We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,300

Open access books available

172,000

International authors and editors

190M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Probability and Sampling in Dentistry

*Yasser Riaz Malik, Muhammad Saad Sheikh
and Shakeela Yousaf*

Abstract

Probability and sampling in dentistry are two fundamentals which have great importance in clinical research. Many research works in dentistry shows lack of proper understanding and use of these two factors. The definition of probability is incredibly significant in daily life. Statistical analysis is based on this particularly useful definition. In fact, the function of probability in modern science is that of substituting for certainty. Probabilities are numbers that represent the probability that a specific occurrence will occur. We learn about the odds of many daily cases, ranging from weather predictions (probability of rain or snow) to lotteries (probability of winning a major jackpot). In biostatistical applications, probability theory underlies the statistical inference. Statistical inference means drawing generalizations or inferences on unknown population parameters. After selecting a sample from the population of interest, we calculate the characteristics under analysis, summarize the characteristics in our sample, and then draw inferences about the population based on what we find in the sample. Population and sampling are two critical aspects of study design. The population is a group of individuals who share common relations. A sample is a population subset. The size of the sample is the number of individuals in the sample. The more representative the sample of the population, the surer the researcher can be about the validity of the data. In this module, we will explore sampling methods, basic principles of probability, and applications of probability theory. The definition of probability is introduced, and the function of probability distributions is discussed in the statistical theory, with reference to the normal distribution and its characteristics. Sampling and sampling variations are defined, along with the sampling error, the standard error of the mean and the confidence intervals for determining the likely magnitude of the population mean. Medical study typically includes patients with an illness or disorder. The generalization of clinical research results is focused on several factors linked to the internal and external validity of the research methods. The sampling process is the key methodological problem that affects the generalizability of clinical research results. In this educational article, we also clarify the various methods of sampling in clinical research.

Keywords: probability, sampling, dentistry, restorative, statistical analysis, clinical research, biostatistics, confidence interval, standard error, probability, non-probability, normal distribution, sampling errors, sampling types, probability distribution

1. Introduction

The theory of probability was developed in the 17th century. It has got its origin from game of poker after a dispute. It led two famous French mathematicians, Blaise Pascal and Pierre de Fermat to create a theory of probability. Antoine Gombaud, Chevalier de Méré, a French nobleman with an interest in gaming and gambling questions, called Pascal's attention to an apparent contradiction concerning a popular dice game. The game consisted of tossing a pair of dice 24 times; the issue was determining whether to bet even money on the occurrence of at least one "double six" during the 24 spins. A seemingly well-established gambling law led de Méré to conclude that betting on a double six in 24 spins would be profitable, but the reverse was implied by his own estimates. Based on questions posed by de Méré, Pascal began to correspond with his friend Pierre Fermat about these problems, in which the basic concepts of probability theory were drawn for the first time. While a few special gambling problems had been resolved by some Italian mathematicians in the 15th and 16th centuries, no general theory had been developed before this famous correspondence. The Dutch scientist Christian Huygens, Leibniz's teacher, heard of this correspondence and shortly afterward (1657) wrote the first book on probability, *De Ratiociniis in Ludo Aleae*, a treatise on gambling-related problems. Thanks to the innate appeal of gambling, probability theory soon became popular and the subject developed swiftly in the 18th century. The key contributors to this time were Jakob Bernoulli (1654–1705) and de Moivre (1667–1754).

In 1812, Pierre de Laplace (1749–1827) applied several new theories and mathematical methods to his book, *Théorie Analytique des Probabilités*. Prior to Laplace, the theory of probability was concerned primarily with the development of a statistical study of gambling. Laplace has applied probabilistic principles to a variety of theoretical and practical problems. The theory of errors, actuarial mathematics, and statistical mechanics are examples of some of the main applications of probability theory that have been developed. Gornband took an initiation and an interest in this area in 1954. After him, many statistical authors have tried to reshape the idea of the former. The quest for a generally accepted meaning lasted almost three centuries and was marked by a great deal of controversy. The problem was eventually addressed in the 20th century by approaching probability theory on an axiomatic basis, by a Russian mathematician A. Kolmogorov [1].

