with reminders can attain 60% response (McKernan et al. 2015), whereas Internet- based surveys tend to have low response rates (De Gregorio et al. 2015), typically around 25%.

METHODS OF SAMPLING

A. Non probability sampling techniques

1. **Convenience Sampling**: This type of sample is often collected in a busy town centre. Its advantages are that collection can be completed quickly and cheaply. Hence it is a common method of sampling for student projects.

2. **Purposive Sampling**: This is also known as judgmental sampling. Recognising that the population may well contain different types of individuals, with differing measures and ease of access, the experimenter exercises deliberate subjective choice in drawing what he regards as a 'representative' sample.
3. **Quota Sampling**: Here, judgment and convenience are combined, and quota sampling is usually more structured than straight accessibility or judgmental sampling.

B. Probability Sampling Techniques

1. **Simple Random Sampling**: A set of random numbers is generated and the associated individuals are selected from the list.
2**. Systematic Sampling**: The sampling frame is divided into contiguous blocks of, perhaps, five, 10 or 20 patients. The initial individual selected is chosen from the first block at random. Subsequent individuals are then taken from the equivalent position in subsequent blocks.
3. **Stratified Random Sample**: The population is first divided into strata (population groups), based on a particular characteristic such as age or sex (two or more characteristics may be used in larger studies). A simple random sample is then selected from each stratum.
4. **Cluster Sampling**: A list of all the individuals in the study population may not exist, but a list of larger units, such as households, may be available. A random selection of households (for instance) rather than individuals is taken. The people in these households form the sample.
5. **Multistage Random Sampling**: Random sampling takes place at two or more levels. For instance, obtaining a random sample of all Indian dental students would involve contacting all of the institutions concerned. Instead, one could draw a random sample of Indian dental schools, contact only those schools, and take a random sample of students from within each chosen institution.

TABLE 1: OVERVIEW OF TYPES OF SAMPLING METHODS USED IN PROSTHODONTIC RESEARCH

Type of Sampling	Uses	Example
Simple Random Sampling	• Used for homogenous population which is readily available. • Haphazard way of eliminating bias Each unit has chance of being selected.	Number of edentulous persons in the entire hospital being selected.
Stratified Random Sampling	• Used for collecting data from large heterogenous population and dividing it into homogenous sub groups having similar characteristics. Subgroups are though not equally distributed in the population.	Number of edentulous persons among different age groups. OR Number of edentulous persons amongst the high and low socioeconomic strata of population
Cluster Sampling	• Used for a larger population having natural groups like elderly in a village, children of a school • Clusters are heterogenous, but Different cluster are homogenous.	Number of edentulous persons in the total population of elderly in the village X or Y
Convenience Sampling	• Uses readily available data Easy to use but not recommendable	Number of edentulous persons coming across when on to street to select them.
Quota Sampling	A quota or a proportion for inclusion of a particular group is determined by some criteria and within this group anyone is selected.	Number of edentulous persons among pensioner and no pensioners.

Sample is a fraction of the universe. Studying the universe is the best parameter. But, when it is possible to achieve the same result by taking fraction of the universe, a sample is taken. Applying this, we are saving time, manpower, cost, and at the same time, increasing efficiency. Hence, an adequate sample size is of prime importance in biomedical studies. If sample size is too small, it will not give us valid results, and validity in such a case is questionable, and therefore, whole study will be a waste. Furthermore, large sample requires more cost and manpower. It is a misuse of money to enroll more subjects than required. A good small sample is much better than a bad large sample. Hence, appropriate sample size will be ethical to produce precise results. Factors Influencing Sample Size Include: 1. Prevalence of event or characteristics- If the prevalence is high, small sample can be taken and vice versa. If prevalence is not known, then it can be obtained by a pilot study. 2. Probability level considered for accuracy of estimate- If we need more safeguard about conclusions on data, we need a larger sample. Hence, the size of sample would be larger when the safeguard is 99% than when it is only 95%. If only a small difference is expected and if we need to detect even that small difference, then we need a large sample. 3)Availability of money, material, and manpower. 4)Time bound study curtails the sample size as routinely observed with dissertation work in post graduate courses.

SAMPLE SIZE DETERMINATION AND VARIANCE ESTIMATE

To calculate sample size, the formula requires the knowledge of standard deviation or variance, but the population variance is unknown. Therefore, standard deviation has to be estimated. Frequently used sources for estimation of standard deviation are:

i. A pilot or preliminary sample may be drawn from the population, and the variance computed from the sample may be used as an estimate of standard deviation. Observations used in pilot sample may be counted as a part of the final sample.

ii. Estimates of standard deviation may be accessible from the previous or similar studies, but sometimes, they may not be correct.

