1. Sampling Method:

Definition: Sampling is a technique of selecting individual members or a subset of the population to make statistical inferences from them and estimate the characteristics of the whole population. Sampling is the process of selecting a subset of individuals or items from a larger population. The goal is to gather information about the population based on the characteristics of the sample.

Importance: Proper sampling is crucial for obtaining representative data and making inferences about the entire population.

The Formula of Sampling Method

(N-n/N-(n-1)). Here P is a probability, n is the sample size, and N represents the population. Now if one cancels 1-(N-n/n), it will provide P = n/N. Moreover, the chance of a sample getting selected more than once is needed: P = 1-(1-(1/N)) n.

Example of Sampling Method; Assume a firm with 1000 employees, of the 100 are needed to complete an onsite work. Now all their names are in the basket and 100 will be picked from those. Now, in this instance, every employee has an equal chance of getting selected.From this database, one can easily select the probability, once the sample size and population is available. Here is the calculation –The chance of one-time selection is:

P = n/N = 100/1000 = 10%

And, for more than once –P = 1-(1-(1/N))n

P = 1 – (999/1000)100P = 0.952P ≈ 9.5%

Types of Sampling Method

i. Probability sampling methods

Probability sampling means that every member of the population has a chance of being selected. It is mainly used in quantitative research. If you want to produce results that are representative of the whole population, probability sampling techniques are the most valid choice.

There are four main types of probability sample.

1. Simple random sampling

In a simple random sample, every member of the population has an equal chance of being selected. Your sampling frame should include the whole population.

To conduct this type of sampling, you can use tools like random number generators or other techniques that are based entirely on chance.

- **Example1:** Simple random sampling You want to select a simple random sample of 1000 employees of a social media marketing company. You assign a number to every employee in the company database from 1 to 1000, and use a random number generator to select 100 numbers.
- **Example2:** Suppose we want to select a simple random sample of 200 students from a school. Here, we can assign a number to every student in the school database from 1 to 500 and use a random number generator to select a sample of 200 numbers.

2. Systematic sampling

Systematic sampling is similar to simple random sampling, but it is usually slightly easier to conduct. Every member of the population is listed with a number, but instead of randomly generating numbers, individuals are chosen at regular intervals.

Example: Systematic sampling Suppose the names of 300 students of a school are sorted in the reverse alphabetical order. To select a sample in a systematic sampling method, we have to choose some 15 students by randomly selecting a starting number, say 5. From number 5 onwards, will select every 15th person from the sorted list. Finally, we can end up with a sample of some students. If you use this technique, it is important to make sure that there is no hidden pattern in the list that might skew the sample. For example, if the HR database groups employees by team, and team members are listed in order of seniority, there is a risk that your interval might skip over people in junior roles, resulting in a sample that is skewed towards senior employees.

3. Stratified sampling

Stratified sampling involves dividing the population into subpopulations that may differ in important ways. It allows you draw more precise conclusions by ensuring that every subgroup is properly represented in the sample.

To use this sampling method, you divide the population into subgroups (called strata) based on the relevant characteristic (e.g., gender identity, age range, income bracket, job role).

Based on the overall proportions of the population, you calculate how many people should be sampled from each subgroup. Then you use random or systematic sampling to select a sample from each subgroup.

Example: Stratified samplingThe company has 800 female employees and 200 male employees. You want to ensure that the sample reflects the gender balance of the company, so you sort the population into two strata based on gender. Then you use random sampling on each group, selecting 80 women and 20 men, which gives you a representative sample of 100 people.