2. Objective

This module would discuss the techniques of sampling, the fundamental concepts of probability, and the applications of probability theory. The role played by probability in modern science is that of a substitute for certainty, therefore definition of the probability is added, and the function of the probability distribution is discussed in the statistical theory, with reference to the normal distribution and its characteristics. Normal distribution is of great value in evaluation and research in both psychology and education, because it helps us to predict that where cases will fall within a distribution probabilistically. Variations of sampling and sampling types will be explained, along with the error of sampling, the standard error of the mean and the confidence intervals for determining the likely magnitude of the population mean. Generally, sampling allows researchers to obtain enough data to answer the research question(s) without having to query the entire population - saving time and money. For this reason the major types of probability sampling methods and their characteristics would be addressed.

3. Probability

“Probability” has become one of the key methods for statistics in dentistry. Often statistical analysis is paralyzed without theory of probability [2].

The probability of an event is its chance of occurrence, measured on a scale from 0 (never occurs) to 1 (always occurring). There are two views we may take of probability. One is that it is the long-term frequency of an event, e.g., in a long series of coin tosses, heads should occur about half of the time, so we write $P(\text{Head}) = 0.5$. The second and broader, view is that probability is a subjective measure of our belief in the chances of an event occurring, e.g., “I believe that there is a 30% chance of a practical AIDS vaccine by the year 2025”. Here, because the event can only occur once, there is no sensible long-term relative frequency view, but such subjective probabilities are useful when making decisions, e.g. planning future medical facilities. Of course, such subjective probabilities are most likely to be accurate when based on good long-term relative frequency information [3]. Predictions are taking the form of probabilities. We use probabilities to predict the likelihood of an earthquake, rain, or whether you are going to get an A in exams. Dentists use percentages to assess the risk of appliance that may trigger gum disease to use alternates. The investment schemes use the probability to determine the rate of return on the investment of the client. You could use the chance to decide whether to purchase a profitable commercial land. In your analysis of statistics, you will use the power of mathematics through probability calculations to evaluate and interpret your results [4].

4. Probability distributions

A probability distribution is a function that describes the probability of obtaining the possible values that a random variable may assume. In other words, the variable values differ depending on the distribution of the underlying probability. In Statistics, the probability distribution gives the possibility of each outcome of a random experiment or occurrences [5].

Suppose you take a random sample and measure the weight of the subjects. You will establish a distribution of weights when you calculate weights. This form of distribution is useful when you need to know the outcomes are most likely to occur, the spread of possible values, and the probability of various outcomes [6].

The probability distributions suggest the probability of an occurrence or result. Statistics use the following notation to define the probabilities:

$P(X)$ = the probability that the random variable will have a particular value of X .

A probability model needs a measure of the probability, typically written to P . This probability measure must allocate a probability to each case A , a probability $P(A)$.

We require the following properties:

1. $P(X)$ is always a nonnegative real number, between 0 and 1 inclusive.
2. $P(\emptyset) = 0$, i.e., if X is the empty set $\emptyset$, then $P(X) = 0$.
3. $P(S) = 1$, i.e., if X is the entire sample space S , then $P(X) = 1$.
4. P is (countably) additive, meaning that if $X_1, X_2, \dots$ is a finite or countable sequence of disjoint events, then $P(X_1 \cup X_2 \cup \dots) = P(X_1) + P(X_2) + \dots$.

The sum of all probabilities for all potential values must be equal to 1. In addition, the likelihood for a particular value or set of values must be between 0 and 1 [7].

The probability distributions define the dispersion of random variable values. Consequently, the type of variable determines the type of distribution of the likelihood [8]. For a single random variable, the statisticians split the distributions into the following two types:

- Discrete probability distribution
- Normal/Continuous distribution

4.1 Discrete probability distribution

A discrete distribution of probability is made up of discrete variables. Specifically, if a random variable is discrete, the probability distribution would be discrete. For a discrete probability distribution, each potential value of a discrete random variable can be correlated with a non-zero probability [9, 10]. Therefore, a discrete probability distribution is also provided in tabular form.