CALCULATION OF SAMPLE SIZE

Calculation of sample size plays a key role while doing any research. Before calculation of sample size, following five points are to be considered very carefully. First, we have to assess the minimum expected difference between the groups. Then, we must find out standard deviation of variables. Now, set the level of significance (alpha level, generally set at $P < 0.05$) and Power of study (1-beta = 80%). After deciding all these parameters, we must select the formula from computer programs to obtain the sample size. Various software are available free of cost for calculation of sample size and power of study. Lastly, appropriate allowances are given for non-compliance and dropouts, and this will be the final sample size for each group in study.

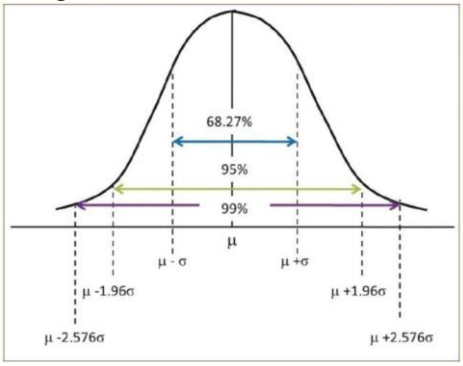
Power Of the Study

It is a probability that study will reveal a difference between the groups if the difference exists. A more powerful study is required to pick up the higher chances of existing differences. Power is calculated by subtracting the beta error from 1. Hence, power is (1-Beta). Power of study is very important while calculation of sample size. Power of study can be calculated after completion of study called as posteriori power calculation. This is very important to know whether study had enough power to pick up the difference if it existed. Any study to be scientifically sound should have at least 80% power. If power of study is less than 80% and the difference between groups is not significant, then we can say that difference between groups could not be detected, rather than no difference between the groups. In this case, power of study is too low to pick up the existing difference. It means probability of missing the difference is high and hence the study could have missed to detect the difference. If we increase the power of study, then sample size also increases. It is always better to decide power of study at initial level of research.

Normal and non-normal distribution of quantitative data

The quantitative data may be distributed in various possible ways. If data are symmetrically distributed on both sides around the mean and form a smooth, bell-shaped curve with a central 'hump' with two equal 'tails' at either side, the distribution of data is called normal or Gaussian (Figure 1); i.e the greatest amount of data values is symmetrically clustered in the centre, around the mean, and the smaller data values, on the right and left tails thus, when the density of distribution follows a unimodal, symmetrical, and bell-shaped distribution, it is called normal distribution. This type of distribution is purely affected by the number of observations in each frequency / class interval, which in turn is affected by the sample size. There could be conditions where the distribution could be bi-modal or asymmetric, and such distributions are said to follow Non-normal Distributions. The data is said to be skewed when an otherwise normal distribution has a long tail to the right or the left.

Fig 2: Normal distribution curve



When conducting a study, there is a chance for error and we cannot argue the results with 100% authority. Hence, the allowable error is assumed to be 5% and we tend to summarize the credibility of the results within a 95 % probability. This 5% error by chance gets divided on either side of this bell-shaped curve, beyond $m \pm 1.96\sigma$.

When the data do not follow normal distribution, then the data can be transformed by taking a logarithm or simply using the statistical methods to test the non-normal data. It is important to note here that the use of the mean along with its SD can be used for statistical analysis only if the sample size is large enough to have a 'normal distribution'.

Normal distribution may serve as a better predictor when the results of a study are to be generalized to the entire population, for example, the longevity of class V composite restoration in treating non-carious cervical lesions (NCCL). However, statistical inference from non-normal data is not obsolete, especially when the results contribute to the understanding of a part of a material or mechanism, and where generalizability to the entire population is not considered; for example, flexural strength of a micro filled composite resin.

Here the flexural strength is a part of the mechanical properties, which can indirectly influence the clinical outcome. Hence, if the research question is to test the clinical performance of micro filled composite resin in NCCL, then phase I of the study would be to test the mechanical properties of different composite resins where the sample size is less critical. Phase II of this study would be to narrow down to two to three composite resins from phase I and test them in a clinical trial, where sample size becomes mandatory.

The term "normal range" is used for the range of a continuous variable (e.g., salivary flow rate) within which we expect measurements for the majority (usually 95% or 90%) of "normal" or disease- free people to be found. It does not mean that the variable has a Normal distribution, but if it does, 95% of the distribution will lie within two standard deviations of the mean or 90% within 1.64 standard deviations.