4. Cluster sampling

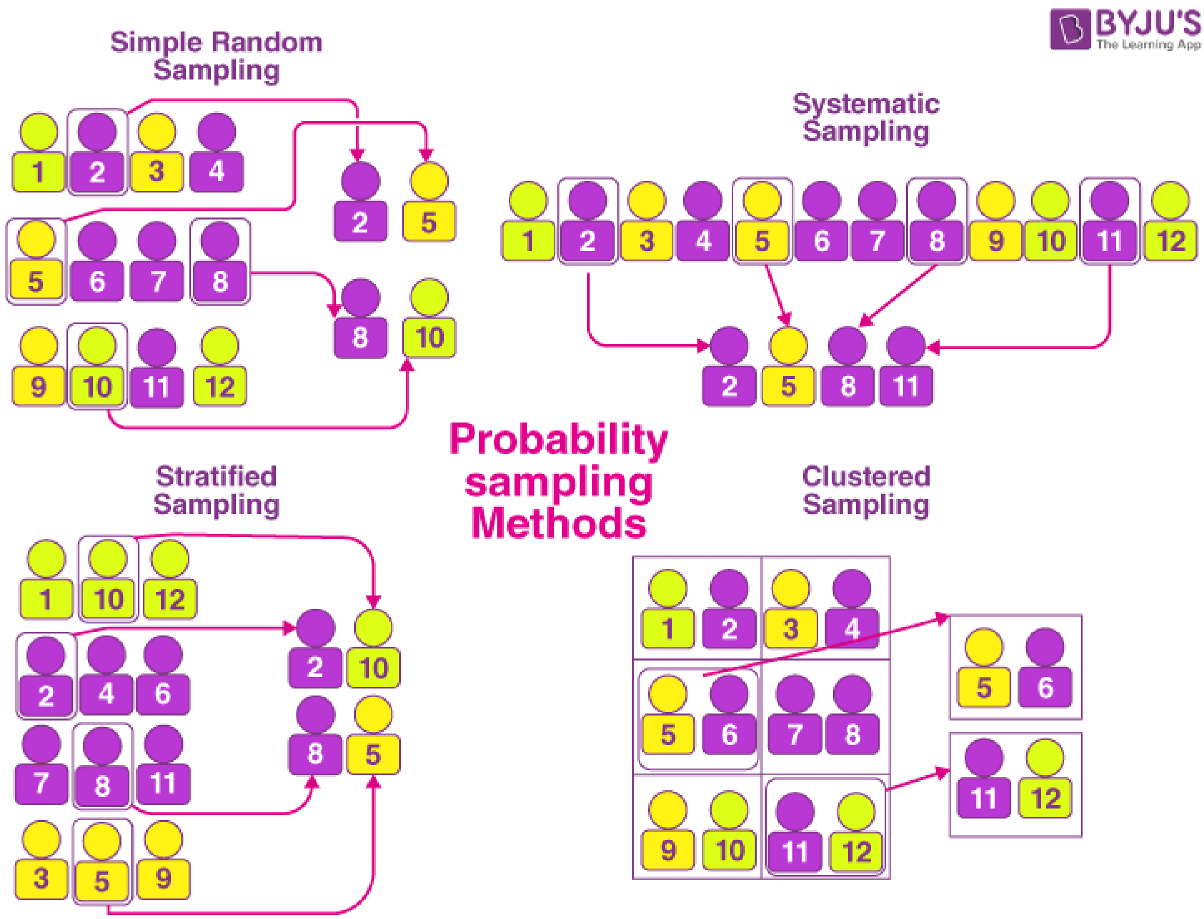
Cluster sampling also involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, you randomly select entire subgroups.

If it is practically possible, you might include every individual from each sampled cluster. If the clusters themselves are large, you can also sample individuals from within each cluster using one of the techniques above. This is called multistage sampling.

This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.

Example: Cluster samplingThe company has offices in 10 cities across the country (all with roughly the same number of employees in similar roles). You don't have the capacity to travel to every office to collect your data, so you use random sampling to select 3 offices – these are your clusters.

All these four methods can be understood in a better manner with the help of the figure given below. The figure contains various examples of how samples will be taken from the population using different techniques.