The following example shows that the probability (relative frequency) of someone randomly chosen from the operating theater staff being a non-smoker is 0.45 (Table 1).

Note that in the right column, the frequencies (counts) have been turned into relative frequencies (percentage). How you do this:

1. Count the total number of items. In this chart the total is 60.
2. Divide the count (the frequency) by the total number. For example, $27 / 60 = 0.45$ or $17 / 60 = 0.28$

4.2 Types of discrete distribution

According to the properties of data, there are other types of discrete probability distributions to display different kinds of data.

- a. Binomial distribution to model binary data.
- b. Poisson distribution to model count data.

4.2.1 Binomial distribution

The binomial distribution is a probability distribution that summarizes the probability that a value will take one of two impartial values under a given set of parameters or assumptions [11].

Smoking habit	Frequency	Relative frequency (%)
Non-Smoker	27	45
Ex-Smoker	17	28
Current smoker	16	27
Total	60	100

Table 1.
Smoking habit of maxillofacial surgery theater staff.

Consider a dichotomous variable, where the outcome for everyone is one of two types (A or B).

1. Sex at birth: M or F.
2. Disease status: diseased or not diseased.
3. Drug response: responded or did not respond.
4. Comparison of two treatments.

4.2.2 Poisson distribution

The Poisson distribution is another discrete distribution, i.e. it applies to a discrete variable. But unlike the binomial there is no strict upper limit to the possible values of the variable. The variable is the count of several independent events that occur randomly in a fixed interval or time or space [12, 13].

1. The number of radioactive emissions from a given source in a given time.
2. The number of particles in a given volume of fluid, samples from a well-mixed bulk quantity.
3. The number of cases of a particular disease that a doctor sees in a week.
4. The number of still births occurring in a hospital per month.

5. Normal distribution

Continuous probability functions are also known as probability density functions or *conjugate prior* [14]. We know that we have a continuous distribution if the variable can assume an infinite number of values between any two values. Continuous variables are often measurements on a scale, such as height, weight, and temperature. Unlike discrete probability distributions where each value has a non-zero likelihood, specific values in continuous distributions have a zero probability. For example, the likelihood of measuring a temperature that is exactly 28 degrees is zero [15, 16].

Just as there are different types of discrete distributions for different kinds of discrete data, there are different distributions for continuous data. Each probability distribution has parameters that determine the shape of the distribution. Most distributions have between 1 and 3 parameters [17]. Specifying these parameters sets out the structure of the distribution and all its probabilities entirely. These parameters reflect the fundamental characteristics of the distribution, such as the central tendency and the variability [18].

The most common is normal distribution which is often also referred to as the Gaussian distribution [19]. The normal distribution has two main features:

- It is symmetrical about its mean.
- It is bell shaped.

The normal distribution is the most important distribution in statistics. One reason is that many continuous variables, such as height, seems to have this distribution

Figure 1.
The standard Normal distribution (source: <https://statisticsbyjim.com/basics/normal-distribution>).

[20]. But the main reason is because of what is known as the central limit theorem. This theorem states that if a random sample is taken from any distribution then the distribution of the sample mean $\bar{x}$ will be approximately Normal. The approximation becomes better as n gets larger [21]. The implication of this result is that inferences, from sample to population, can be based on the Normal distribution [22].

A Normal distribution may be summarized by its mean μ and variance σ^2 . It is often necessary to be able to find the areas under specified parts, particularly the tails, of Normal distribution curve.

This can be done by referring to published tables, which are given in most statistical texts. In order to use these tables, it is necessary to standardize the variable x . This can be done by calculating a standardized Normal deviate z by the formula (**Figure 1**) [23]:

$$Z = (X - \mu) / \sigma \text{ (Standardized Normal Deviate SND)} \quad (1)$$

(where Z is the value on the standard normal distribution, X is the value on the original distribution, μ is the mean of the original distribution, and σ is the standard deviation of the original distribution.)

The standard score is the number of standard deviations above or below the mean where a given observation falls. For example, a standard score of 1.5 means that the observation is 1.5 standard deviations above the average. On the other hand, the negative score is below the average. The mean is a Z -score of 0.

