

* Inferential Statistics

By Dr. Bidisha Bhabani

*Introduction

- * Inferential statistics deals with drawing conclusions about large group of individuals (population) on the basis of observation of a few participants from among them or about the events which are yet to occur on the basis of past events.
- * It provides tools to compute the probabilities of future behavior of the subjects. Inferential statistics throws light on how generalization from sample to population can be made.
- * The fundamental question is: can we infer the population's characteristics from the sample's characteristics?
- * Descriptive statistics remains local to the sample describing its central tendency and variability, while inferential statistics focuses on making statements about the population.
- * Procedure
 - * Estimation
 - * Hypothesis Testing

Dr. Bidisha Bhabani

* **Essence of Sampling**

- * Sampling is a process of selecting subset of observations/records from a population to make inference about various population parameters such as mean, proportion, standard deviation, etc
- * It is an important step in inferential statistics since an incorrect sample may lead to wrong inference about the population

Sampling is necessary when it is difficult or expensive to collect data on the entire population. The inference about the population is made based on the sample that was collected; incorrect sample may lead to incorrect inference about the population.

* **POPULATION PARAMETERS**

- * Measures such as mean and standard deviation calculated using the entire population are called *population parameters*
- * The population parameters mean and standard deviation are usually denoted using symbols μ and σ , respectively

SAMPLE STATISTIC

When population parameters are estimated from sample they are called *sample statistic* or *statistic*

The sample statistic is denoted using symbols \bar{x} (for mean) and s (or s for standard deviation)

* SAMPLING

The process of identifying a subset from a population of elements (aka observations or cases) is called **sampling process** or **simply sampling**

Steps used in any Sampling process:

- Identification of target population that is important for a given problem under study
- Decide the sampling frame.
- Determine the sample size
- Sampling method

- * Availability of population elements
- * Lower cost
- * Greater accuracy of result
- * Greater speed of data collection

* Need of Sampling

Dr. Bidisha Bhabani

- * **Random sampling/probability sampling**

- * Every element of the population has the same probability of being selected in the sample.

- * **Non-random sampling/nonprobability sampling**

- * Not every element of the population has the same probability of being included in the sample.

* Classification of Sampling Methods

Dr. Bidisha Bhabani

* **Random Sampling**

- * Shewhart (1931) defines random sample as a ‘sample drawn under conditions such that the law of large number applies’
- * Random sampling is usually carried out **without replacement**, that is, an observation which is selected in the sample is removed from the population for further consideration
- * Random samples can also be created **with replacement**, that is, an observation which is selected for inclusion in the sample can again be considered since it is replaced (not removed) in the population.

- * Simple Random Sampling
- * Systematic Random Sampling
- * Stratified Random Sampling
- * Cluster Random Sampling

* Methods of Random Sampling

Dr. Bidisha Bhabani

- * Each element of the population has the same probability of being included in the sample as every other element
- * Every sample of fixed size has the same chance of selection as every other sample of that size

* Simple Random Sampling

Dr. Bidisha Bhabani

*The method of random sampling is applicable when the elements of a population are available as a list, such as telephone directory or a list of names in alphabetical order of successful candidates in some competitive examination.

***Systematic Random Sampling**

Dr. Bidisha Bhabani

* **Stratified Sampling**

- * The population can be divided into mutually exclusive relatively homogeneous groups using some factor (for example, age, gender, marital status, income, geographical regions, etc.). The groups, thus, formed are called **stratum (plural from is strata)**
- * It is important that the groups are mutually exclusive and exhaustive of the population.
- * There are two approaches
 - * In the first approach, we select a simple random sample from each stratum and combine them to get the required stratified sample
 - * In the second approach we choose the simple random sample of the same size from each stratum and give weights to the elements of the strata according to the proportion of the strata in the population

Stratified Sampling -Examples

- a) Amount of time spent by male and female users in sending messages in a day. Here the strata are male and female users.
- b) Efficacy of a drug among different age groups. Age group can be classified into categories such as less than 40, between 41 and 60, and over 60 years of age.
- c) Performance of children in school and the parents' marital status. Here, marital status can be (a) Single, (b) Married, (d) Divorced. In this case we assume that the parent's marital status may influence children's academic performance.
- d) Television rating points for a program across different geographical regions of a country. For India, geographical regions could be different states of the country.

* Steps in creating stratified sample

- a) Identify the factor that can be used for creating strata (for example: factor = Age; Strata 1: age less than 40; Strata 2: age between 41 and 60; and Strata 3: Age more than 60).
- b) Calculate the proportion of each stratum in the population (say p_1 , p_2 , and p_3 for three strata identified in step 1).
- c) Calculate the sample size (say N). The sample size for strata 1, 2, and 3 identified in step 2 are $p_1 \times N$, $p_2 \times N$, and $p_3 \times N$, respectively.
- d) Use random sampling procedure to generate random samples in each strata.
- e) Combine samples from each stratum to create the final sample.

- * It has the ability to provide a representative sample
- * On the basis of a stratified sample, we can devise different solutions to different strata
- * For example, a restaurant can find new ways to attract the youth and at the same time retain the aged customer.