In other words, the 95% normal range is from the mean minus two standard deviations to the mean plus two standard deviations. Note that the term "disease- free" is usually specific to the condition under consideration; it does not necessarily imply that the individual has excellent overall health. For non- Normal distributions, centiles (sometimes called percentiles) are calculated to estimate the normal range. Centiles: The 10th centile is the number for which 10% of the data has a lower value, the 5th centile is the number for which 5% of the data has a lower value; the 95th centile has 95% of the data below it and 5% above it. The 50th centile or median is a type of average based on the middle value of the ordered observations. In order to work out the 95% normal range, the 2.5% and 97.5% points of the distribution would need to be estimated (this would give tails of equal size – 2.5%).

Frequently used Statistical test: Broadly statistical tests are divided into two groups:

- Parametric
- Non parametric

Selection of appropriate statistical tests:

Selection of appropriate statistical test is very important for analysis of research data. Use of wrong or inappropriate statistical test is a common phenomenon observed in articles published in many medical journals (study by Jay karan 2011) Wrong statistical tests can be seen in many conditions like use of paired test for unpaired data or use of parametric statistical tests for the data which does not follow the normal distribution or incompatibility of statistical tests with the type of data, etc. Because of the availability of different types of statistical software, performing the statistical tests become easy, but selection of appropriate statistical test is stil

Following are the different parametric test used in analysis of various types of data.

1) **Student's 't' Test**

Mr. W. S. Gosset, a civil service statistician, introduced 't' distribution of small samples and published his work under the pseudonym 'Student.' This is one of the most widely used tests in pharmacological investigations, involving the use of small samples. The 't' test is always applied for analysis when the number of sample is 30 or less. It is usually applicable for graded data like blood sugar level, body weight, height etc. If sample size is more than 30, 'Z' test is applied. There are two types of 't' test, paired and unpaired.

2) **ANOVA** When we want to compare two sets of unpaired or paired data, the student's 't' test is applied. However, when there are 3 or more sets of data to analyse, we need the help of well-designed and multi-talented method called as analysis of variance (ANOVA). This test compares multiple groups at one time.

a) **One way ANOVA** It compares three or more unmatched groups when the data are categorized in one way. For example, we may compare a control group with three different doses of aspirin in rats. Here, there are four unmatched group of rats

b) **Two-way ANOVA** Also called two factors ANOVA, determines how a response is affected by two factors. For example, you might measure a response to three different drugs in both men and women

Following are the non-parametric tests used for analysis of different types of data.

1)Chi-square test The Chi-square test is a non-parametric test of proportions. This test is not based on any assumption or distribution of any variable. This test, though different, follows a specific distribution known as Chi-square distribution, which is very useful in research. It is most commonly used when data are in frequencies such as number of responses in two or more categories. This test involves the calculations of a quantity called Chi-square (x2) from Greek letter 'Chi'(x) and pronounced as 'Kye.' It was developed by Karl Pearson

Applications

a) Test of proportion: This test is used to find the significance of difference in two or more than two proportions.

b) Test of association: The test of association between two events in binomial or multinomial samples is the most important application of the test in statistical methods. It measures the probabilities of association between two discrete attributes.

2**) Wilcoxon-Matched-Pairs Signed-Ranks Test** This is a non-parametric test. This test is used when data are not normally distributed in a paired design. It is also called Wilcoxon-Matched Pair test. It analyses only the difference between the paired measurements for each subject.

3) **Mann-Whitney test** It is a Student's 't' test performed on ranks. For large numbers, it is almost as sensitive as Student's 't' test. For small numbers with unknown distribution, this test is more sensitive than Student's 't' test.

4) **Friedman test** This is a non-parametric test, which compares three or more paired groups. In this, we have to rank the values in each row from low to high.

5) **Kruskal-Wallis test** It is a non-parametric test, which compares three or more unpaired groups. Non-parametric tests are less powerful than parametric tests. Generally, P values tend to be higher, making it harder to detect real differences.

Table 2: Parametric tests (follow normal distribution and normal curve)

Statistical test	Description	Example
Student t test	Investigates whether the expected values of two groups are the same. The test can be used for paired or unpaired groups. **Unpaired t test**-Difference between means of two unpaired samples. **Paired t test**-Is applied to paired data of observations from one sample only when each individual gives a paired of observations	• To compare electromyographic data of two groups of patients (open and deep skeletal bite) before and after surgical orthodontic treatment
Z test	Similar to 't' test in all aspect except that the sample size should be > 30	-
ANOVA	Useful for comparison of means of several groups. Is an extension of student's 't' test for more than two groups	
One way ANOVA	Under one way ANOVA, we consider only one factor and observe the mean of single variable	Comparison of pharyngeal airway volume among different vertical skeletal patterns-high, medium and low angle cases
Two way ANOVA	Two way ANOVA is used to study the impact of two factors in different groups(3 or more).	-To evaluate the length and orientation of masseter in 3 different types of malocclusions quantitatively using Cone Beam Computed Tomography (CBCT).