6. Sampling

A population in any collection of individuals in which we may be interested, e.g. all people in Saudi Arabia, all females in Hail Region, all diabetic children in Sydney Hail City under 12 years of age. Usually the population is too large for us to examine every individual, so we take a *sample* from the population. If the sample is **representative** of the population, we can then make *inferences* about the population from the sample.

For example, in a study of the incidence of schistosomiasis in a particular region, the population would be all adults living in the region and might consist of many thousands of individuals. The sample might be a few hundred people from the total population, and we wish to be able to generalize from the sample to the population [24]. The advantage of studying just a sample is a saving of labour and costs. The disadvantage of a sample is that precision is lost by not observing the complete population. The sample mean is unlikely to equal exactly the population mean; that is, the sample estimate will have some error [25].

It is useful to distinguish two kinds of error: sampling errors and non-sampling errors.

7. Sampling errors

Sampling errors are statistical errors, they arise because only a fraction of population has been observed. Different samples will give different results. Sampling errors become less important as the sample size increases [26].

$$\text{Sampling Error} = Z \times \left(\sigma / \sqrt{n} \right) \tag{2}$$

(sampling Error is calculated by dividing the standard deviation of the population by the square root of the size of sample and then multiplying the resultant with the Z score value which is based on confidence interval).

Step by Step Calculation of Sampling Error:

Collected all the data set called the population. Calculate population mean and population standard deviation.

Now, it is important to decide the size of the sample, and further, the size of the sample must be less than the population, and it should not be larger.

Determine the confidence level and, as a result, the value of the Z score can be calculated from its table.

Now multiply the Z score by the population standard deviation and divide the same by the square root of the sample size in order to arrive at an error margin or an error in the sample size.

8. Non-sampling errors

They arise if the sampling procedure is not representative of the total population. Such errors do not necessarily decrease as sample size increases. Examples of this type of error are the failure to include people with no permanent home because their existence is not recorded, or the refusal of some individuals to participate in the study. These errors constitute *bias* [27].

9. Types of sampling methods

Methods of sampling are as follows:

9.1 Probability sampling

Probability Sampling is a method in which each member of the population has the same chance of being part of the sample. Probability or random sampling is the most impartial but can be the most expensive sample in terms of time and energy for a given amount of sampling error [28].

9.2 Simple random sampling

A simple random sample means that each case in the population has the same probability of being included in the sample. This approach is the most straightforward of all probability sampling methods, because it includes only a single random sample and requires no specialized knowledge of the population. Since randomization is used, any research conducted on this sample should have a high internal and external validity.

Simple random sampling can be difficult to implement in practice. There are some prerequisites for using this method:

- I.If chosen, you can contact or access each member of the population.
- II.You have the time and resources to collect the data from the appropriate sample size.
- III.Standard errors of estimators may be high and a complete framework (a list of all units in the whole population) is needed [29].

Simple random sampling works best if you have a lot of time and resources to carry out your study, or if you are studying a limited population that can be easily sampled (Figure 2).

9.3 Stratified sampling

Stratified sampling is where the population is divided into sub-groups and a random sample is collected from each sub-group.

Figure 2.
Simple random sampling (source: <https://www.datasciencemadesimple.com/simple-random-sampling-in-sas/>).

(Stratified sampling formula)

$$\text{Stratified random sampling} = \frac{\text{Total Sample size}}{\text{Entire population}} \times \text{Population of subgroups.}$$

Mostly used when the population is heterogeneous and includes a variety of different classes, some of which are related to the subject of the analysis. The advantage of stratified sampling that it ensures a high degree of representativeness of all strata or strata in the population. Disadvantage is that it increases the work for planning and analysis for keeping the uncertainty within an acceptable level (Figure 3) [30].

9.4 Cluster sampling

Cluster sampling is where the entire population is divided into clusters or groups. Subsequently, a random sample of these clusters is taken, all of which are used in the final sample [31, 32].
It is simple and convenient, but the downside is that the members of the groups can be different from each other, decreasing the efficiency of the techniques (Figure 4).