* Advantages of Stratified Sampling

Dr. Bidisha Bhabani

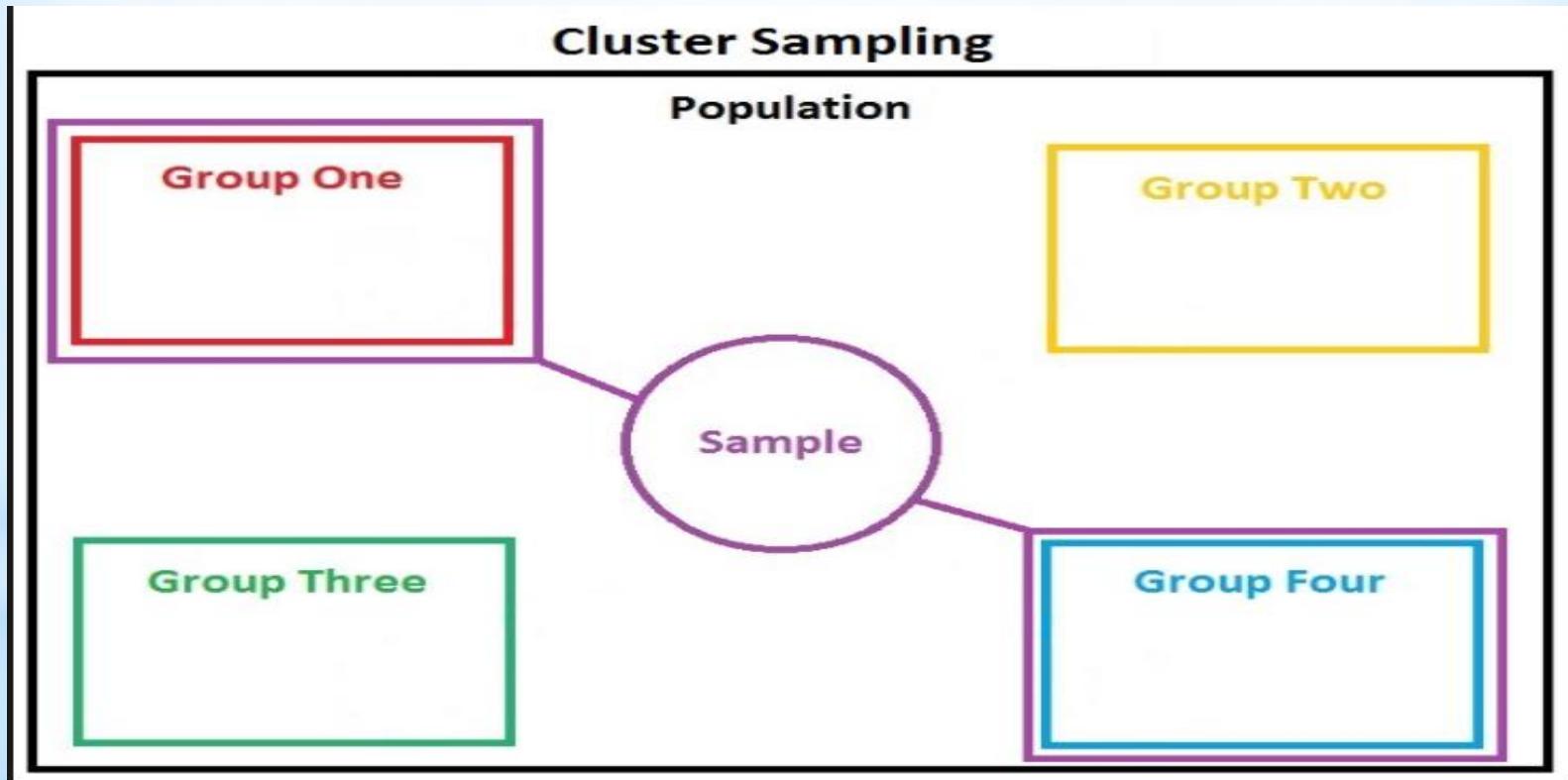
- * The population is divided into non-overlapping groups called clusters, each cluster is representative of the entire population, but elements within a cluster differ in their characteristics. In short, a cluster is population of a smaller scale.
- * The strata are homogeneous within themselves, but heterogeneous between them whereas the clusters are heterogeneous within themselves but homogeneous between them.

* Cluster Sampling

Dr. Bidisha Bhabani

* Cluster Sampling

- * In cluster sampling, the population is divided into mutually exclusive clusters



Cluster Sampling - Steps

- * Identify the clusters (example: different models of smart phones sold by a manufacturer, customers from different geographical locations).
 - * Using random sampling select the clusters.
 - * Select all units in the clusters selected in step 2 and form the sample. If the size is too large, a random sampling within the clusters identified in step 2 may be used for final sample.
-
- **Stratified sampling** and **cluster sampling** are similar; the major difference is that in a stratified sample, all strata will be represented in the sample, whereas in a cluster sampling, not all clusters will be represented