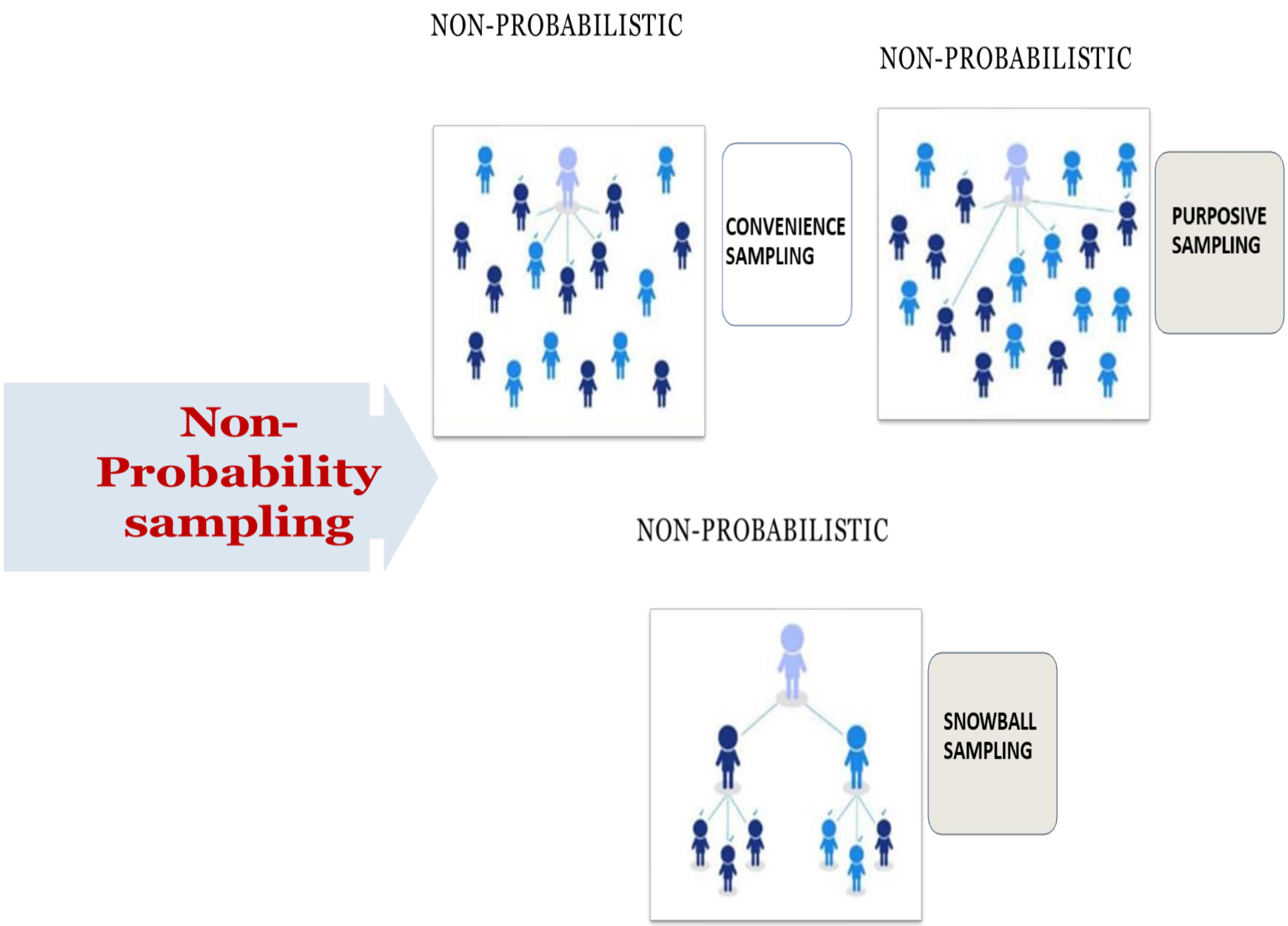


ii. Non-probability sampling methods

In a non-probability sample, individuals are selected based on non-random criteria, and not every individual has a chance of being included.