9.5 Systematic sampling

Systematic sampling is the selection of a sample on an orderly basis. To build a sample, look at the target population and choose every fifth, tenth, or twentieth name, based on the size of the sample.
Systematic sampling can be used by statisticians if they want to save time or are disappointed with the results obtained from a simple random sampling process. If a fixed starting point has been established, the statisticians choose a constant interval to encourage the selection of the participant.
It ensures a high degree of representativeness, and no need to use a table of random numbers. Disadvantage is that it is Less random than simple random sampling (Figure 5) [33].

Figure 3.
Stratified sampling–(source: <https://www.datasciencemadesimple.com/stratified-sampling-in-sas/>).

Figure 4.
Cluster sampling (source: <https://research-methodology.net/sampling-in-primary-data-collection/cluster-sampling/>).

Figure 5.
Systematic sampling (source: <https://www.wallstreetmojo.com/systematic-sampling/>).

9.6 Non-probability sampling

Non-probability Sampling is a process in which each member of the population has no fair chance of being chosen. If the researcher chooses to choose participants selectively, non-probability sampling shall be considered [34].

9.7 Quota sampling

Quota sampling is a non-random sampling method in which participants are selected based on predetermined characteristics such that the total sample will have the same distribution of characteristics as the larger population. Selection of participants who meet certain characteristics, such as gender, age, fitness, etc. Scientists will also go back over the preferred sample population to guarantee that subgroups are represented to the same degree as those in the wider population.

There are several drawbacks to quota sampling since a subgroup’s meanings are usually limited to a few characteristics. Other characteristics associated with the population could also become over-expressed (**Figure 6**) [35].

9.8 Purposive sampling

Pre-selected research hypotheses criteria are used to classify research subjects, such as a lung cancer study for individuals having repeated exposure to asbestos. That is where the researcher includes cases or participants in the study because they think they should be included. The purposive sampling approach may prove to be successful when only a limited number of people can serve as primary data sources due to the nature of the research design and objectives.

Purposive sampling is one of the most cost-effective and time-effective sampling methods available. The disadvantages of purposive sampling are vulnerability of errors of judgment by the researcher, poor level of reliability and high degree of bias and inability to generalize the results of study (**Figure 7**) [36].

9.9 Snowball sampling

Participants in the study are referred to the researcher by other individuals who match the characteristics needed for the study. This method is most applicable to

Figure 6.
Quota sampling (source: <https://deepai.org/machine-learning-glossary-and-terms/quota-sample>).

Figure 7.
Purposive sampling (source: <https://research-methodology.net/sampling-in-primary-data-collection/purposive-sampling/>).

small communities that are difficult to reach due to cultural professional or other reasons.

Snowball sampling is also sometimes referred to as chain-referral sampling. Snowball sampling is low-cost and easy to implement. Snowball sampling is low-cost and easy to implement. It does not require a research team to hire recruiters for the study since the initial subjects act as the recruiters who bring in additional subjects. The disadvantage is that sampling bias can occur because, when initial subjects recruit additional subjects, it is possible that several subjects may share similar characteristics or characteristics that may not be representative of the broader population under research (Figure 8) [37].

9.10 Convenience sampling

Convenience sampling appears to be a favorite sampling technique among students, as it is cheap and simple compared to other sampling techniques. It is collection of participants since they are often readily and conveniently available.

Convenience sampling is not the preferred form of sampling for successful research as samples are taken from a particular segment of the population, so the degree of generalizability is questionable (Figure 9) [38].

Figure 8.
Snowball sampling (source: <https://cuttingedgepr.com/find-mobilize-unofficial-opinion-leaders/>).

Figure 9.
Convenience sampling (source: <https://www.mathstopia.net/sampling/convenience-sampling>).

10. Standard error of a sample mean

Suppose we have taken a sample of n measurements of some continuous variate, such as blood pressure or hemoglobin level. The sample mean $\bar{x}$ may be different from the population mean because it is based on a sample. Different samples of the same size n would give different values of the mean. These differences are due to sampling error. The sample mean therefore has its own distribution.