Stratified Sampling Vs Cluster Sampling



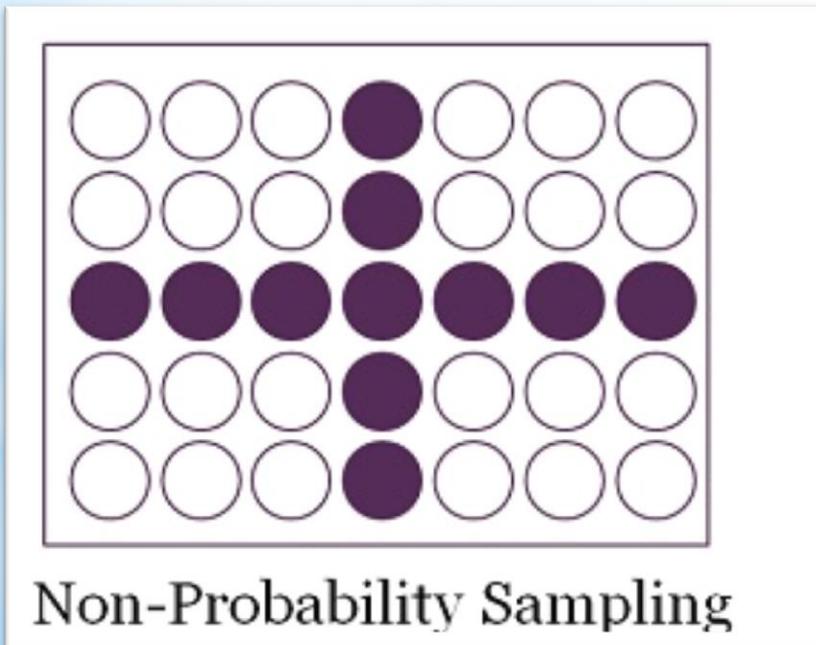
Stratified Sampling Vs Cluster Sampling

* **Bootstrap Aggregating (Bagging)**

- * Bootstrap Aggregating (known as Bagging) is sampling with replacement used in machine learning algorithms, especially the random forest algorithm (Breiman, 1996)
- * The size of each sample and the number of samples are determined based on factors such as population size, target accuracy of the model developed using bagging and convergence, etc
- * Bagging is frequently used in ensemble methods (in which several models are developed and the final prediction is usually based on the majority voting)

* **Non-random/Non-Probability Sampling**

- * In a non-probability sampling, the selection of sample units from the population does not follow any probability distribution
- * Sample units are selected based on convenience and/or on voluntary basis.



- * Convenience Sampling
- * Judgement Sampling
- * Quota Sampling
- * Snowball Sampling

* Convenience Sampling

*Convenience sampling is a non-probability sampling technique in which the sample units are not selected according to a probability distribution. We choose the elements which are readily available or willing to be included in the sample.

*Example-

- * A railway squad checking ticketless traveler
- * TV channel eliciting the opinion of the audience coming out of a theatre

* **Judgement Sampling**

- * Elements are selected on the basis of knowledge/experience of the person selecting the sample

* **Quota Sampling**

- * We divide the population into homogeneous groups as in stratified sampling. But the selection of elements from each group is not random.

* **Snowball Sampling**

- * This method is used when it is very difficult to trace the elements of a population. In snowball sampling, a person/object of the population is identified. Information about other members of the population are sought from the first person. The process continues until we get a sizeable sample of required size. After few steps we get a sizeable sample just as a snowball gathers more and more ice on being rolled.

* If a sample of size n is drawn from a normal population with mean μ and standard deviation σ , then the sample mean $\mu_{\bar{x}} = \mu$ follows a normal distribution with standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

*Sampling Distribution

* Central Limit Theorem (CLT)

- * Let S_1, S_2, \dots, S_k be samples of size n drawn from an independent and identically distributed population with mean μ and standard deviation σ .
- * Let $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$ be the sample means (of the samples S_1, S_2, \dots, S_k).
- * The sampling distribution of mean will follow a normal distribution with mean μ (same as the mean of the population) and standard deviation σ/\sqrt{n}
- In simple terms, central limit theorem states that for a large sample drawn from a population with mean μ and standard deviation σ , the sampling distribution of mean, \bar{X} follows an approximate normal distribution with mean μ and standard deviation (standard error) σ/\sqrt{n} irrespective of the distribution of the population.

Alternative Version of CLT

Let X_1, X_2, \dots, X_n be n random variables that are independent and identically distributed with mean μ and standard deviation σ .

Then for large n , mean

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

follows a normal distribution with mean μ standard error σ/\sqrt{n}

* Central Limit Theorem - Implications

- * The variable $\frac{X - \mu}{\sigma/\sqrt{n}}$ will be a standard normal distribution (mean = 0, standard error = 1).
- * If $S_n = X_1 + X_2 + \dots + X_n$, then $E(S_n) = n\mu$ and Standard error $\sigma\sqrt{n}$ is The random variable $\frac{X - n\mu}{\sigma\sqrt{n}}$ is a standard normal variate.
- * Regardless of the population distribution, the sampling distribution of large sample ($n > 30$) will follow the normal distribution with mean same as population mean and standard error
- *Central limit theorem is the basis for hypothesis tests such as Z test and t test. In many cases, we will have access to only a sample and the inference about the population has to be made based on sample statistic.*
- *An important assumption of CLT is that the random variables have to be independent and identically distributed.*

* Central Limit Theorem for Proportions

* If X_1, X_2, \dots, X_n are counts from a Bernoulli trials with probability of success p , $E(X_i) = p$ and $\text{Var}(X_i) = p \times (1 - p)$, then the sampling distribution of probability of success (say \hat{P}) follows an approximate normal distribution with mean p and standard error

$\sqrt{p(1-p)/n}$ where n is the sample size.