This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias. That means the inferences you can make about the population are weaker than with probability samples, and your conclusions may be more limited. If you use a non-probability sample, you should still aim to make it as representative of the population as possible.

Non-probability sampling techniques are often used in exploratory and qualitative research. In these types of research, the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small or under-researched population.



1. Convenience sampling

A convenience sample simply includes the individuals who happen to be most accessible to the researcher.

This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population, so it can't produce generalizable results. Convenience samples are at risk for both sampling bias and selection bias.

Example: Convenience samplingYou are researching opinions about student support services in your university, so after each of your classes, you ask your fellow students to complete a survey on the topic. This is a convenient way to gather data, but as you only surveyed students taking the same classes as you at the same level, the sample is not representative of all the students at your university.

2. Voluntary response sampling

Similar to a convenience sample, a voluntary response sample is mainly based on ease of access. Instead of the researcher choosing participants and directly contacting them, people volunteer themselves (e.g. by responding to a public online survey).

Voluntary response samples are always at least somewhat biased, as some people will inherently be more likely to volunteer than others, leading to self-selection bias.

Example: Voluntary response samplingYou send out the survey to all students at your university and a lot of students decide to complete it. This can certainly give you some insight into the topic, but the people who responded are more likely to be those who have strong opinions about the student support services, so you can't be sure that their opinions are representative of all students.

3. Purposive sampling

This type of sampling, also known as judgement sampling, involves the researcher using their expertise to select a sample that is most useful to the purposes of the research.

It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences, or where the population is very small and specific. An effective purposive sample must have clear criteria and rationale for inclusion. Always make sure to describe your inclusion and exclusion criteria and beware of observer bias affecting your arguments.

Example: Purposive samplingYou want to know more about the opinions and experiences of disabled students at your university, so you purposefully select a number of students with different support needs in order to gather a varied range of data on their experiences with student services.

4. Snowball sampling

If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people you have access to "snowballs" as you get in contact with more people. The downside here is also representativeness, as you have no way of knowing how representative your sample is due to the reliance on participants recruiting others. This can lead to sampling bias.

Example: Snowball samplingYou are researching experiences of homelessness in your city. Since there is no list of all homeless people in the city, probability sampling isn't possible. You meet one person who agrees to participate in the research, and she puts you in contact with other homeless people that she knows in the area.

5. Quota sampling

Quota sampling relies on the non-random selection of a predetermined number or proportion of units. This is called a quota.

You first divide the population into mutually exclusive subgroups (called strata) and then recruit sample units until you reach your quota. These units share specific characteristics, determined by you prior to forming your strata. The aim of quota sampling is to control what or who makes up your sample.

Example: Quota sampling You want to gauge consumer interest in a new produce delivery service in Boston, focused on dietary preferences. You divide the population into meat eaters, vegetarians, and vegans, drawing a sample of 1000 people. Since the company wants to cater to all consumers, you set a quota of 200 people for each dietary group. In this way, all dietary preferences are equally represented in your research, and you can easily compare these groups.You continue recruiting until you reach the quota of 200 participants for each subgroup.

2. Point Estimation:

Definition: Point estimation involves using a single value (a point) to estimate an unknown parameter of a population. This single value is often derived from the sample data.

Example: If you want to estimate the average height of all students in a school, you might use the sample mean height as a point estimate.

The formula of Point Estimation is p = x/n

Now, to the first estimator. This is the **sample mean**, $\bar{x}$, of the population mean, μ. Its formula is

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n},$$

where

- x_i are the data points (observations) of a sample;
- n is the sample size.

As you have already read, this is the best unbiased estimator of the population mean. This is an estimator based on the arithmetic mean.