If the distribution of x in the population has mean μ and standard deviation σ and a sample of size n is taken, then the *sampling distribution* of the sample mean $\bar{x}$ has the following properties:

1. The mean of the distribution of $\bar{x}$ is the same as that of the whole population, i.e.

$$E(\bar{x}) = \mu \text{ (see Appendix for proof)}$$

or the sample mean is an unbiased estimate of the population mean.

2. The standard deviation of $\bar{x}$ is equal to $\sigma / \sqrt{n}$. The standard deviation of an estimate is referred to as the standard error (SE). The standard error of the mean is therefore [39]

$$SE(\bar{x}) = \sigma / \sqrt{n}$$

3. By the central limit theorem, $\bar{x}$ is approximately Normally distributed, i.e. the distribution of $\bar{x}$ tends to be Normal even if the distribution in the population is markedly non-Normal. The distribution of means becomes closer to the Normal distribution as n increases.

The standard error of the mean is a measure of the sampling error [40]. For example, consider the measurement of lung function, forced expiratory volume, measured in liters. This is known to have a standard deviation of 0.5 in a population. If we select various sample sizes from this population, we have:

$$n = 1 \quad SE(\bar{x}) = 0.5/1 = 0.50$$

$$n = 9 \quad SE(\bar{x}) = 0.5/3 = 0.17$$

$$n = 25 \quad SE(\bar{x}) = 0.5/5 = 0.10$$

$$n = 100 \quad SE(\bar{x}) = 0.5/10 = 0.05$$

The larger the sample size, the smaller the sampling error. The standard error of the mean is used when we want to indicate how precise our estimate of the mean is. The standard deviation s , on the other hand, is used to show how widely spread our measurements of x are.

11. Confidence intervals

From property (3) above, and using tables of the Normal distribution, it can be stated that there is a 95% probability that $\bar{x}$ (the sample mean) will be no further than $1.96 \times SE(\bar{x})$, from μ , the (unknown) population mean. There is also a 99% chance that it will be within $2.58 \times SE(\bar{x})$ of μ . The quantities $1.96 \times SE(\bar{x})$ and $2.5 \times SE(\bar{x})$ are referred to as the *maximum likely error*.

11.1 Known standard deviation

The foregoing statements describe a property of the sample mean in terms of the population mean. But the sample mean is known, and the population mean unknown, so we require a statement describing the population mean in terms of the known sample mean.

This is achieved by reversing the statement above, viz., if the sample mean is within $1.96 \times SE(\bar{x})$ of the population mean, then the population mean is within $1.96 \times SE(\bar{x})$ of the sample mean. That is, for 95% of samples it is true that μ lies in the interval.

$$\bar{x} - 1.96 \times SE(\bar{x}) \text{ to } \bar{x} + 1.96 \times SE(\bar{x})$$

This interval is called the 95% confidence interval for μ and the ends of the interval are called the 95% *confidence limits*. The single value $\bar{x}$ is called a *point estimate* of the population mean μ , in contrast with the above interval *estimate*.

Example 1

The respiratory health of a sample of 25 men exposed to fumes in a dental laboratory was assessed by measuring the forced expiratory volume (FEV). The sample mean was 3.20 liters. From previous work it is known that the standard deviation of FEV is 0.5 liters.

$$\bar{x} = 3.20$$

$$SE(\bar{x}) = 0.5 / \sqrt{n} = 0.1$$

$$\begin{aligned} \text{95\% confidence interval for } \bar{x} \text{ is } & 3.20 \pm 1.96 \times 0.1 \text{ liters} \\ & = 3.00 \text{ to } 3.40 \text{ liters} \end{aligned}$$

Conclusion We are 95% confident that the interval 3.00 to 3.40 liters contains the unknown population mean μ .

$$\begin{aligned} \text{99\% confidence interval for } \bar{x} \text{ is } & 3.20 \pm 2.5 \times 0.1 \text{ liters} \\ & = 2.94 \text{ to } 3.46 \text{ litres} \end{aligned}$$

Conclusion We are 99% confident that the interval 3.00 to 3.46 liters contains the unknown population mean μ .