* The variable $\frac{\hat{p} - p}{\sqrt{p(1-p)/n}}$ converges to a standard normal distribution.

* Example

It is believed that college students in Bangalore spend on average 80 minutes daily on texting using their mobile phones and the corresponding standard deviation is 25 minutes. Data from a sample of 100 students were collected for calculating the amount of time spent in texting. Calculate the probability that the average time spent by this sample of students will exceed 84 minutes.

Solution

Using the central limit theorem, the mean of the sampling distribution is 80 and the corresponding standard deviation is $25/\sqrt{100} = 2.5$

The probability that the sample average is more than 84 minutes is given by

$$P\left(Z > \frac{84 - 80}{2.5}\right) = P(Z > 1.6) = 0.05479$$

* Sample Size Estimation for Mean of the population

- * From the central limit theorem, we know that the sampling distribution of mean follows a normal distribution with mean μ and standard deviation σ/\sqrt{n}
- * The standard normal variate of the sampling distribution of mean is given by $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$

The difference between the sample mean and the population mean

$\bar{X} - \mu$ is error in estimation of the population mean. Above equation can be written as

$$n = \left[\frac{Z_{\alpha/2} \times \sigma}{D} \right]^2$$

where $Z_{\alpha/2}$ is the critical value for normal distribution or $(1 - \alpha)$ is the desired confidence in estimating the population mean and $D = \bar{X} - \mu$ is the error in estimating the population mean.

* Example

A hospital is interested in estimating the time it takes to discharge a patient after the clearance (discharge note) by the doctor. Calculate the sample size at a confidence of 95% and maximum error in estimation of 5 minutes. Assume that the population standard deviation is 30 minutes.

Solution

- We know that $D = 5$, $\sigma = 30$, $\alpha = 0.05$, and $|Z_{\alpha/2}| = 1.96$ for $\alpha = 0.05$ we get

$$n = \left[\frac{Z_{\alpha/2} \times \sigma}{D} \right]^2 = \left[\frac{1.96 \times 30}{5} \right]^2 \approx 138$$

* Estimation

- * In estimation a sample is drawn and studied and inference is made about the population characteristics on the basis of what is discovered about the sample.
- * There may be sampling variations because of chance fluctuations, variations in sampling techniques, and other sampling errors.
- * We, therefore, do not expect our estimate of the population characteristics to be exactly correct. We do, however, expect it to be close.
- * The real question in estimation is not whether our estimate is correct or not but how close is it to be the true value.

Dr. Bidisha Bhabani

* **Estimation of Population Parameters**

- * Estimation is a process used for making inferences about population parameters based on samples
- * **Point Estimate:** Point estimate of a population parameter is the single value (or specific value) calculated from sample (thus called statistic).
- * **Interval Estimate:** Instead of a specific value of the parameter, in an interval estimate the parameter is said to lie in an interval (say between points a and b) with certain probability (or confidence).

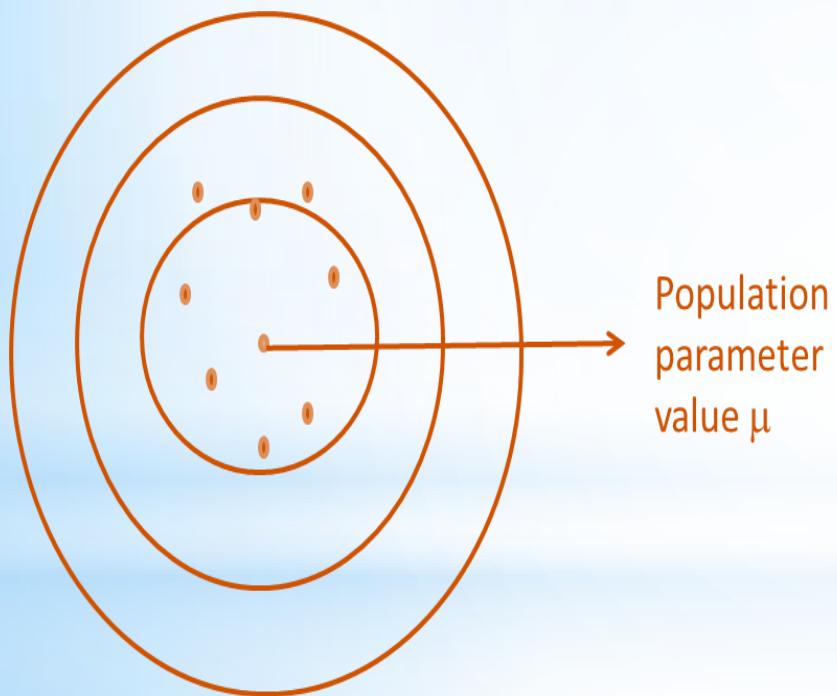
* Criteria for measuring Quality of Estimates

A. Unbiased estimate of a population parameter is an estimator whose expected value is equal to the population parameter.