Example

Given the values below, find the best point estimate for the population mean μ.

$$7.61, 7.17, 9.06, 6.305, 7.805, 7.11, 9.705, 6.11, 8.56, 7.11, 6.455, 9.06$$

Solution:

The idea is simply to calculate the sample mean of this data.

$$\begin{aligned}
\bar{x} &= \frac{\sum_{i=1}^{n} x_i}{n} \\
&= \sum_{i=1}^{n} \frac{x_i}{n} \\
&= \frac{7.61}{12} + \frac{7.17}{12} + \frac{9.06}{12} + \frac{6.305}{12} + \frac{7.805}{12} \\
&\quad + \frac{7.11}{12} + \frac{9.705}{12} + \frac{6.11}{12} + \frac{8.56}{12} \\
&\quad + \frac{7.11}{12} + \frac{6.455}{12} + \frac{9.06}{12} \\
&= \frac{92.06}{12} \\
&= 7.67
\end{aligned}$$

The best point estimate for the population mean μ is $\bar{x} = 7.67$.

Example

A survey was conducted using a sample of 300 teacher trainees in a training school to determine what proportion of them view the services provided to them favorably. Out of 150 trainees, 103 of them responded that they viewed the services provided to them by the school as favorable. Find the point estimation for this data.

Solution:

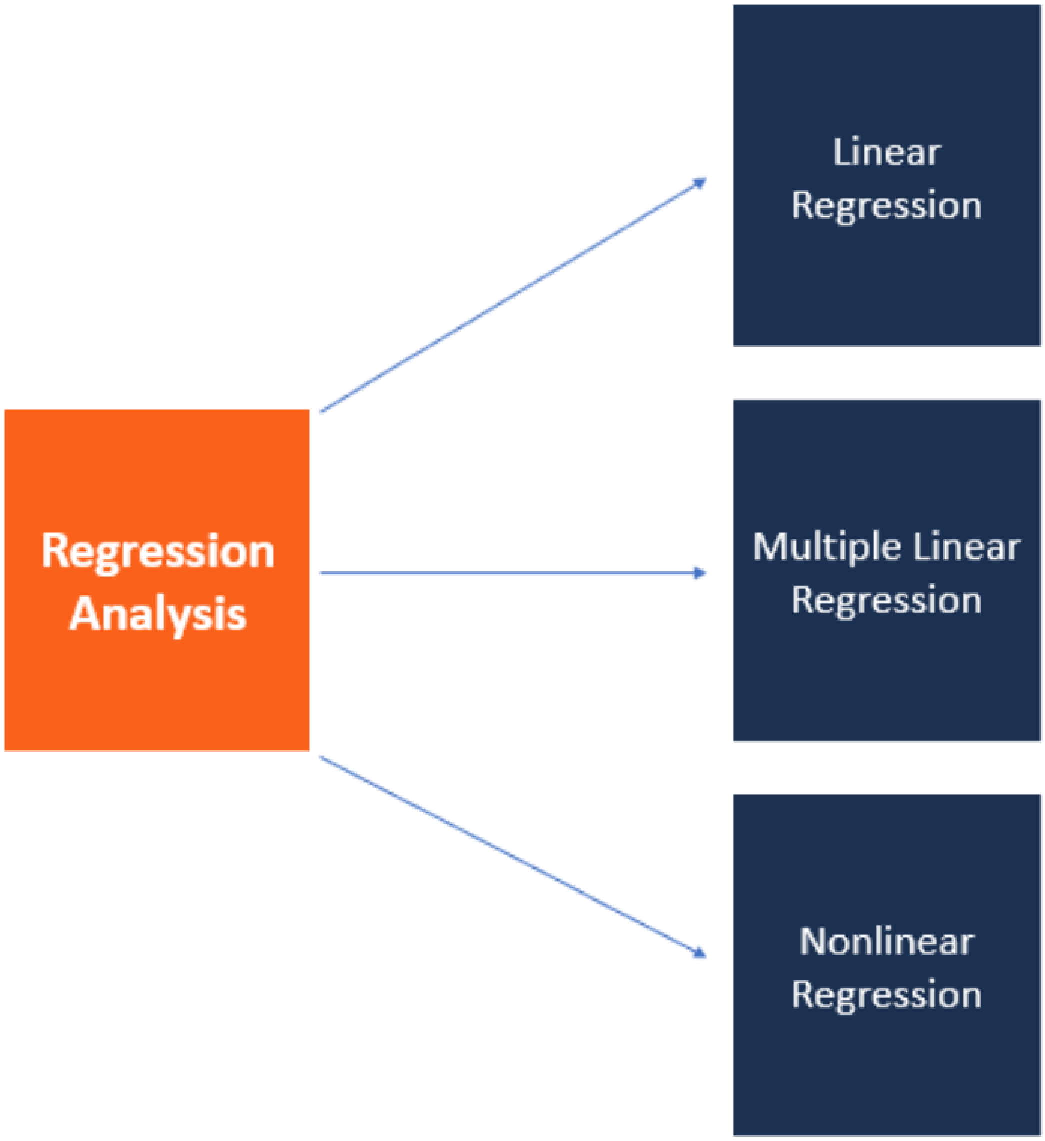
The point estimation here will be of the population proportion. The characteristic of interest is the teacher trainees having a favorable view about the services provided to them. So, all trainees with a favorable view are successes, $x = 103$. And $n = 150$. that means