Note: The 99% confidence interval is wider because of the extra confidence that the interval contains the population mean μ .

11.2 Unknown standard deviation

In the above it was assumed that the standard deviation in the population was known. In many practical situations this will not be the case and the standard deviation has to be estimated from the sample data. It therefore seems natural to replace σ by its estimate s and argue exactly as above. However, there is a loss of precision because the standard deviation has its own sampling error. This extra imprecision is included by widening the confidence intervals by using larger constants than 1.96 (for 5%) or 2.58 (for 1%).

Example 2. In (example 1) the values of FEV were:

2.82	3.49	2.72	2.59	2.30
2.64	3.16	3.59	2.44	4.10
3.66	3.35	3.62	3.24	2.82
2.79	2.92	4.20	3.21	3.80
3.62	4.00	3.53	2.76	2.68

$$\bar{x} = 3.20 \pm 0.54 \text{ (24df)}.$$

The sample mean is 3.20 liters and the sample standard deviation 0.54 liters. The standard error of the mean is 0.11 liters. The estimated standard deviation has 24 degrees of freedom and the 5% and 1% points of the t distribution with 24 df are 2.06 and 2.80 respectively.

Thus the 95% confidence interval is:

$$3.20 \pm 2.06 \times 0.11 = 2.97 \text{ to } 3.43 \text{ liters}$$

and the 99% confidence interval is:

$$3.20 \pm 2.80 \times 0.11 = 2.89 \text{ to } 3.51 \text{ liter}.$$

12. The significance of choosing an effective method of sampling

Sampling determines the outcome of analysis. However, with the variations that may be present between the population and the sample, there may be errors in the sample (**Figure 10**).

Sampling bias happens because certain members of the population are more likely to be included in a survey than others. It is often referred to as determination bias in medical fields. *Sampling frame error* occurs when a sample is having a certain type of possible respondent relative to the population of interest. *Systematic errors* mostly affect the precision of the calculation. It is of prime importance to use the most appropriate and useful form of sampling to attain accurate outcome of research.

Figure 10.
Sources of error (source: <https://creativemaths.net/blog/sampling-and-non-sampling-error/>).

A. Appendix

A.1. Proof of the basic formulae of statistical inference

$$\begin{aligned} E(x) &= E[(x_1 + x_2 + \dots + x_n)/n] \\ &= [E(x_1) + E(x_2) + \dots + E(x_n)]/n \\ &= \mu, \text{ since } E(x_i) = \mu \text{ for all } i. \end{aligned}$$

$$\begin{aligned} \text{Variance}(x) &= [\text{var}(x_1) + \text{var}(x_2) + \dots + \text{var}(x_n)] / n^2 \\ &= [n\sigma^2] / n^2 \\ &= \sigma^2 / n \end{aligned}$$

Therefore Standard error of x is: $SE(x) = \sigma / \sqrt{n}$

Author details

Yasser Riaz Malik^{1,2*}, Muhammad Saad Sheikh³ and Shakeela Yousaf⁴

1 Department of Preventive Dentistry, Dental Public Health, University of Hail, Hail, Saudi Arabia

2 Department of Community Dentistry, Sir Syed College of Medical Sciences for Girls, Karachi, Pakistan

3 Department of Oral Biology, Sindh Institute of Oral Health Sciences, Jinnah Sindh Medical University, Karachi, Pakistan