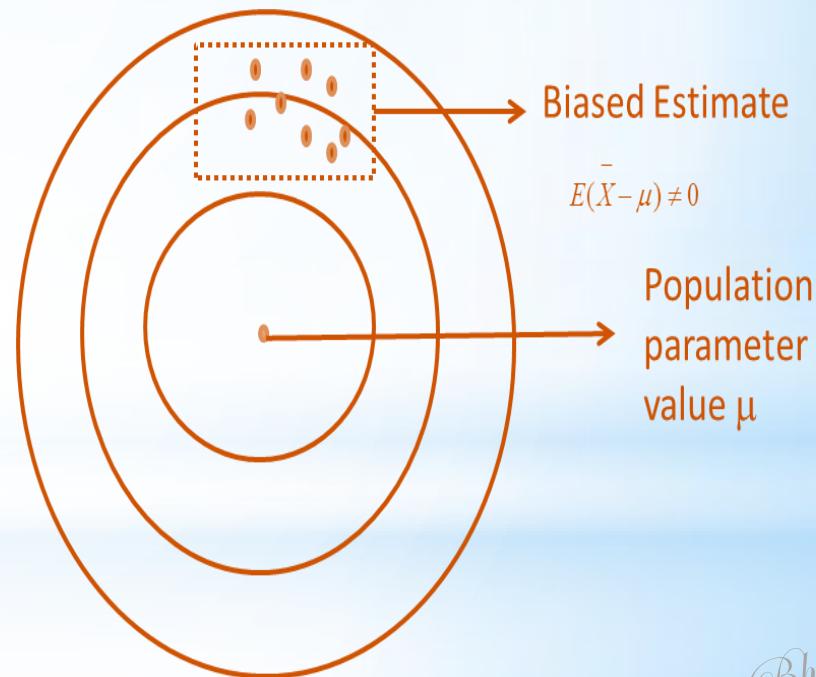
Let \bar{X} be an estimate of the population mean μ . If \bar{X} is an unbiased estimate of μ , then

$$E(\bar{X}) = \mu \quad \text{or} \quad E(\bar{X} - \mu) = 0$$

*Unbiased estimate (the estimates are randomly scattered around the actual value).



*Biased estimate (all estimated values are to the one side of the actual value).



B. Consistency: An estimator of population parameter (say \bar{X}) is said to be consistent if it converges to the true value of the parameter (μ) as the size of the sample increases .That is, a consistent estimator implies

$$\lim_{n \rightarrow \infty} \bar{X} = \mu$$

C. Efficiency: An efficient estimator implies that resulting estimate of the population parameter has the minimum variance.

D. Sufficiency: An estimator is sufficient if it makes use of more information available in the sample than any other estimator.

* Point Estimate

* Point Estimate of Population Mean

* The sample mean \bar{x} is an unbiased and consistent estimator of population mean.

* Point Estimate of Population Variance and Population Standard Deviation

* Population variance $s^2 = \frac{\sum(x_i - \bar{x}^2)^2}{n-1}$ is an unbiased estimator of s^2

$$* s^2 = \frac{\sum x^2}{n-1} - \frac{n\bar{x}^2}{n-1}$$

* Population deviation $s = \sqrt{\frac{\sum(x_i - \bar{x}^2)^2}{n-1}}$ is an unbiased estimator of s

* Interval Estimate

- * An interval estimate of a population parameter is a range of values within which the population parameter is likely to lie.
- * We have 3 cases in interval estimation:
 - * Case1: In the case of large samples (30 or more) from any population, we can use properties of normal distribution in construction of interval estimates.
 - * Case2: In the case of small samples of normal population, we can use properties of normal distribution in construction of interval estimates.
 - * Case3: In the case of small samples of non-normal population, we can use properties of t-distribution in construction of interval estimates.

- * The probability associated with an interval estimate is called the confidence level
- * The interval is called the confidence interval associated with the confidence level
- * Properties:
 - * All the confidence intervals are centered at \bar{x} , the sample mean.
 - * If $\alpha < \beta$, then the confidence interval corresponding to β contains the confidence interval corresponding to α
 - * The width of the confidence interval increases as the confidence level decreases.
 - * When the confidence level is one, the confidence interval is the entire line
 - * When the confidence level approaches to zero, the confidence interval reduces to point estimate \bar{x} .

* Confidence Level and Confidence Interval

- * The life of a particular TV model produced by a reputed TV manufacturer has a standard deviation of 2.5 years. In order to estimate the mean life of the TV model, a sample of 40 TVs was taken and found to have a mean life of 9 years.
 - * Compute the standard error of the sample mean
 - * Construct a 90% confidence interval for the mean life of the particular model
 - * Construct a 98% confidence interval for the mean life of the particular model

* Example 1

Dr. Bidisha Bhabani

Solution: We are given the following data

$$\sigma = \text{Population standard deviation} = 2.5$$

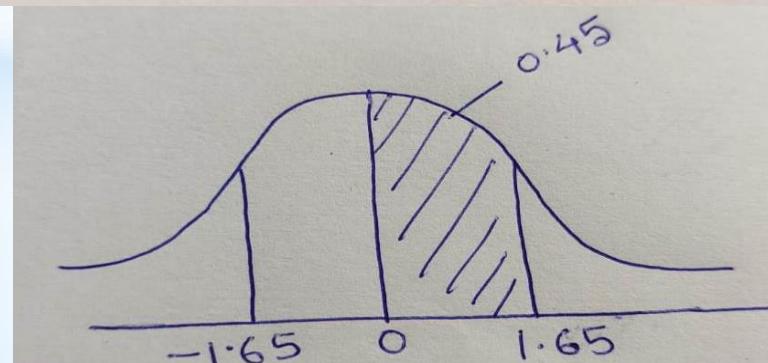
For the sample,

$$n = 40 \text{ and } \bar{x} = 9$$

(a) S.E. = $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{40}} = 0.396$

- (b) For the confidence level of 0.9, the z -value corresponding to 0.45 is 1.65. So, the 90% confidence interval is $\bar{x} \pm (1.65)$ (S.E.)
i.e., 9 ± 0.653
So, the 90% confidence interval is [8.347, 9.653]