Table 3:

NON-PARAMETRIC TESTS - are applicable to almost all types of distribution

Statistical test	Description	Example
Chi square test	• To test association between two events in binomial or multinomial samples • As a test of goodness of fit The test is most commonly used on data that are presented in a 2 X 2 table of frequencies.	To evaluate the subjects perceived satisfaction of their dental appearance in different groups. Patient response to vaccum formed splint compared to heat cured acrylic splint.
Fisher's exact test	Fisher's exact test is used for small samples or when the expected cell frequency is <5 subjects	-
McNemar test	Similar to Fischer exact test but for paired samples	-
Mann Whitney U test (Wilcoxon's rank sum test)	It is a alternative to student 't' test & requires at least ordinal or nominal measurement	Evaluation of subgingival plaque by Plaque index in pregnant and non pregnant women
Wilcoxon's signed rank test	Test is used for paired comparison	Speech evaluation before and after insertion of denture.
Kruskal Wallis test	To compare more than two unpaired samples	Quality assessment of colour stability of different bleaching techniques (4 techniques) on vital tooth.

Table 4;

Examples of parametric and non-parametric tests :

S.No	Objective	Parametric test	Non parametric test
1.	To compare difference between two groups	Unpaired t test	Wilcoxon rank sum test (Mann Whitney U test)
2.	To test difference between paired observation	Paired t test	Wilcoxon signed rank test
3.	To compare difference between several groups	One way ANOVA	Kruskal Wallis test
4.	To compare group values on two variables	Two way ANOVA	Friedmann test
5.	To measure association between two variable	Pearson's correlation coefficient	Spearman's correlation coefficient

Inferential analysis: hypothesis tests, and confidence intervals

Testing a Hypothesis:

This involves assuming about the parameter and checking the plausibility of that assumption using sample data. We can also identify two types of research procedures with reference to hypothesis testing:

1. Comparison of two or more populations, e.g., patients with disease A and disease B.

2.Relationships between two or more factors in a population, e.g., smoking and denture stomatitis.

In hypothesis testing also, we must deal with sampling variation as most often we have to make an inference based on samples. The process of statistical inference involves the following principles. We must determine the sampling variation when the samples are drawn from statistically identical populations, or determine the correlation between factors which the samples are likely to reveal when, in the population, they are not related. Only if the observed sample difference or correlation is larger than that which the sampling variation would produce, can we conclude that differences (or correlation) exist in the populations. Compare the actual difference that we observe in our experiment with the corresponding sampling distribution, under the null hypothesis. If the probability of observing that and greater differences in the sampling distribution is small, reject the null hypothesis. In this case, we say that the difference between the two treatments is significant. This means that the samples came from statistically different populations.Usually, a probability of 5% or less is considered as rare or small and sometimes 1% is used. On the other hand, if the probability of observing that and greater differences as seen from the sampling distribution is large, we do not reject the null hypothesis, and say that the difference is not significant. This means that the difference that is observed in the actual experiment is not big enough to reject the null hypothesis.

While drawing conclusions as stated above, we are likely to commit two types of errors. These two types of errors are involved in the test procedure itself. We may conclude that the difference is significant when in fact there is no real difference in the population; and so, reject the null hypothesis when it is true. This error is called Type I error whose magnitude is usually denoted by the Greek letter a. On the other hand, we may give our final verdict that the difference is not significant when in fact there is real difference between the populations; that is, the null hypothesis is not rejected when it is false. This error is called the error of the second kind or Type II error whose magnitude is denoted by beta.

The P- value is the probability of obtaining at least the difference observed between the sample and null hypothesis values of a measure (e.g., the mean), assuming that the null hypothesis is true. Usually, it is the magnitude of the difference (regardless of whether it is positive or negative) that is of concern. The smaller the P-value, the more evidence there is against the null hypothesis. For the case of two groups where the P- value is small (e.g., 0.001), under the null hypothesis a difference between the groups of at least the size observed in the sample has a small probability; there is strong evidence against the null hypothesis. For larger P- values (e.g., 0.6), the evidence against the null hypothesis is weak because the difference between the two groups is small.