4 Sir Syed College of Medical Sciences, Karachi, Pakistan

*Address all correspondence to: docyrm@gmail.com

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

References

- [1] Apostol, T. M. (1969). *Calculus, Volume II*. (2nd edition, John Wiley & Sons.
- [2] Smeeton, N. (2005). *Dental Statistics Made Easy*. Radcliffe Publishing.
- [3] Martin, B. (2000). *Introduction to medical statistics 3rd edition*. oxford university press.
- [4] Dean, B. I. (2013, July 18). Introductory Statistics. OpenStax.
- [5] Frost, J. (2000). *Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries*. James D. Frost.
- [6] Charles Miller Grinstead, J. L. (2012). *Charles Miller Grinstead, James Laurie Snell*. American Mathematical Soc., 2012.
- [7] Rosenthal, M. J. (2004). Probability Models. In *Probability and Statistics* (pp. 4-6). Macmillan.
- [8] Armitage P, B. G. (2001). *Statistical methods in medical research. 4th edition*. Oxford: Blackwell Scientific Publications.
- [9] Bryman, A. &. (2003). *Business research methods*. Oxford: Oxford University Press .
- [10] Altman, D. G. (1999). *Practical statistics for medical research*. London: Champan & Hall.
- [11] Barbara Illowsky, S. D. (2017). *Introductory Statistics*. Samurai Media Limited.
- [12] Palmisano, A. (2018). *The Encyclopedia of Archaeological Sciences*, S. L. López Varela (Ed.) (pp.1–4). Wiley.
- [13] Günter Last, M. P. (2017). *Lectures on the Poisson Process*. Cambridge University Press.
- [14] Koch, K.-R. (2007). *Introduction to Bayesian Statistics*. Springer Science & Business Media.
- [15] Catherine Forbes, M. E. (2010). *Statistical Distributions*. John Wiley & Sons.
- [16] Wesolowski, D. J. (2018). *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. Thousand Oaks: SAGE Publications, Inc.
- [17] Bryc, W. (2012). *The Normal Distribution: Characterizations with Applications*. Springer Science & Business Media.
- [18] Jagdish K. Patel, C. B. (1982). *Handbook of the Normal Distribution*. M. Dekker.
- [19] J.S.Chitode, D. (n.d.). *Digital Communication*. 2009: Technical Publications.
- [20] DasGupta, A. (2010). *Fundamentals of Probability: A First Course*. Springer Science & Business Media.
- [21] Agarwal, B. L. (2006). *Basic Statistics*. New Age International.
- [22] Silvey, S. (2017). *Statistical Inference*. Routledge.
- [23] Michael C. Fleming, J. G. (2000). *Principles of Applied Statistics*. Cengage Learning EMEA.
- [24] Dhivyadeepa, D. E. (n.d.). *Sampling Techniques in Educational Research*. Lulu.com.
- [25] Arijit Chaudhuri, H. S. (2005). *Survey Sampling: Theory and Methods, Second Edition*. CRC Press.
- [26] Sedgwick, P. (June 2012). BMJ, 344.

- [27] Slutsky, D. J. (November 2013). Statistical Errors in Clinical Studies. *Journal of wrist surgery*, 285–287.
- [28] ZIKMUND. (2002). *Business research methods*. Dryden: Thomas learning.
- [29] GHOURI, P. & GRONHAUG, K. 2005. *Research Methods in Business Studies*, Harlow, FT/Prentice Hall
- [30] ACKOFF, R. L. 1953. *The Design of Social Research*, Chicago, University of Chicago Press
- [31] WILSON, J. 2010. *Essentials of business research: a guide to doing your research project*, SAGE Publication
- [32] DAVIS, D. 2005. *Business Research for Decision Making*, Australia, Thomson South-Western
- [33] Michael Köhl, S. S. (2006). *Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory*. Springer Science & Business Media.
- [34] Uprichard, E. (2011). Sampling: bridging probability and non-probability designs. . *International Journal of Social Research Methodology - INT J SOC RES METHODOLOGY* , 16(1):1–11.
- [35] Moser, C. A. (1952). Quota sampling. *Journal of the Royal Statistical Society. Series A (General) Vol. 115, No. 3*, pp. 411–423 (13 pages).
- [36] MAXWELL, J. A. 1996. *Qualitative Research Design: An Intractive Approach* London, Applied Social Research Methods Series
- [37] Irina-Maria Dragan, A. I.-M. (2013). Snowball Sampling Completion. *Journal of Studies in Social Sciences , Volume 5, Number 2*, 160–177.
- [38] Henry, G. T. (1990). *Practical Sampling Volume 21 of Applied Social Research Methods*. SAGE Publications.
- [39] Allen, M. P. (2007). *Understanding Regression Analysis*. Springer Science & Business Media.
- [40] Caldwell, S. (2012). *Statistics Unplugged edition 4*. Cengage Learning.