- (c) For 98% confidence interval, the confidence level is 0.98
The z -value corresponding to 0.49 is 2.33. So, the 98% confidence interval is $\bar{x} \pm (2.33)$ (S.E.)
i.e., $9 \pm (2.33)(0.653) = 9 \pm 1.522$
So, the 98% confidence interval is [7.478, 10.522]



- * Use the following information to construct confidence intervals for estimating mean.
 - * 80% confidence interval for $\bar{x} = 55$ $\sigma = 12$ and $n = 32$
 - * 90% confidence interval for $\bar{x} = 3.5$ $\sigma = 0.97$ and $n = 30$
 - * 95% confidence interval for $\bar{x} = 25$ $\sigma = 4.5$ and $n = 60$

* Example 2

Dr. Bidisha Bhabani

The confidence level corresponding to 80% confidence interval is 0.8. The z value corresponding to 0.4 is 1.28

$$\text{S.E.} = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{32}} = \frac{12}{5.657} = 2.12$$

So, 90% confidence interval is:

$$\bar{x} \pm 1.28 (\text{S.E.})$$

i.e. $55 \pm (1.28) (2.12)$

i.e. 55 ± 2.714

So, 80% confidence interval is [52.286, 57.714];

The confidence level corresponding to 90% confidence interval is 0.9. The z value corresponding to 0.45 is 1.645

$$\begin{aligned} \text{S.E.} &= \frac{\hat{\sigma}}{\sqrt{n}} = \frac{s}{\sqrt{n}} = \frac{0.97}{\sqrt{30}} = \frac{0.97}{5.477} \\ &= 0.177 \end{aligned}$$

So, 90% confidence interval is:

$$\bar{x} \pm (1.645) (\text{S.E.})$$

i.e. $3.5 \pm (1.645) (0.177)$

i.e. 3.5 ± 0.291

So, the 90% confidence interval is:

$$[3.209 \ 3.791]$$

The confidence level corresponding to 98% is 0.98 and the z value corresponding to 0.49 is 2.33

$$S.E. = \frac{\sigma}{\sqrt{n}} = \frac{4.5}{\sqrt{60}} = \frac{4.5}{7.746} = 0.581$$

So, 98% confidence interval is:

i.e.

$$25 \pm (2.33)(0.581)$$
$$25 \pm 1.354$$

So 98% confidence interval is

$$[23.646, 26.354]$$

- * The customers coming to a branch of LIC complain that they have to wait for a long time before paying their premium. The manager knows from his past experience that the standard deviation of the waiting time is 2.5 minutes. He took a sample of 49 persons, found their average waiting time to be 12.3 minutes. Construct a confidence interval for the average waiting time with confidence level of
- * 90%
 - * 99%

* Example 3

Dr. Bidisha Bhabani

Solution: We are given the following data:

$$\sigma = 2.5 \quad n = 49 \quad \bar{x} = 12.3$$

$$\text{S.E.} = \frac{\sigma}{\sqrt{n}} = \frac{2.5}{\sqrt{49}} = \frac{2.5}{7} = 0.357$$

- (a) For the confidence level of 0.9, the z value corresponding to 0.45 is 1.645. So, 90% confidence interval is:

$$12.3 \pm (1.645)(0.357)$$

i.e. 12.3 ± 0.587

So, 90% confidence interval is [11.713, 12.887]

- (b) For a confidence level of 0.99 the z value corresponding to 0.495 is 2.575. So, 99% confidence interval is:

$$12.3 \pm (2.575)(0.357)$$

i.e. 12.3 ± 0.919

So, 99% confidence interval is [11.381, 13.219]

Estimation Approaches

The estimation of parameters is usually carried out using the following approaches:

- * Method of Moments
- * Maximum Likelihood Estimate (MLE)
- * Bayesian Estimation

* Method of Moments

*Moments are measures used in statistics. According to method of moments, a theoretical curve $Y = f(X, c_1, c_2, \dots, c_n)$, where c_1, c_2, \dots are the model parameters, can be fitted given a set of observations by equating the area and first $(n - 1)$ moments of the observations (Schultz, 1925). The n^{th} order moment, $E(X^n)$, is given by

$$E(X^n) = \sum_i x_i^n \times p(x_i) \quad \text{when } X \text{ is discrete}$$

Where $p(x_i)$ is the probability mass function

* For a continuous random variable, the n^{th} moment is given by

$$E(X^n) = \int_{-\infty}^{+\infty} x^n f(x) dx$$

where $f(x)$ is the probability density function.