$$\hat{p} = \frac{x}{n} = \frac{103}{150} = 0.686.$$

The researchers of this survey can establish the point estimate, which is the sample proportion, to be 0.686 or 68.7%.

3. Regression Analysis:

Definition: Regression analysis is a statistical method used to examine the relationship between one dependent variable and one or more independent variables. It is commonly used for prediction and understanding the strength and nature of relationships.



Example: Predicting a person's salary based on their years of experience and education level using a regression model.

Formula

$$Y_i = f(X_i, \beta) + e_i$$

Y_i = dependent variable

f = function

X_i = independent variable

β = unknown parameters

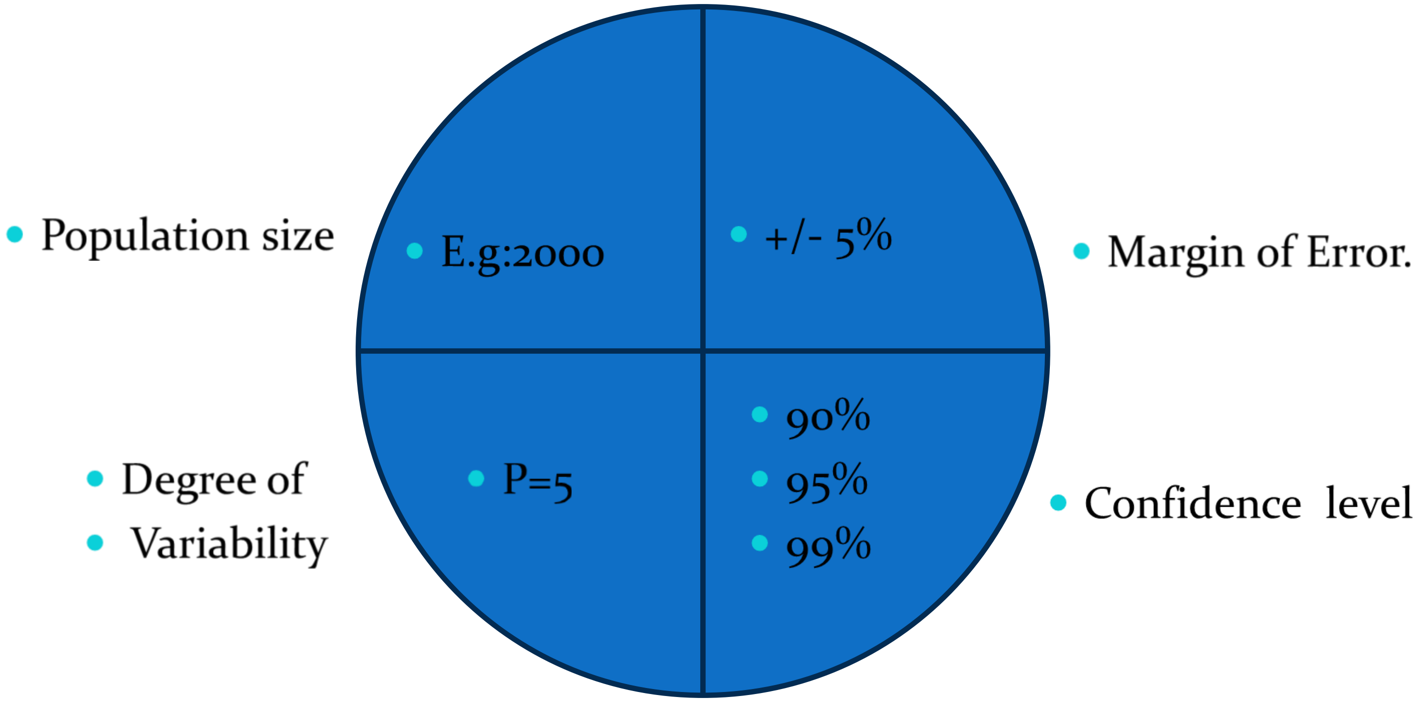
e_i = error terms

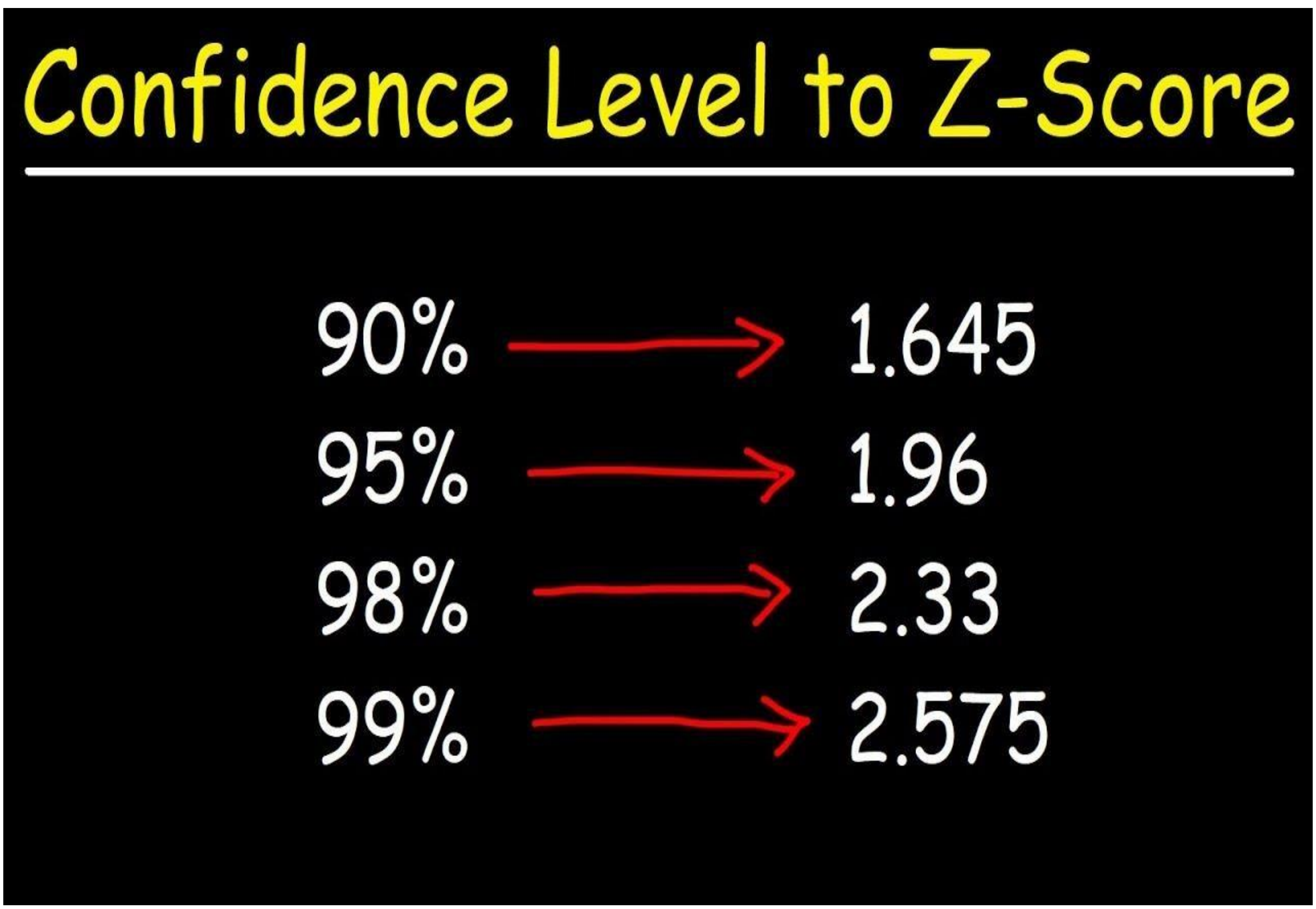
4. Confidence Intervals:

Definition: A confidence interval is a range of values derived from sample data that is likely to include the true population parameter with a certain level of confidence. It provides a measure of the uncertainty associated with the point estimate.

- **Also known as risk level**
- **Based on the central limit theory, which states that when the population is repeatedly sampled, the average value of the attribute obtained by those samples is the true population value**
- **It usually expressed in percentages (e.g: 95%)**

Example: A 95% confidence interval for the average weight of a certain species of bird might be (45 grams, 55 grams).


Population size
E.g:2000
+/- 5%
Margin of Error.
Degree of
Variability
P=5
90%
95%
99%
Confidence level


Confidence Level to Z-Score
90% 1.645
95% 1.96
98% 2.33
99% 2.575

1) Cochran formula
(Infinite and large population)

$$n_0 = \frac{Z^2 \times p \times q}{e^2}$$

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

2) Yamane formula
(Finite population)

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

3) Mean formula
(Finite population)

$$n = \frac{Z^2 \sigma^2}{e^2}$$

$$n = \frac{p(1-p)Z^2}{e^2}$$

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

1) Cochran formula

$$n_0 = \frac{Z^2 \times p \times q}{e^2}$$

$$n = \frac{n_0}{1 + \frac{(n_0 - 1)}{N}}$$

n_0: Sample size
Z: Value found in the Z-Table at a given confidence level
p: Estimated proportion of an attribute that is represented in the population
q: 1-p
e: Desired level of precision

Assume there is a large population with no known variability in the proportion that will adopted the land reform policy. Therefore, we assume p=0.5 (Max Var.). Suppose we desire a 95% confidence level and ±5% precision. The resulting sample is as follows:

$$n = \frac{(1.96)^2 \times 0.5 \times 0.5}{(0.05)^2} = 385$$

Example: Conducting a survey to assess how much customers are fully satisfied with the product and services.

- Assume:

Variability (p)=0.5

Confidence level=95%

Sampling error =5%

Find the sample simple of 100000 population

5. Hypothesis Testing:

Definition: Hypothesis testing is a statistical method used to make inferences about population parameters based on sample data. It involves formulating a hypothesis, collecting data, and assessing whether the data provides enough evidence to reject or fail to reject the null hypothesis.

Example: Testing whether a new drug is effective by comparing the outcomes of a treatment group to a control group.