Confidence intervals :Confidence intervals (CI) are also used in hypothesis testing. There is a close relationship between confidence intervals and significance tests. Confidence limits are two extremes of a measurement within which 95% observations would lie. These describe the limits within which 95% of the mean values if determined in similar experiments are likely to fall. The value of 't' corresponding to a probability of 0.05 for the appropriate degree of freedom is read from the table of distribution. By multiplying this value with the standard error, the 95% confidence limits for the mean are obtained as per formula below.

Lower confidence limit = mean - ($t_{0.05}$ × SEM)

Upper confidence limit = mean + ($t_{0.05}$ × SEM)

If n > 30, the interval M ± 2(SEM) will include M with a probability of 95% and the interval M ± 2.8(SEM) will include M with probability of 99%. These intervals are, therefore, called the 95% and 99% confidence intervals, respectively. The important difference between the 'p' value and confidence interval is that confidence interval

represents clinical significance, whereas 'p' value indicates statistical significance. Therefore, in many clinical studies, confidence interval is preferred instead of 'p' value, and some journals specifically ask for these values. Various medical journals use mean and SEM to describe variability within the sample. The SEM is a measure of precision for estimated population mean, whereas SD is a measure of data variability around mean of a sample of population. Hence, SEM is not a descriptive statistic and should not be used as such. Correct use of SEM would be only to indicate precision of estimated mean of population. Most authors recommend the regular reporting of CI because simply mentioning the p value by itself is not sufficient, as it does not provide information on the size of the effect. CI provide more meaningful evidence on the magnitude of the effect because they do not only contain information from p values, but additionally demonstrate the direction of the treatment effect, the size of the effect estimate, and its degree of precision. Therefore, both p and CI should be reported.

II. CONCLUSION:

Following the rapid growth of computer technologies, digital innovation has been driven to be a global imperative. The development of computing and informatics has a major impact on the role of statistics, in which statistical approaches have become more robust and effective. The advancement of data analytics recently plays a significant role in a variety of areas including healthcare. For instance, prevention of errors in diagnosis and treatment, following the improvement of decision-making process, seems to be one of the important benefits of data analytics. Therefore, the implementation of statistics is unavoidable in healthcare professions including dental education. It is important that a researcher knows the concepts of the basic statistical methods used for conduct of a research study. This will help to conduct an appropriately well-designed study leading to valid and reliable results. Inappropriate use of statistical techniques may lead to faulty conclusions, inducing errors and undermining the significance of the article. Bad statistics may lead to bad research, and bad research may lead to unethical practice. Hence, an adequate knowledge of statistics and the appropriate use of statistical tests are important. An appropriate knowledge about the basic statistical methods will go a long way in improving the research designs and producing quality dental research which can be utilised for formulating the evidence-based guidelines. Hence, it's been well stated by

Karl Pearson that "After all, high statistics are numerical expression of common sense."

BIBLIOGRAPHY

[1]. Research Methods for Clinical Therapists: Applied Project Design and Analysis. by Carolyn M. Hicks, 3RD Edition.
[2]. Research methodology methods and techniques by C R Kothari, third edition.
[3]. Introduction to Biostatistics and Research Methods, P.S.S. Sundar Rao, 5th edition.
[4]. Dental Statistics Made Easy, Nigel C, Third Edition.
[5]. Biostatistics A Methodology for the Health Sciences, Gerald Van Belle Lloyd D. Fisher, Second Edition.
[6]. Navarro P, Alemán I, Sandoval C, Matamala C, Corsini G. Statistical Testing Methods for Data Analysis in Dental Morphology. Int J Morphol. 2020 Oct 1;38(5).
[7]. Ali Z, Bhaskar SB. Basic statistical tools in research and data analysis. Indian J. Anaesth. 2016 Sep;60(9):662.
[8]. Omar Abdulmohsin Ali, Advanced Biostatistics in Dentistry. Research gate publication. 2017 March.
[9]. Singh AS, Masuku MB. Sampling techniques & determination of sample size in applied statistics research: An overview. Int.J.economics.commerce manag. 2014 Nov;2(11):1-22.
[10]. Zodpey, S P (2004): Sample size and power analysis in medical research, Indian J. Dermatol. Venereol.Leprol, 70 (2), 123-128.
[11]. Garrocho-Rangel JA, Ruiz-Rodríguez MS, Pozos-Guillén AJ. Fundamentals in Biostatistics for Research in Pediatric Dentistry: Part I - Basic Concepts. *J Clin Pediatr Dent*. 2017;41(2):87-94