* Central moments are moments about the mean, μ , and are given by

$$E((X - \mu)^n) = \sum_i (x_i - \mu)^n \times p(x_i) \quad \text{when } X \text{ is discrete}$$

* For a continuous random variable, the n^{th} central moment is given by

$$E((X - \mu)^n) = \int_{-\infty}^{+\infty} (x - \mu)^n f(x) dx$$

* The moments can be connected to various measures of population. The zero-order moment is the total probability.

$$E(X^0) = \int_{-\infty}^{+\infty} x^0 f(x) dx = 1$$

Similarly, the first-order moment is the mean:

$$E(X^1) = \int_{-\infty}^{+\infty} x f(x) dx$$

Second-order moment about the mean is the variance:

$$E(X - \mu)^2 = \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

Thus, moments can be used for estimating the population parameters

* Estimation of Parameters using Method of Moments

* Consider a uniform distribution between points a and b with probability density function $[1/(b - a)]$. We can use the method of moments to estimate the mean and standard deviation as shown below:

$$E(X) = \int_a^b x \times \frac{1}{b-a} dx = \frac{1}{b-a} \int_a^b x dx = \frac{1}{b-a} \left[\frac{b^2}{2} - \frac{a^2}{2} \right] = \frac{a+b}{2}$$

* The estimate of variance is

$$E(X - \mu)^2 = \int_a^b [x - (a+b)/2]^2 \times \frac{1}{b-a} dx = \frac{1}{b-a} \times \frac{1}{3} \times \left[\left(b - \frac{a+b}{2} \right)^3 - \left(a - \frac{a+b}{2} \right)^3 \right] = \frac{1}{12}(b-a)^2$$

* Example

Estimate the expected value of Poisson and exponential random variable using method of moments.

Solution:

The probability density function of Poisson distribution is given by

$$f(x) = \frac{e^{-\lambda} \times \lambda^x}{x!}$$

The expected value is given by

$$E(X) = \sum_{i=0}^{\infty} i \times \frac{e^{-\lambda} \times \lambda^x}{x!} = \sum_{i=1}^{\infty} i \times \frac{e^{-\lambda} \times \lambda^i}{i!} = \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!}$$

Now

$$\sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} = e^{\lambda}$$

Thus , $E(X) = \lambda e^{-\lambda} e^{\lambda} = \lambda$

The probability density of exponential distribution is given by

$$f(x) = \lambda e^{-\lambda x}$$

The expected value is given by $E(X) = \int_0^{\infty} x \times \lambda e^{-\lambda x} dx$

We have to solve the above integration by parts. Let $u = x$ and $dv = \lambda e^{-\lambda x} dx$. Then $du = dx$ and $v = -e^{-\lambda x}$. Integrating by parts we get

$$\int_0^{\infty} x \times \lambda e^{-\lambda x} dx = [uv]_0^{\infty} - \int_0^{\infty} v du = 0 - \left[\frac{e^{-\lambda x}}{\lambda} \right]_0^{\infty} = \frac{1}{\lambda}$$

* **Estimation of Parameters Using Maximum Likelihood Estimation**

- * One of the frequently used methods for estimation of parameters of probability distribution is called maximum likelihood estimation (MLE).
- * The main advantages of MLE are that it is mathematically rigorous and less susceptible to individual values as every data in the sample has equal weight in calculation of the estimates of the parameters
- * The method is very robust and thus can be used for any distribution

Dr. Bidisha Bhabani

Maximum Likelihood Estimation - STEPS

- 1.** Start with a belief about the population (say exponential distribution).
- 2.** Derive the likelihood function that estimates probability of observing the data using the belief in step 1.
- 3.** Do a natural logarithmic transformation of the likelihood function (Log likelihood function). Log likelihood function is used to simplify the computation.
- 4.** Estimate the parameters that maximized the log likelihood function derived in step 3.

* Estimation of Binomial Distribution Parameter

Consider a binomial distribution with n Bernoulli trials and each with probability of success p . The objective is to estimate the probability p . The probability density function of binomial distribution is given by

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}$$

where

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

Let x_1, x_2, \dots, x_m be the number of success obtained out of n successive trials repeated m times. The corresponding joint probability is the likelihood of observing x_1, x_2, \dots, x_m successes out of n trials repeated m times and is given by

$$L(x_1, x_2, \dots, x_m | p, n) = \prod_{i=1}^m f(x_i) = \prod_{i=1}^m \binom{n}{x_i} p^{x_i} (1-p)^{n-x_i}$$

The log likelihood function is given by

$$LL(x_1, x_2, \dots, x_m | p, n) = \sum_{i=1}^m \ln(n!) - \sum_{i=1}^m \ln(x!) - \sum_{i=1}^m \ln(n-x!) + \sum_{i=1}^m x_i \ln(p) + \sum_{i=1}^m (n-x_i) \ln(1-p)$$

Taking derivative and setting it to zero, we get

$$\frac{dLL(x_1, x_2, \dots, x_m | p, n)}{dp} = \frac{1}{p} \sum_{i=1}^m x_i - \frac{1}{1-p} \sum_{i=1}^m (n - x_i) = 0$$

That is

$$(1-p) \sum_{i=1}^m x_i - p \sum_{i=1}^m (n - x_i) = 0 \Rightarrow \sum_{i=1}^m x_i - p \sum_{i=1}^m x_i - p \times m \times n + p \sum_{i=1}^m x_i = 0$$

Thus, the estimate \hat{p} is given by

$$\hat{p} = \frac{\sum_{i=1}^m x_i}{m \times n}$$

That is, the estimate \hat{p} is average of proportions

* Example

* A talent acquisition company is interested in estimating the probability of successful recruitment of top executives for their clients. Table shows the number of successful recruits out of 10 persons interviewed during the past 8 recruitment cycles. Estimate the probability of success p using the maximum likelihood estimation

Recruitment cycle number	1	2	3	4	5	6	7	8
Number of people recruited	4	2	5	4	2	1	5	3

Solution

- The estimate of p is given by

$$\hat{p} = \frac{\sum_{i=1}^8 x_i}{m \times n} = \frac{26}{8 \times 10} = 0.325$$

* Estimation of Scale Parameter of Exponential Distribution

- * Assume that a data set $\{X_1, X_2, \dots, X_n\}$ follows an exponential distribution with scale parameter λ . The objective of MLE is to estimate the value of λ that will maximize the likelihood of the data $\{X_1, X_2, \dots, X_n\}$. The likelihood of observing the data $\{X_1, X_2, \dots, X_n\}$ from an exponential distribution is given by the joint probability given in Eq.

$$L(\lambda) = f(X_1, X_2, \dots, X_n) = \lambda e^{-\lambda X_1} \times \lambda e^{-\lambda X_2} \times \dots \times \lambda e^{-\lambda X_n} = \lambda^n e^{-\lambda \sum_{i=1}^n X_i}$$

where $L(\lambda)$ is the likelihood function, which is same as the joint probability of observing the data $\{X_1, X_2, \dots, X_n\}$ that follows an exponential distribution. In Eq. (4.19), we assume that the events X_1, X_2 , etc. are independent. The objective of MLE is to find the value of λ that will maximize the likelihood function, that is

$$\underset{\lambda}{\text{Max}} \left(L(\lambda) = \lambda^n \times e^{-\lambda \sum_{i=1}^n x_i} \right)$$

- To find the optimal value of λ , we have to take the derivative of the likelihood function in Eq. (4.20) and equate that to zero. However, the derivative is mathematically intractable, thus we take log likelihood function instead of likelihood function defined in Eq.
- The log likelihood function is given by

$$LL(\lambda) = n \ln(\lambda) - \lambda \times \sum_{i=1}^n X_i$$

*The derivative of Eq with respect to λ is given by

$$\frac{dLL(\lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n X_i$$

*Equating Eq to zero and rearranging, we get

$$\lambda^* = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\frac{1}{n} \sum_{i=1}^n X_i} = \frac{1}{\bar{X}}$$

*Where \bar{X} is the mean value of the observed data.

Example

- * Time to failure of an electronic component is assumed to follow an exponential distribution. Data of 20 failures measured in days are given in Table 4.6. Estimate the time between failure and the failure rate.

Failure times of 20 electronic components

40	72	56	95	32	12	64	120	145	89
26	37	69	78	98	44	7	21	76	102

Solution

- Making the assumption that these times are exponentially distributed, we can find the MLE of the parameter as

$$\lambda = \frac{1}{\bar{X}} = \frac{1}{\frac{1}{20}(40 + 72 + \dots + 102)} = \frac{1}{64.15} = 0.01558$$

- The estimate of the mean time between failure is 64.15 days and the corresponding failure rate (λ) is 0.01558

* MLE of Normal Distribution Parameters

- * The probability density function for the normal distribution is given by

$$f(x : \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- * The likelihood function of the normal distribution given the data $\{X_1, X_2, \dots, X_n\}$ is

$$L(X_1, X_2, \dots, X_n; \mu, \sigma) = \prod_{i=1}^n \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

- * Log likelihood function is given by

$$LL(X_1, X_2, \dots, X_n; \mu, \sigma) = \sum_{i=1}^n \left\{ \frac{1}{2} \ln(2\pi) + \ln(\sigma) + \frac{(X_i - \mu)^2}{2\sigma^2} \right\}$$

- * The maxima of the log likelihood function occurs when

$$\frac{\partial LL}{\partial \mu} = \frac{1}{2\sigma^2} \sum_{i=1}^n -2(X_i - \mu) = 0$$

$$\frac{\partial LL}{\partial \mu} = \frac{\partial LL}{\partial \sigma} = 0$$

* which can be reduced to

$$\mu = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$$

* That is, the maximum likelihood estimator of the mean is simply the sample mean:

$$\frac{\partial L}{\partial \sigma} = \frac{n}{\sigma} - \frac{1}{\sigma^3} \sum_{i=1}^n (X_i - \mu)^2$$

* which can be reduced to

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

* **Summary**

- * Sampling is a process of creating a subset from the population since collecting the data from the entire population is either expensive or impossible.
- * Sampling process start by identifying the target population, identifying sampling frame, calculating the sample size and choosing the method of sampling.
- * Sampling frame which identifies the source of data is important for correct inference about the population. An incorrect sampling frame can result in incorrect inference about the population as demonstrated in the example of Literary Digest.

Dr. Bidisha Bhabani

- * Random sampling, stratified sampling, cluster sampling and convenient sampling are few frequently used sampling techniques.
- * In a random sampling, every case in the population has equal probability of being selected in the sample. Random sampling is one of the most popular sampling techniques.
- * According to the central limit theorem, sampling distribution of mean and proportion for a large sample follows a normal distribution.
- * Central limit theorem forms the basis of many hypothesis tests and test statistic are derived based on CLT.
- * The estimation of various parameters of probability distributions can be derived using method of moments or using method of maximum likelihood estimation

